FACULTAD DE FILOSOFÍA

GRADO EN FILOSOFÍA

Lógica e Inteligencia artificial.
Problemas del razonamiento retractable en Pollock.

Trabajo Fin de Grado presentado por Nerea Muñiz Castro, siendo la tutora del mismo la profesora Dra. Dª Cristina Bares Gómez

Sevilla. Agosto de 2022

GRADO EN FILOSOFÍA

FACULTAD DE FILOSOFÍA

TRABAJO FIN DE GRADO

CURSO ACADÉMICO [2021-2022]

TÍTULO: **Lógica e Inteligencia artificial. Problemas del razonamiento retractable en Pollock.**

AUTOR: **Nerea Muñiz Castro**

TUTOR: **Dra. Dª Cristina Bares Gómez**

DEPARTAMENTO: **FILOSOFÍA Y LÓGICA Y FILOSOFÍA D ELA CIENCIA**

ÁREA DE CONOCIMIENTO: **Lógica y Filosofía de la Ciencia**

RESUMEN:

En este trabajo veremos el estudio sobre el razonamiento retractable realizado por John L. Pollock. Pollock une IA y Filosofía para crear una teoría sobre un razonamiento dinámico que partiendo de entradas de percepción va desarrollando un conjunto de creencias que irán siendo retractadas, reforzadas o desechadas por la llegada de nuevas informaciones. Pollock elaborará una teoría de la garantía donde se mostrará como el razonamiento procediendo mediante el uso de argumentos, a través de reglas inferencia se llevan a cabo estos cambios donde las creencias seguirán siendo invictas o serán derrotadas. La teoría de Pollock es computacional, el le da importancia a que la teoría pueda ser implementada en agentes artificiales, por lo que, acabará creando un marco de representación del funcionamiento del razonamiento retractable llamado OSCAR donde se mostrará de forma gráfica como debe actuar un agente inteligente capaz de llevar a cabo un razonamiento retractable.

PALABRAS CLAVE:

Conocimiento, Razonamiento, IA, Pollock, Lógica

ABSTRACT:

In this work we will see the study on defeasible reasoning carried out by John L. Pollock. Pollock links AI and Philosophy to create a theory about a dynamic reasoning that, starting from perception inputs, develops a set of beliefs that will be retracted, reinforced or discarded according to new information. Pollock will develop a theory of guarantee where it will be shown how reasoning proceeds through the use of arguments, through inference rules these changes are carried out where beliefs will remain undefeated or will be defeated. Pollock's theory is computational, he attaches importance to the fact that this theory can be implemented in artificial agents, so he will end up creating a representation framework for the operation of retractable reasoning called OSCAR, where it will be shown graphically how an intelligent agent should act in order to be able of defeasible reasoning.

KEYWORDS: Knowledge, Reasoning, AI, Pollock, Logic

ÍNDICE

INTRODUCCIÓN

1.1 ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?

La Inteligencia Artificial es una de las ciencias más recientes en la actualidad. Surgió después de la Segunda Guerra Mundial y recibió su nombre al ser sugerido por John McCarthy en los años cincuenta. No hay una única definición de lo que es la IA. Como la IA trata diferentes aspectos de la inteligencia cada investigador suele dar una definición asociada al campo que trabaja. En (Russell, Norving,1978) se dan cuatro definiciones sacadas de ocho textos diferentes en la Figura 1.1[1]:

Sistemas que piensan como humanos	Sistemas que piensan racionalmente
«El nuevo y excitante esfuerzo de hacer que los computadores piensen… máquinas con mentes, en el más amplio sentido literal». (Haugeland, 1985)	«El estudio de las facultades mentales mediante el uso de modelos computacionales». (Charniak y McDermott, 1985)
«[La automatización de] actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje…» (Bellman, 1978)	«El estudio de los cálculos que hacen posible percibir, razonar y actuar». (Winston, 1992)
Sistemas que actúan como humanos	**Sistemas que actúan racionalmente**
«El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia». (Kurzweil, 1990)	«La Inteligencia Computacional es el estudio del diseño de agentes inteligentes». (Poole *et al.*, 1998)
«El estudio de cómo lograr que los computadores realicen tareas que, por el momento, los humanos hacen mejor». (Rich y Knight, 1991)	«IA… está relacionada con conductas inteligentes en artefactos». (Nilsson, 1998)

Figura 1.1 Algunas definiciones de inteligencia artificial, organizadas en cuatro categorías.

La IA es una ciencia que trata de dar respuesta a la pregunta formulada en (Turing,1950). Alan Turing es considerado el padre de la IA. Turing sugirió una prueba para considerar o no inteligente a una máquina, ésta consiste en tener una conversación con un interlocutor que no vemos, si este interlocutor es capaz de hacer que creamos que es una persona, entonces se puede entender que es una computadora inteligente. Según (Delgado,1996), Mc Culloch en un artículo titulado *Cálculo lógico de las ideas inmanentes en la actividad nerviosa* proponía que una red de neuronas trabaja como una máquina de Turing que usando un sistema de códigos binario puede realizar cualquier operación. Es decir, el cerebro puede ser visto como un ordenador. Esta idea contribuyó al desarrollo de la IA y de la Ciencia de la Información, aunque con el tiempo fue decayendo.

El nacimiento de la Inteligencia Artificial se suele situar en 1956, en la llamada Conferencia de Dartmouth. En esta conferencia se reunieron científicos de diferentes

[1] Este cuadro se encuentra en Rusell, S.J; Norving, *Inteligencia Artificial. Un enfoque Moderno.*,P.p.2. Estas definiciones pertenecen a Bellman, R.E (1978) *Artificial Intelligence: Can Computers Think?,* Nilsson, N.J. (1998), *Artificial Intelligence: A New Synthesis*, Rich, E. and Knight. K (1991) *Artificial Intelligence* (second edition) y Winston, P. H (1992) *Artificial Intelligence* (Third Edition)

disciplinas: neurólogos, psicólogos, ingenieros eléctricos y matemáticos. Aunque pertenecían a distintas disciplinas todos estaban usando computadoras para tratar ciertos aspectos de la inteligencia humana. La conferencia fue organizada por John McCarthy, Marvin L. Minsky, Nathaniel Rochester y Claude E. Shannon. Los objetivos de esta conferencia era examinar los distintos aspectos de la inteligencia de forma tan precisa que se pudiera construir una máquina que la simulara.

Así nace la IA, que como veremos en el siguiente apartado, junto con la Lógica, acabarán cooperando para lograr avanzar en sus objetivos a través del estudio de la inteligencia y del razonamiento humano.

1.2 LÓGICA E INTELIGENCIA ARTIFICIAL

La IA como hemos visto del punto de vista computacional y hay muchos desarrollos que siguen éste, pero, la base es el estudio del razonamiento y la observación del comportamiento racional. El primer estudio del razonamiento, o más bien de la validez de los argumentos fue tratado por Aristóteles, el fundador del silogismo. Los silogismos son esquemas de argumentos que representan la argumentación correcta, donde partiendo de premisas verdaderas se llegan a conclusiones necesariamente verdaderas. De aquí partió la lógica. Como se cuenta en (Carnota, 2005), en un principio la IA y la Lógica eran campos independientes unos de otros, pero a finales de los años setenta y comienzo de los años ochenta se comenzó a tener interés por otros campos que antes no se tomaban en cuenta. Entonces, la IA empezó a utilizar resultados de la Lógica y los proyectos de la IA motivaron a la Lógica a buscar respuestas sobre ciertas preguntas. A pesar de que la lógica clásica ya no es los silogismos aristotélicos, sigue una estructura de razonamiento que no permite un razonamiento cercano a los agentes reales. La lógica clásica partía de un contexto vacío, recursos ilimitados por parte del agente para procesar información e inferencias infalibles. Pero el mundo real cambia, hay limitaciones y los agentes crean nueva información, también la descubren y la intercambian. Pero al partir la Lógica Clásica de unas premisas para llegar a la conclusión, al añadir nuevas premisas, al obtener nueva información, no se podía revocar la conclusión. Surge entonces la necesidad de nuevas lógicas que contemplen esto. Crear sistemas que permitan la revisión de creencias, las lógicas no-monótonas, lógicas dinámicas que permitan manejar el cambio de las inferencias, lógicas lineales que toman la memoria y las limitaciones que tenemos para razonar. Esto lleva a un nuevo concepto de racionalidad que es lo que ha promovido muchos de los avances en Lógica y en IA, aunque también es fuente de muchos problemas abiertos.

1.2.1 Problemas y metas de la IA

La IA tiene distintas áreas de investigación, la cuales, tienen distintos objetivos y conllevan una serie de problemas. Siguiendo a (Delgado, 1996) vamos a ver algunas áreas de la IA. Una de estas áreas recibe el nombre de “Game Playing”. Esta área se encarga de desarrollar programas de ordenador capaces de jugar a juegos intelectuales de alto nivel como podría ser el ajedrez. Se considera que tiene comienzo en los sesenta con Arthur L. Samuels. Intentó emplear la capacidad y la velocidad de un ordenador para realizar secuencias de movimientos propios de las partidas de ajedrez. Esto solo llegó a funcionar con juegos más básicos como el tres en raya. Esto se debe a que en el tres en raya las posibilidades de juego son muy reducidas, entonces el ordenador puede analizar cada una de ellas, pero, en los juegos más complejos como el ajedrez las posibilidades aumentan a un número considerable que la máquina no puede analizar. Por eso surge la necesidad de añadir a estos programas la capacidad de aprendizaje y el “conocimiento experto”. De este modo, se logran reducir las posibilidades a tratar de un modo similar al del razonamiento humano. Este tipo de

programas suele superar a los humanos cuando juegan con un tiempo limitado, pero, inversamente, si no hay límite de tiempo, suelen acabar ganando los humanos.

Otra área para tratar por la IA es la Lógica y la demostración de teoremas. El interés por esta área ha ido creciendo a lo largo de los años. Las técnicas de mecanización de la Lógica varían de la forma en que trabaja el razonamiento humano en ciertos aspectos. El método elaborado por John Alan Robinson nombrado como "Método de Resolución" es uno de los más populares, aunque cuando los conjuntos de posibilidades son muy grandes, la búsqueda comienza a dar problemas en el camino de la deducción. Vuelve a surgir la necesidad que vimos en los juegos de ordenador, se necesita reducir la búsqueda a niveles más factibles incorporando otros métodos.

La Resolución de problemas es una de las disciplinas que más ha interesado porque parece requerir un grado de inteligencia. Al final, el objetivo no se ha centrado tanto en la resolución de problemas, sino, más bien, en la representación del conocimiento y en desarrollar estrategias para dar soluciones.

Ahora vamos a hablar de los "Sistemas Expertos", los cuales, están basados en el conocimiento. Por lo general, estos sistemas usan el conocimiento para poder, dentro de un dominio concreto, resolver distintos problemas como lo haría, por ejemplo, un experto en su materia. Además, por lo general, no solo tienen que ser capaces de resolver los problemas, sino que, además, deben ser capaces de explicar los pasos razonamiento que ha llevado a cabo hasta llegar a resolverlos y de los datos que ha usado para ello. Pero hay algunos sistemas de este tipo que no tiene que dar estas explicaciones, son sistemas basados en conocimiento que no son explicativos.

Otro campo muy importante es el de Aprendizaje automático, el cual, desde la práctica puede justificarse por los avances en los procesos de adquisición de conocimiento, aunque, desde el punto formal, el desarrollo de la capacidad de aprendizaje es uno de los más importantes de cara a conseguir proporcionar inteligencia a las máquinas, puesto que, es una parte crucial en la inteligencia humana. El problema es que de esto hay poca información, como veremos próximamente cuando hablemos del conocimiento, hay distintas teorías acerca del conocimiento, de su estructura y de la forma de obtenerlo, por lo que se dificulta el desarrollo de sistemas de aprendizaje automático. Otra dificultad para estos sistemas viene del poco conocimiento de la memoria humana. Las máquinas tienen una memoria distinta a la humana y son capaces de encontrar información en ésta al momento, pero, los humanos, tenemos problemas para hacerlo, pero, somos capaces en dividir la información según su relevancia y de hacer uso de información incompleta o incluso contaminada, cosa que cuesta trasladar a los sistemas, ya que, no sabemos muy bien acerca del funcionamiento de estos procesos.

También es importante de cara a crear un sistema inteligente hacer que la máquina sea capaz de crear frases y de comprender el lenguaje. De esto se encargan áreas que investigan el proceso del lenguaje natural. Esto es complejo puesto que, aunque teóricamente se establecen algunas reglas para el uso del lenguaje, el uso que se lleva a cabo realmente no las sigue y puede llevar a ambigüedades, o en ocasiones, incluso hacemos un uso incorrecto de éste. En los primeros programas de traducción se observó precisamente, que para llevar a cabo una interpretación del lenguaje no es suficiente con tener un vocabulario y una gramática. Para llevar a cabo una traducción es necesario poder comprender el texto que se ha de traducir. Muchos dieron por imposible la creación de un programa que fuera capaz de llevar a cabo una traducción automática. Pero R. Shank propuso que se creara una especie de intermediario que representara el sentido del texto y luego el programa tradujera esta representación. Por lo tanto, hacían fata dos programas; el Analizador, que ha de comprender el texto, y el

Generador, que llevaría a cabo la traducción. Otros problemas volverían a ser, la búsqueda de información y la recuperación.

Estas son algunas de las áreas más importantes a tratar por la IA. Como hemos visto todas tienen sus problemas pues el funcionamiento de la inteligencia humano es un poco misterioso para nosotros y, por lo tanto, crear inteligencia lo es aún más. Pero algunos problemas ocurren en distintas áreas, por lo que las soluciones son trabajadas desde distintos campos, e incluso, podría parecer que algunos de los problemas se podrían reducir incorporando métodos de otras áreas como hemos visto por ejemplo en los juegos de ordenador donde se introduce "conocimiento experto" y la capacidad de aprendizaje. Al final, aunque las áreas se dividan, el conjunto de ellas, incluidas las que no se han nombrado aquí, son las que podrían posibilitar la creación de inteligencia artificial.

A continuación vamos a pasar a sobre el razonamiento retractable desde John L.Pollock, un filósofo estadounidense conocido por sus trabajos en epistemología, lógica, ciencia cognitiva e inteligencia artificial. Primero hablaremos un poco sobre el conocimiento y de algunos problemas que se han tratado con respecto a este y la forma en la que adquirimos y actualizamos el conocimiento. Después entraremos en la teoría de la garantía de Pollock para acabar este apartado entrando en el marco general de razonamiento retractable creado por Pollock, el cual, es nombrado como OSCAR.

RAZONAMIENTO RETRACTABLE

2.1 CONOCIMIENTO

Según la RAE el razonamiento se trata de "acción y efecto de razonar" y razonar es "ordenar y relacionar ideas para llegar a una conclusión". Pero estas definiciones acerca del razonamiento pueden parecer un poco incompletas. El razonamiento no actúa de la misma forma, no todo es deducir algo partiendo de unas premisas. Existen distintos tipos de razonamiento aparte del deductivo como puede ser el razonamiento subjuntivo o el razonamiento retractable[2].

A través del razonamiento vamos adquiriendo conocimiento y según nuestro conocimiento el razonamiento se mueve de cierta forma y extrae cierta información. Según Pollock (1975) el conocimiento puede separarse en distintas áreas; el conocimiento del mundo físico es adquirido mediante percepción, la fuente del conocimiento del pasado es la memoria, el conocimiento de verdades contingentes proviene de la inducción, el conocimiento de otras mentes nos llega a través de la interpretación de los cuerpos y el conocimiento a priori y de las verdades morales cuya procedencia no se sabe muy bien.

Pero hablar de conocimiento lleva consigo ciertos problemas. Los problemas más relevantes surgen entorno a la justificación, ya que la definición clásica de conocimiento es creencia justificada verdadera, pero… ¿Qué nos justifica para creer en algo? En (Gettier,1963) Trata de establecer cuáles son las condiciones necesarias para estar justificado a tener una creencia verdadera. Rechaza las condiciones que se habían dado hasta el momento por autores como Chisholm y precisamente pone en tela de juicio la definición clásica. Gettier dice que la justificación no es razón suficiente para tener una creencia verdadera y expone dos casos para ilustrarlo.

[2] Téngase en cuenta que vamos a tratar los diferentes tipos de razonamiento desde el punto de vista de Pollock, ello no implica que sean los únicos. Es más, otros autores usarán una clasificación diferente.

En el primer caso expuesto hay dos personas, Smith y Jones, las cuales acuden a una entrevista de trabajo. Parece que Smith tiene una fuerte evidencia para creer que Jones obtendrá el trabajo porque el director le ha dicho que va a contratar a Jones y que Jones tiene diez monedas en el bolsillo porque Smith las ha contado. Esto lleva a la implicación de que el hombre que obtendrá el trabajo tiene diez monedas en el bolsillo. Smith está justificado a creer en esta implicación partiendo de la evidencia. Pero entonces Gettier nos dice que imaginemos que el que va a obtener el trabajo sin saberlo es Smith y que sin saberlo tiene diez monedas en su bolsillo. De aquí se llega a que la segunda proposición es verdadera, Smith está justificado para creer que es verdadera, pero no sabe que es verdadera puesto que esa verdad se basa en que tiene diez monedas en el bolsillo y él no lo sabe. Smith cree en la proposición porque cree que Jones va a obtener el trabajo y cree que tiene diez monedas en el bolsillo. Smith tiene la creencia y está justificado para tenerla, pero, sin embargo, no tiene conocimiento, no sabe.

En el segundo caso, Gettier nos dice que supongamos que Smith cree que Jones tiene un Ford porque en su memoria tiene recuerdos de este conduciendo uno y se ha montado con él en un Ford. También tenemos que suponer que Smith tiene un amigo que no sabe dónde está. Smith construye tres proposiciones con lugares aleatorios:

Jones tiene un Ford o Brown está en Boston.

Jones tiene un Ford o Brown está en Barcelona.

Jones tiene un Ford o Brown está en Francia.

La primera proposición tiene evidencias para ser cierta, por lo que asume todas como verdaderas. Está justificado para creer las tres, pero imaginemos que el Ford no es de Jones pero que sí que está en Barcelona. La creencia sigue siendo verdadera y Smith está justificado para creerla, pero, no lo sabe, no tiene conocimiento de ello.

Así demuestra Gettier que para tener conocimiento no es suficiente con que una creencia sea verdadera y tengamos evidencias. Nuestras creencias influyen en nuestra forma de inferir. Podemos estar justificados para tener una creencia, pero no implica que ésta sea verdadera.

Pollock (1975) dice que está por un lado el problema de nombrar las condiciones de justificación para cada una de las áreas de conocimiento y por otro en demostrar que esas son las condiciones de justificación. Por lo general, los epistemólogos no dudan de que obtenemos conocimiento partiendo de argumentos inductivos y de la experiencia, pero el problema es el demostrarlo. Los escépticos llevan a la duda de que podamos creer o no en lo que percibimos y en aquello que extraemos de la percepción. Pero Pollock dice que el argumento de los escépticos es falso, debe haber en alguno de los pasos de su argumento un error. Hay que dar con ese error para acabar con este problema.

Nuestro conocimiento del mundo nos llega a través de modos de intuición. Obtenemos conocimiento del mundo a través de lo que percibimos por los cinco sentidos. Los filósofos dicen, según Pollock, que los modos de intuición no son solo las causas de nuestras creencias, sino que son su fundamento lógico. Tenemos de primeras unas creencias sobre el contenido de las intuiciones. A cada intuición se le asocia unas creencias. Se ha supuesto que estas creencias son incorregibles y en estas se basan el resto de las creencias. Los modos de intuición conectan con su creencia incorregible correspondiente. Se genera así una teoría piramidal del conocimiento, donde, la base

de este está compuesta por un tipo de proposiciones que no necesitan justificación externa. Para creer en este tipo de proposiciones no se necesitan razones. Esas proposiciones las cuales podemos creer sin ninguna razón parece que en cierto modo se justifican a ellas mismas.

Los filósofos a menudo han pensado que la justificación va ligada a un hecho. Pero más que en el hecho en sí, la justificación va ligada a la creencia de una persona en un hecho. Por ejemplo, en el segundo caso que nos exponía Gettier, Smith pensaba que Jones tenía un Ford porque recordaba haber visto a Jones conduciendo uno e incluso porque se había montado con él. Finalmente resultó que era alquilado y que la proposición que decía que Jones tenía un Ford no era verdadera, pero, Smith estaba justificado para creer en ella basándose en los hechos que el recordaba.

Pero el tener una creencia no justifica directamente otra creencia, se ha de tener una creencia que esté justificada. Ocurre que no se sabe realmente si hay proposiciones que se autojustifiquen, las cuales son llamadas por Pollock “proposiciones básicas”. La teoría de la nebulosa niega la existencia de estas proposiciones, lo que lleva a que, al no haber una base de conocimiento, las creencias se encuentran todas al mismo nivel y el ciclo de justificación va dando vueltas de una creencia a otra terminando en cualquier lugar. La justificación no se detiene. La idea de que para que una persona pueda tener una creencia justificada deba apelar conscientemente a otras creencias que además también deben estar justificadas llevaría a que esa persona debería llevar a cabo una cantidad inmensa de pasos para estar justificados para creer en algo. Esto sería imposible, por lo que, no se podría justificar ninguna creencia. Pollock dice que como esto es falso se seguiría que la teoría de la nebulosa es falsa. Pero podría darse que no sea necesario pensar explícitamente en la conexión entre varias creencias, sino que, solo sea necesario tener otra creencia justificada que sea razón para justificar la nueva creencia. De esta forma se anularía el problema del retroceso infinito para llevar a cabo la justificación. No es necesario pensar conscientemente en la justificación para estar justificado para tener una creencia.

Pero, aunque la regresión infinita no sea el problema puede que haya otro en la teoría de la nebulosa. Al final la teoría de la nebulosa nos dice que tenemos una serie de creencias y que una creencia se puede justificar si alguna de estas la respalda. Es decir, S tiene una secuencia de creencias $Q_1,\ldots,Q_n$, para creer en P. Se necesita que algunas de estas creencias la respalden y cada una de estas creencias están respaldadas por otras creencias de la secuencia. Esto haría que las creencias no tuvieran que corresponderse necesariamente con el mundo, simplemente se debería tener una secuencia coherente de creencias. Pollock nos dice que imaginemos a una persona con creencias que la justifican a rechazar las evidencias de sus sentidos. Esta persona creería justificadamente que aquello que percibe como caliente es frío, lo estrecho como ancho y lo alto como bajo. Esto estaría bien siempre y cuando sus creencias formen un conjunto coherente. Si observamos, esto realmente no tiene mucho sentido. Alguien puede estar justificado a creer que algún sentido lo engaña, como una persona daltónica, por ejemplo, pero para esto ha de tener evidencias empíricas de estos sentidos. Pero que una persona rechace todo el tiempo lo que le dicen sus sentidos justificadamente es imposible. Al permitir esto, se demuestra que la teoría de la nebulosa es errónea. No puede haber una cadena infinita de justificación, tiene que haber unos pasos finitos. Pollock nos dice que esto no tiene porqué llevarnos de nuevo a la teoría tradicional de la incorregibilidad. Si se acepta que la justificación siempre procede de otras creencias parece que necesariamente deban existir proposiciones epistemológicamente básicas.

Pero ¿podríamos pensar en la justificación de otra forma? En (DeVries, 2011) se muestra como Willfrid Sellars lleva a cabo una crítica a lo que llama “el mito de lo dado” donde niega la idea tradicional de que ha de haber algunos estados cognitivos en contacto directo con la realidad que sirvan de base sobre la que construirse el resto del conocimiento. Para Sellars no solo no debe haber algo dado, sino que no puede haberlo. Pollock no estará de acuerdo con esta tesis y supondrá que existen proposiciones epistemológicamente básicas.

Hasta ahora lo que se ha dicho es que estas proposiciones no necesitan razones, se justifican ellas mismas. Esto, para Pollock, no quiere decir que tengan que ser incorregibles, sino que, si una persona S tiene la creencia incorregiblemente justificada de que P, quiere decir que S crea P es lógicamente suficiente para que su creencia esté justificada. Igual que una creencia falsa puede estar justificada, una creencia incorregiblemente justificada puede ser falsa, no tiene por qué ser incorregiblemente verdadera. Pero es cierto que, aunque no tengamos razones para justificarlas, sí que pueden existir razones para no creerlas. Esto quiere decir que, nuestras creencias epistemológicamente básicas solo pueden justificarse prima facie si tenemos estas creencias y no tenemos otras creencias que constituyan razones para rechazarlas. Una razón prima facie es una razón que por sí misma sería una buena razón para creer algo y garantizar la justificación, pero puede dejar de ser una buena razón cuando se toma con otras creencias adicionales. Es decir, para que S tenga como creencia epistemológicamente básica P, no necesita ninguna creencia que justifique P y no puede existir ninguna creencia en S que lleve a ~P o derrote P. La idea detrás de una creencia prima facie justificada radica en que hay una “presunción lógica” a favor de que la creencia está justificada. Hay teorías como la de Armstrong que rechazan la teoría de la incorregibilidad. Armstrong (1963) defiende que el modelo introspectivo no puede ser incorregible y que además no tenemos un acceso lógicamente privilegiado a nuestra existencia mental. Para Armstrong el conocimiento de la mente es análogo al perceptual. A través de la percepción sensorial accedemos a los objetos del mundo y a través del “sentido interno” mediante la introspección accedemos a los objetos de la mente. EL proceso de la introspección puede producir errores. Si en un futuro en base a una teoría neurofisiológica de la percepción se pudiera construir un “super electroencefalograma” (SEEG) que fuera capaz de decir en qué estados de percepción nos encontramos, una persona podría decir que le parece ver algo rojo y la SEEG le informaría de que en realidad le parece verde, entonces, si esto pasara, tendríamos razones fuertes para pensar que la persona está equivocada y su creencia no podría ser incorregible. La persona hace un informe incorrecto sobre su propia experiencia y el SEEG representa una autoridad mayor que la propia persona. Armstrong (1963) presenta otro argumento donde una persona informa que tiene dolor ahora, pero, tardamos un tiempo en expresarnos. Por lo tanto, ¿A qué tiempo se refiere la palabra “ahora”? El tiempo de antes puede ser equivocado y el tiempo de después también, por lo cual, parece que el conocimiento indubitable debe abarcar solo el instante actual. Pero, no puede ser indubitable que al acabar la oración la persona seguirá teniendo dolor. A ese momento se le puede llamar “instante introspectivo” y el defensor de la incorregibilidad debería decir que el conocimiento incorregible solo es aquel que es conocimiento del “instante introspectivo” actual. Los instantes pasados se recuerdan, por lo que, puedo estar equivocado, y los instantes futuros solo se prevén, por lo tanto, también se pueden dudar de ellos. Pero esto los tendría que llevar a reconocer que, al menos en la práctica, no podríamos hacer declaraciones sobre el estado lógico de las propias experiencias internas, ya que, al terminar de hablar ya nos estaríamos refiriendo al pasado. Otro argumento que realiza es que para hablar de adquisición de conocimiento es necesario que exista el error, pues, para que exista la victoria tiene que existir la derrota. El último argumento que presenta Armstrong en lo que respecta a la negación de la

incorregibilidad en (Armstrong,1963) consiste en que una cosa es estar en un cierto estado mental y otra cosa es ser consciente de estar en él. La adquisición de conocimiento introspectivo debe consistir en la elaboración de informes sinceros de ocurrencias mentales o en la aprehensión no verbal de estas ocurrencias, pero estas pueden existir independientemente de que sean reportados o aprehendidas, por lo que no podrían llamarse “experiencias” y no habría conocimiento introspectivo.

Hasta aquí hemos visto que nuestro conocimiento tiene una especie de estructura piramidal con distintos niveles de creencias. La base del conocimiento, hemos dicho, está constituida por creencias epistemológicamente básicas. Los niveles superiores de creencias se van justificando en las creencias de niveles inferiores. Cuando una creencia justifica a otra se dice que esta es una razón para la segunda creencia. Pollock hablará de forma indistinta de creencias y razones porque las razones suelen ser creencias.

Hay razones que digamos, tienen más fuerza que otras. Si, por ejemplo, llevo a cabo una cantidad de experimentos para probar una hipótesis, cuantos más experimentos hago más fuerte es la razón para confirmar o rechazar la hipótesis. La realización de doce experimentos es una razón más débil para comprobar una hipótesis que la realización de cincuenta experimentos. Pero, por ejemplo, si mi hipótesis implica que todos los experimentos que lleve a cabo debieran tener el mismo resultado, en el caso de que un solo experimento llevara a algo distinto, este a pesar de ser solo uno, sería una razón más fuerte para llegar a la creencia de que la hipótesis es errónea. Es decir, una razón es buena si es lo suficientemente buena para justificar la creencia de la que es razón. Las razones no siempre tienen que ser buenas razones. Hay algunas que no justifican aquello para lo que son razones. Puede darse el caso de que una razón sola sea buena porque no hay razones contradictorias, pero, al juntarla con otras razones ya no sea buena.

Las razones buenas pueden ser de distintos tipos. Un tipo de razones buenas son las razones lógicas. Estas se basan en el significado de los enunciados para justificar la creencia. P es una creencia lógica para que S crea Q según Pollock (1975) si (1) S cree justificadamente que P, (2) P es una razón lógica para que S crea Q y (3) P es una buena razón para que S crea Q. Son como lo que llamamos verdades analíticas, las cuales contienen la consecuencia en las propias premisas y las razones están en los significados de estas. Por ejemplo, una verdad analítica sería que los cuervos son negros y la creencia de que todos los cuervos son negros puede ser una razón lógica para que una persona S crea que el pájaro negro que está viendo sea un cuervo. Hay que tener en cuenta que el hecho de que P sea en este momento una razón lógica para que S crea Q, no quiere decir que vaya a serlo siempre. John L Pollock (1987) habla de razones que no son anulables porque son concluyentes, pero también habla de razones que, si son anulables, las cuales, al añadir información se puede destruir la conexión de la razón.

Otro tipo de razones son las razones contingentes. Estas son razones que no son lógicas y a menudo se han pasado por alto. Pero es cierto que ambas suelen estar estrechamente relacionadas. Suele ocurrir que para que una creencia sea una razón contingente una persona debe tener también una razón lógica para creerlo. Aquí surge el problema de si las razones contingentes pueden reducirse a razones lógicas, que, al parecer, si se puede. Por definición, para que P sea una razón contingente para que S crea Q, S debe tener una razón para creer (P→Q). Es decir, para que P sea una razón contingente para que S crea Q, S debe tener una razón donde relacione P con Q. Por ejemplo, si el hecho de que mi madre me haya dicho que alguien se ha comprado la casa de al lado es una buena razón para creer que alguien se ha comprado la casa de al lado quiere decir que creo justificadamente en que si mi madre me dice que alguien

ha comprado la casa de al lado, alguien ha comprado la casa de al lado. La creencia que conecta P y (P→Q) es una razón lógica para que S crea que Q. Esto lleva a que, aunque existen dos tipos de buenas razones, siempre se pueden reducir a razones lógicas.

Pollock dice que las implicaciones simples son implicaciones lógicas, pero, esto no quiere decir que todas las implicaciones lo sean. En implicaciones más complejas podríamos creer P sin saber que P→Q y no estar justificados para creer Q. Para que una implicación simple sea una buena razón basta con remitirse a los significados de los enunciados, pero, en implicaciones más complejas las buenas razones solo pueden conocerse después de dar una demostración. Dar esta demostración necesaria es adquirir una razón independiente para creer que P es verdadero y que entonces Q también lo es. Estas demostraciones son buenas razones, sin embargo, no son razones lógicas.

Hay otra clase de razones lógicas llamada razones concluyentes. Se dice que P es una razón concluyente para que S crea Q si y solo si P implica y es una razón lógica para que S crea Q. Al parecer se ha pensado por parte de muchos filósofos que las razones lógicas son todas concluyentes, pero, Pollock muestra como la inducción, por ejemplo, no se puede explicar en termino de razones concluyentes, pero, nos proporciona buenas razones. Esta suposición de que las razones lógicas son todas concluyentes se debe a la idea tradicional de que los problemas epistemológicos pueden solucionarse a través de análisis reductivos. Esto se apoya en la teoría piramidal de conocimiento, pero Pollock dice que cada vez más esta idea parece improbable porque en general no se pueden dar análisis reductivos. Entonces la probabilidad de que la justificación se de en términos de vinculación disminuye, llevándonos a suponer que hay muchas razones lógicas que, como dijimos con anterioridad, no son concluyentes. El no reconocimiento de la presencia de estas razones lógicas no concluyentes es lo que ha dado lugar a las problemáticas en las distintas áreas filosóficas. Muchas razones lógicamente buenas tienen estructuras de razones prima facie, por ejemplo, la inducción puesto que es una razón claramente lógica, pero, puede ser anulada por un contra ejemplo o por el descubrimiento de la no validez de una de las muestras en las que se basó la inducción. Otro ejemplo sería el de los juicios perceptuales. Siguiendo el ejemplo de Pollock (1975) si S ve x de color rojo, el percibir x como rojo es una razón lógica prima facie para que S crea que x es rojo. Parece que debe haber una conexión entre ambas porque de lo contrario no habría forma de saber que algo es rojo. Esta conexión debe ser lógica de algún tipo, pero no contingente porque esto llevaría a establecer la conexión inductivamente y no podemos debido a que sería necesario que pudiéramos decir de forma independiente qué cosas son rojas y solo somos capaces de hacerlo tomando las cosas que nos parecen rojas. Por lo general, si veo un objeto como rojo de forma clara y no tengo ningún motivo para creer que hay algo que está haciendo que se vea así o que mi ojo me está engañando voy a juzgar que el objeto que veo como rojo es rojo. Si no tengo creencias que lleven a lo contrario si veo algo como rojo estoy justificado para pensar que es rojo. Pero si tengo otras creencias como que hay una luz encima del objeto que lo hace verse rojo y que con luz natural se vería blanco, el hecho de que vea el objeto de color rojo no me justificaría para creer que el objeto es rojo. Es por esto por lo que S tenga la creencia “veo x de color rojo” es una razón prima facie para que S crea que “x es rojo”, porque puede ser anulada por otras creencias. Estas son llamadas por Pollock (1987) “derrotadores” y las define de la siguiente forma:

“R is a defeater for P as a prima facie reason for Q if and only if P is a reason for S to believe Q and R is logically consistent with P but (P&R) is not a reason for S to believe Q.” (Pollock, 1987, p.484)

Es decir, R es un derrotador para P como razón prima facie para Q si y solo si P es una razón para que S crea Q y R es lógicamente consistente con P, pero (P&R) no es una razón para que S crea Q.

Volviendo al caso anterior, si hubiera una luz encima del objeto que vemos como rojo que hiciera que lo viéramos así, la creencia "x está iluminada por una luz roja" sería un derrotador de "veo x rojo" y, por lo tanto, ya no estaríamos justificados para creer que "x es rojo".

Decir que P es una razón prima facie para que S crea que Q quiere decir que mientras S no tenga otra información está justificado para creer que no sería cierto P a menos que fuera cierto Q, es decir, para creer que P→Q. Un derrotador para este condicional de subjuntivo sería una razón para que S creyera que este condicional es falso, es decir, ~(P→Q). Hay dos tipos de derrotadores para negar un condicional. Los derrotadores de tipo I derrotan la verdad del condicional. Si P es una razón prima facie para que S crea que Q, entonces cualquier razón para que crea que Q es falsa, independientemente de que P sea verdadera, es un derrotador. En el caso de "x me parece rojo" luego "x es rojo", si un amigo de confianza me dijera "x no es rojo" podría ser suficiente para negar "x es rojo" siendo entonces derrotado el condicional P→Q. Este condicional se construye porque la mayoría de los objetos que nos parecen rojos son rojos. Creamos un vínculo y hacemos una generalización. Para que un condicional se cumpla la teoría lingüística según (Pollock, 1976) tiene que haber unas leyes físicas y unas circunstancias. La conjunción L^C^P implicaría Q. Si un amigo de confianza me dice que "x no es rojo", se convierte en una razón para ~Q y pensar que la generalización no se cumple ~(P→Q). La conexión se rompe si las circunstancias y las leyes físicas no se cumplen.

Los derrotadores de tipo II atacan a la conexión entre P y Q haciendo que saber P no tenga que llevar a Q. Es el caso por ejemplo en el que hay una luz roja sobre el objeto que vemos rojo. No es suficiente para saber que x no es rojo, pero si para romper con que si x me parece rojo entonces x es rojo.

En (Pollock,1976) habla de que es cierto que en algunos condicionales de subjuntivo existe conexión entre el antecedente y el consecuente y que tal conexión es condición suficiente para la verdad del subjuntivo, pero, esta conexión no es necesaria. Hay otros tipos de subjuntivos donde no existe esta conexión entre el antecedente y el consecuente, sino que precisamente porque el consecuente ya es verdadero y no existe conexión con el antecedente el condicional se cumple. Los derrotadores de tipo II atacarían a los condicionales de subjuntivos simples donde si existe esa conexión.

En la inducción, los derrotadores de tipo II consisten en razones para pensar que la muestra de la inducción no es justa, es decir, que su análisis no constituye una buena forma de averiguar si el predicado se satisface de forma universal. Un ejemplo puede ser el buscar demostrar que ningún gato tiene pelo mirando solo a los gatos esfinge y descubrir luego que esta es una raza muy particular y que el resto de raza de gatos tienen pelo.

Para que un derrotador actúe como tal no es necesario que la creencia de S en este sea justificada. Si P es una razón prima facie para que S crea Q, si S cree una proposición R, de forma justificada o no, que lleva a ~(P→Q), es suficiente para que sea derrotado y P no sea una buena razón para Q. Esto podría llevar a pensar que, si S tiene una buena razón para creer en un derrotador, pero no lo hace esto puede ser una buena razón para evitar la creencia de Q. Pero no debiera darse, puesto que, sería irracional que S no creyera en R siendo una buena razón, si lo hiciera, debería deberse a la ignorancia de esa buena razón para creer en el derrotador R y siendo así, S está justificado para creer Q.

En este apartado hemos hablado sobre el conocimiento, los distintos tipos de razones, la justificación y sobre los derrotadores. Las razones prima facie y los derrotadores son los responsables del carácter no monótono del razonamiento. Ahora en el siguiente apartado vamos a pasar hablar del razonamiento retractable y los problemas ligados a este.

2.2 RAZONAMIENTO RETRACTABLE EN POLLOCK

En este apartado vamos a pasar a hablar del razonamiento retractable y de los problemas que hay con respecto a este desde John L. Pollock . A lo largo de los siglos se le ha otorgado una gran importancia al razonamiento deductivo, se buscaba el llegar a unas conclusiones que fueran inamovibles. Pero ocurre que, como hemos visto, nuestro conocimiento no es absoluto ni cerrado. Vamos adquiriendo conocimiento, se trata de algo que va renovándose, se añaden creencias, se corrigen otras, se refuerzan o incluso se desechan. Como hemos visto en el apartado anterior, podemos tener unas creencias de forma justificada hasta este momento, con el conocimiento que tengo actualmente, pero, si añadimos información puede ser que ya no estemos justificados para creer en ciertas creencias. El razonamiento retractable consiste en esto, tengo una serie de creencias y si hay algún cambio en el contexto porque obtengo nueva información, la información que tengo cambia o pierdo alguna información, puedo desechar algunas creencias.

Pollock dice que el razonamiento retractable es algo similar al razonamiento no-monótono. En IA el razonamiento no-monótono se convirtió en un tema de gran interés casi al mismo tiempo en el que se comenzaron a desarrollar teorías sobre el razonamiento retractable. Según (Morado,2004) hay muchas personas que cometen el error de tratar el razonamiento no-monótono y el razonamiento retractable como si fueran lo mismo. Morado dice que no se debe reducir a lo mismo, puesto que, el razonamiento no-monótono lleva a retractar una conclusión debido a un aumento de información, mientras que, en el razonamiento retractable la puede retractar por distintos cambios en el contexto, no solo por un incremento de información, sino también por perdida o cambio de ésta. De hecho, la mayoría de las reglas default, por ejemplo, son no-monótonas, pero, algunas no son retractables. Las inferencias default están ligadas al contexto, por eso cuando ocurre un cambio en éste cambia el grado de razonabilidad. El razonamiento sigue unas reglas, pero la retractabilidad no quiere decir que estas reglas estén mal o se haya llevado a cabo un mal uso de estas, sino que esas reglas ya no son aplicables, quedan desactivadas. El conocimiento es incompleto, se modifica y se aumenta, es por eso que somos susceptibles de cometer errores y, por lo tanto, nuestras conclusiones y creencias han de poder ser revisadas y desechadas. Del mismo modo que los humanos cometemos errores, los sistemas artificiales también los cometen, por eso, es necesario que estos sean capaces de revisar sus interpretaciones frente a la llegada de nuevos datos.

2.2.1 Teoría de la Garantía

El razonamiento humano comienza partiendo de estados de entradas, percepciones, inputs. Con ayuda de la memoria se construyen argumentos, de los cuales surgen creencias y partiendo de estas se llega a otras creencias y así sucesivamente. A través del razonamiento podemos llegar a una creencia y luego llegar a rechazarla porque se rechazó otra creencia en la que se apoyaba. Pollock elabora una teoría del razonamiento en (Pollock, 1987) basándose en un ideal de máquina donde no hay límites. Existen argumentos lineales e indirectos. Los argumentos lineales son secuencias de proposiciones finitas donde cada una de esas proposiciones describen

la base epistémica o base de datos, o, donde el conjunto de miembros anteriores de la secuencia constituye una razón para esto. Los argumentos indirectos actúan tomando como premisas unas suposiciones para llegar a conclusiones, luego, esas premisas son descargadas usando alguna regla como la reducción al absurdo o la condicionalización, dejando así, de depender de las suposiciones. Se dice que un argumento apoya la conclusión P si y solo si P es apoyada por alguna línea de argumento.

Los argumentos se forman usando unas reglas de inferencia donde se justifica. Pollock define dos reglas de inferencia, la regla F y la regla R:

"Rule **F**: *Foundations*

If P expresses a doundation state contained in the epistemic basis, <P,F,Ø> can be entered as any line of the argument.

Rule **R**: *Closure under reasons*

If $\{P_1,...,P_n\}$ is a reason for Q and $<P_1,\ldots>,\ldots,<P_n,\ldots>$ occur as lines $i_1,\ldots,i_n$ of an argument, $<Q,R,\{i_1,\ldots,i_n\}>$ can be entere don any later line." (Pollock,1987, p.491)

La justificación siempre es relativa a la base epistémica. Se ha de pasar un proceso donde argumentos tratan de derrotarse unos a otros. Los argumentos van ascendiendo del nivel 0 hacia arriba según van superando el proceso sin ser derrotados por argumentos de distintos niveles, es decir, en un principio todos pertenecen al nivel 0, si un argumento sube al nivel 1 significa que ningún argumento de nivel 0 ha podido derrotarlo, si pasa al nivel 2 significa que ningún argumento de nivel 1 ha podido derrotarlo y así sucesivamente. Cuando llegamos a un punto en el que el argumento ha pasado por todo esto y no ha podido ser derrotado se dice que la conclusión está justificada.

Sin embargo, si a través del uso de una regla de inferencia llegamos a un argumento que es una razón prima facie para llegar a una conclusión, si otro argumento llega a negar la conclusión o niega alguna inferencia por la que el otro argumento llega a la conclusión, entonces el argumento queda derrotado. Pollock lo define así:

"(4.1) $<\eta,j>$ *defeats* $<\sigma,i>$ if and only if $<\sigma,i>$ is obtained by rule R using $\{P_{1,...,}P_n\}$ as a prima facie reason for Q, and η_j is either ~Q or $[(P_1\&\ldots\&P_n)\rightarrow Q]$" (Pollock,1987, p.491)

Pero en ocasiones ocurre lo que Pollock llama "derrota colectiva". Sucede que a veces tenemos argumentos para apoyar una conclusión y para negarla. Es decir, siendo Q una proposición en la que creemos y P un conjunto de proposiciones, tenemos argumentos para creer P, pero a su vez, dentro de cada P en P hay un subconjunto finito P_p cuya conjunción con Q proporciona un argumento para ~P. P y ~P tienen el mismo valor y no son derrotadas por ningún otro argumento, solo pueden derrotarse por su interacción. Esto llevaría que no podríamos creer razonablemente en ninguno de los dos, ninguno estaría justificado siguiendo el modo de justificación de argumentos que habíamos establecido, pues, hay argumentos de nivel 0 que respaldan P pero a su vez se pueden combinar para crear nuevos argumentos de nivel 0 que derroten P.

Pollock en (Pollock,1987) hablará de dos paradojas del razonamiento retractable donde tienen lugar derrotas colectivas. La primera paradoja es conocida como "la paradoja de la lotería". Tenemos una lotería que sortea un número n de boletos, donde uno de estos será seleccionado al azar y resultará ganador. Entonces, tenemos en primer lugar la creencia P, la cual dice, hay una lotería justa donde se sortea n boletos y uno de ellos será ganador. Pero ocurre que, las probabilidades de que el boleto 1 salga es mucho menor que la probabilidad de que no salga, y lo mismo se repite con cada uno de los boletos de la lotería. Esto hace que estemos prima facie justificados para pensar que el boleto 1 no saldrá, que el boleto 2 no saldrá y así con cada uno de los boletos. Si existen razones para pensar que los boletos no saldrán, entonces, estamos justificados para pensar que la lotería no es justa. Veámoslo más detenidamente, creemos P porque nos

lo han dicho y lo que nos dicen mayoritariamente suele ser verdad. Entonces, tenemos un argumento σ que apoya P. Por otro lado, tenemos la proposición "se sacará el boleto x" que representaremos con Bx. Si alargamos el argumento σ podemos llegar a un argumento η que lleve a ~P. Es decir, de ($\sim B_1$, $\sim B_2$,…,$\sim B_n$) llegamos al argumento η que dice ~Bn, el cual lleva a ~P, puesto que, si no se saca ningún boleto, no es una lotería justa. Los argumentos σ y η , se derrotan entre sí P y ~P.

La segunda paradoja es llamada "la paradoja de la inducción estadística". Esta consiste en, si observamos un número de elementos n en A, de los cuales se observa que r son B, llegamos por inducción estadística a que probablemente (Bx/Ax) ≈ r/n. La inducción estadística consiste en que si una proporción de elementos en A han sido B podemos llegar a la conclusión de que la probabilidad de que A sea una B es aproximadamente la misma proporción. Suponemos que r/n es alto. Si observamos un conjunto k de elementos que son A, sin saber si son B, podemos inferir por silogismo que cada uno de ellos es B. Aplicando nuevamente la inducción estadística llegamos a que probablemente (Bx/Ax)≈(r+k)/(n+k). Si resulta que k es lo suficientemente grande llegamos a (n+k)/(r+k)≠r/n y podemos inferir que probablemente (Bx/Ax)≠r/n. Podemos ejemplificarlo de la siguiente forma. Tenemos algunas cajas cerradas. Cogemos 7 de ellas y las abrimos. Al abrirlas, vemos que hay 5 cajas que están llenas de naranjas en buen estado y hay 2 cajas de naranjas en mal estado. Por inducción estadística podemos inferir que la probabilidad de que una caja esté llena de naranjas en buen estado es de 5/7 y la probabilidad de que esté llena de naranjas en mal estado es de 2/7. Podemos observar que la probabilidad de encontrarnos con cajas llenas de naranjas en buen estado es bastante mayor que la probabilidad de encontrarnos con cajas de naranjas en mal estado. Por lo tanto, si abrimos otras 7, debido a esta probabilidad estamos prima facie justificados para pensar que las cajas van a estar llenas de naranjas en buen estado. Ahora tenemos 14 cajas de naranjas y suponemos que 12 de ellas van a estar llenas de naranja en buen estado, es decir, la probabilidad ahora es de 12/14. Esto quiere decir, que la probabilidad no es 5/7 como habíamos dicho al principio. Llegamos entonces a una contradicción de la conclusión por lo que el razonamiento ha quedado destruido. Sin embargo, esto se ve intuitivamente erróneo. Se ha llegado a una derrota colectiva.

Pollock dice que esto parece demostrar una insuficiencia en la teoría de la garantía. Comenzábamos con un argumento suponiendo una conclusión P. Extendemos el argumento para llegar a otro argumento η donde se apoya ~P. A simple vista vemos que P debería estar garantizada, pero, dice Pollock que debido a (4.1) llegamos a una derrota colectiva. Pollock cree que esto se debe a que el argumento η ha de ser defectuoso porque, siendo un subargumento dentro de un argumento, ataca el argumento al que pertenece. Como el argumento es auto-derrotable no se ha de dejar que entre en conflicto con otros argumentos. Pollock define esto de la siguiente forma:

"(4.5) σ is self-defeating if and only if σ supports a defeater for one of its own defeasible steps, that is, for some i and j, <σ,i> defeats <σ,j>."(Pollock, 1987, p.495)

Así las paradojas no acabarían en derrota colectiva debido a que los argumentos autoderrotables son defectuosos y no pueden entrar en un conflicto. Se requiere que los argumentos de nivel 0 no sean defectuosos, no pueden atacarse a sí mismos. Un argumento es autoderrotable si apoya a un derrotador dentro de sus propios pasos. Siguiendo el ejemplo de (Koons, 2005) si Robert dice que el elefante a su lado es rosa pero la visión del color de Robert se vuelve poco confiable en presencia de elefantes rosas, normalmente la creencia de que Robert dice que el elefante a su lado es rosa apoyaría a la conclusión de que el elefante es rosa, pero la conclusión es socavada por la creencia de que la visión del color de Robert se vuelve poco confiable en presencia de elefantes rosa socava la conclusión.

Hasta aquí la teoría de la garantía había tomado únicamente argumentos lineales, por lo que, ahora vamos a introducir argumentos indirectos donde se pueden tomar premisas como suposiciones. Esto añade dificultad. Hay que decidir cómo se integraran argumentos subsidiarios en sus argumentos adjuntos. Solemos incorporar al argumento subsidiario las conclusiones respaldadas por el argumento adjunto sin defenderlo puesto que sería volver a repetir el argumento. Pero Pollock dice que hay que poner algunas restricciones a esto.

Supongamos que tenemos un argumento subsidiario que procede de la suposición P, pero, el argumento adjunto es ~P. No podemos incorporar tal cual ~P sin hacerlo autoderrotable. No podemos incorporar al argumento subsidiario ningún derrotador para ningún paso para obtener aquello que respalda el argumento adjunto. Para evitar esto, Pollock dice que es mejor volver a repetir el argumento al añadirse al argumento subsidiario y así evitamos equivocarnos.

Para los argumentos no lineales tendremos que modificar las reglar R y F de la siguiente forma:

"*Rule* **F**: *Foundations*

If P expresses a foundation state contained in the epistemic basis, and Γ is any finite set of propoditions, <Γ,P,**F**,Ø> can be entered as any line of the argument.

Rule **R**: *Clousure under reasons*

If $\{P_1,\ldots,P_n\}$ is a reason for Q and $<\Gamma, P_1,\ldots>$,…, $<\Gamma,P_n,\ldots>$ occur as lines $i_1,\ldots,i_n$ o fan argument, $<\Gamma,Q,\mathbf{R},\{ i_1,\ldots,i_n\}>$ can be entere don any later line." (Pollock, 1987, p.497)

Además, debemos añadir dos reglas para los argumentos indirectos:

"*Rule **P**: Premise introduction*

For any finite set Γ and any P in Γ, <Γ,P,**P**,Ø> can be entered as any line of the argument.

*Rule **C**: Conditionalization*

If <ΓU{P},Q,…> occurs as the line of an argument then <Γ,(PↃQ),**C**, {i}> can be entere don any later line." (Pollock, 1987, p.497)

La condicionalización es una forma de inferencia muy generalizada. Existen condicionales de distintas clases y cada uno de ellos puede inferirse mediante una forma de condicionalización. Además, existe la condicionalización débil, la cual, requiere la ejecución de cada una de las suposiciones a la vez. Pollock va a tomar los condicionales que se relacionan con la socavación de los derrotadores a través de la siguiente regla:

"Rule WC: Weak conditionalization

If <{P}, Q, …> occurs as the line of an argument then <Ø,P→Q, SC, {i}> can be entered on any line." (Pollock, 1987, p.498)

Todas estas reglas de inferencias constituyen solo una pequeña parte de las reglas que son usadas para inferir por los humanos, pero, son las que necesitaremos para continuar con la teoría de la garantía de Pollock.

Esta teoría nos permite demostrar que los conjuntos de proposiciones garantizadas tienen una serie de propiedades que son importantes. Para simbolizar que una proposición está justificada en relación a la base epistémica usaremos "$I=>_E P$". Para que un conjunto Γ $I=>_E P$ tiene que haber un argumento invicto en la base epistémica que contenga <Γ,P,…>. P será una consecuencia deductiva de Γ si en Γ hay un argumento deductivo principal que lleva a la conclusión P, entonces, Γ $I=>_E P$. Pollock hace uso del razonamiento deductivo porque piensa que hay razones concluyentes que se lo permiten. Un conjunto de proposiciones será consistentemente deducible si y solo si no hay una consecuencia deductiva que sea una contradicción. Si se derivara una contradicción del conjunto de proposiciones justificadas, el razonamiento de algunas de

esas proposiciones derrotaría a otras proposiciones del conjunto, por lo que, no estarían justificadas. Si en Γ todo P $I=>_E$P y Q es una consecuencia de P, Q $I=>_E$P.

Supongamos P_1,P_2, P_n como argumentos garantizados y Q como una consecuencia deductiva de estos. Basándonos en esto, podemos construir mediante la combinación de P_1,P_2,P_n un argumento secundario Q, siempre que, añadamos al final del razonamiento un paso donde se cree un argumento donde de P_1,P_2, P_n se deduzca Q. Para que Q no estuviera garantizado debería haber un derrotador para alguno de los pasos del razonamiento. Los últimos pasos no son anulables, por lo que los derrotadores deberían atacar a los argumentos P_1,P_2, P_n, pero, P_1,P_2, P_n están garantizados, lo que quiere decir que , no hay derrotadores, por lo tanto, Q debe estar garantizado.

De aquí se puede deducir según Pollock el teorema de deducción estándar de la lógica clásica que contrasta con la lógica no-monotona de McDermott y Doyle (1980):

"(5.7) If Γ∪{P}$I=>_E$Q then Γ $I=>_E$(P⊃Q)." (Pollock, 1987, p.499)

Ahora que hemos introducido los argumentos indirectos el principio de derrota colectiva sigue funcionando en la teoría de la garantía, pero, este solo se refería a casos en los que los argumentos apoyan tanto a la conclusión P como a ~P.

Existe también el principio de derrota colectiva por socavación. Este consiste en que si estamos justificados para creer R y existe un conjunto Γ donde tenemos argumentos igualmente buenos para creer cada proposición dentro del conjunto, tenemos un argumento de apoyo que implica un paso retractable que procede de algunas premisas S_1,…,S_n a una conclusión Q en cada P de Γ, hay en Γ un subconjunto $Γ_P$ cuya conjunción entre cada uno de sus miembros y R proporciona un argumento deductivo para ~[(S_1,…,S_n)→T] igual de fuerte como el argumento inicial P, y ninguno de los argumentos tiene derrotadores, solo se pueden derrotar al interactuar entre ellos, entonces, ninguna de las proposiciones dentro del conjunto Γ está garantizada en la base de estos argumentos retractables.

Pollock (1987) pone el siguiente ejemplo, tenemos un argumento R que nos dice que por lo general la gente dice la verdad, tenemos la premisa P donde Jones dice que Smith no es confiable y la premisa Q donde Smith dice que Jones no es confiable. La conjunción entre P y R constituye una razón prima facie para creer que Smith no es confiable, lo cual simbolizaremos con S. La conjunción entre Q y R constituye una razón prima facie para creer que Jones no es confiable. Simbolizaremos "Jones no es confiable" con J. Es decir, tenemos por un lado (P&R)→S y por otro lado tenemos (Q&R)→J. Pero, J es un derrotador socavador para S y S un derrotador socavador para J, es decir si tomamos J no podemos confiar en Smith, que es lo que nos dice S, y si tomamos S no podemos confiar en Jones, que es lo que dice J. Por lo tanto, parece que no deberíamos confiar en que ninguno de los dos diga la verdad, tenemos, por lo tanto, una derrota colectiva por socavación.

En ocasiones ocurre que nos encontramos una disyunción garantizada de dos derrotadores, pero los derrotadores individualmente no están garantizados. Supongamos que tenemos una razón prima facie P para creer Q y una razón prima facie R para creer S. Por otra parte, tenemos T que dice ~(P→Q) y V que dice ~(R→S), las cuales no están garantizadas. Si tenemos (T v U) de forma garantizada, existe un principio de derrota conjunta que dice que podemos tomar cualquiera de las dos razones a pesar de que no estén garantizadas. Entonces el argumento queda invicto siguiendo la regla **R**.

En IA han estudiado el razonamiento no monótono centrándose en razonamientos que parten de suposiciones predeterminadas, sin embargo, la monoticidad del razonamiento retractable se debe al uso de razones prima facie, los cuales, pueden verse como

suposiciones, pero, no en el sentido de creencias. Las razones prima facie no tienen que aparecer representadas de forma explícita en el razonamiento, no tengo que pensar que “si x me parece rojo, entonces es rojo”. En IA el papel de esas suposiciones es similar al de las creencias que se mantienen hasta que son derrotadas.

Pollock cree en la existencia de proposiciones justificadas prima facie. Una proposición está prima facie justificada si en relación con toda base epistémica, está justificada en ausencia de una razón que lleve a su negación. Dice que esto se ve de forma clara si tenemos una razón prima facie para Q y formamos el condicional (P→Q) partiendo de la regla **WC,** a menos que tengamos un derrotador que la derrote. Pero en realidad, es irrelevante que haya o no un derrotador para (P→Q), el derrotador niega la razón, pero la formación del condicional está justificada. En un principio Pollock creía que las bases epistémicas incluían proposiciones prima facie justificadas, pero, luego vio que no era así, concluyendo que probablemente las únicas proposiciones prima facie justificadas son condicionales, como el que hemos visto, que se derivan de razones prima facie.

Las proposiciones prima facie justificadas son muy parecidas a la interpretación que hacen de las suposiciones predeterminadas en IA y aunque probablemente solo se encuentren en forma de condicional, pueden ser usadas para llevar a cabo la formalización del lenguaje cotidiano.

Se podría pensar en crear una máquina donde se almacenen condicionales prima facie justificados y se razone de forma deductiva mediante modus ponens partiendo de estos. Pero esto derivaría en el uso de razones en forma de esquemas, los cuales tienen una cantidad de posibilidades enormes, requiriéndose entonces un almacenamiento de condicionales prima facie justificados demasiado grande para ser posible. También cabría pensar en realizar generalizaciones universales para reducirlos, pero, las generalizaciones universales se rompen con encontrar un solo caso en el que no se de la generalización. En el momento que el universal quede anulado por ese caso la generalización no volverá a usarse con el resto de los objetos. Pero para que esto pasara de forma legítima se deberían encontrar muchos casos en los que no se cumpla el esquema. Por lo tanto, no parece buena idea llevar a cabo la implantación de un razonamiento basado en proposiciones prima facie en lugar de un razonamiento retractable.

Ocurre que del mismo modo que veíamos en Gettier donde se demostraba que podíamos tener creencias justificadas que no son verdaderas, también podemos tener creencias no garantizadas, al menos no de la forma que hemos estado hablando hasta ahora. La forma de razonamiento que hemos estado tratando hasta ahora está bien de cara a un ideal con memoria infinita y una capacidad de procesamiento muy superior a la que tenemos. En la realidad una persona no puede comparar cada argumento contenido en su conocimiento para poder creer algo, ni entrar en un ciclo infinito de argumentos derrotados y reincorporados al llegar a una derrota colectiva.

Debemos razonar sobre aquello que nos es más inmediato, aquello que nos es más accesible en la situación epistémica. Se dice que el uso de esas reglas internas que tenemos para razonar es lo que hace que nuestras creencias estén justificadas, pero puede darse el caso de que esto no se cumpla. El razonamiento aspira a llevar a cabo la justificación de la forma ideal que hemos visto, pero, está limitado. Nuestras reglas para formar creencias justificadas presuponen que hemos hecho un buen razonamiento.

Solemos suponer que las reglas de formación de creencias nos llevan a formar nuevas creencias si ya tenemos unas creencias que están bien relacionadas. Pero no tenemos por qué crear una nueva creencia cada vez que ocurre esto. Tenemos la posibilidad, pero, no la obligación. Podemos tener una creencia, pero no tenemos que creer todo lo que se infiere de ellas. Normalmente llevamos a cabo este tipo de inferencias cuando es de nuestro interés. Es por lo que Pollock dice que este modus ponens debería llamarse algo así como “regla de permiso”. Pero, aunque las reglas no nos obliguen,

tampoco deben dejarnos hacer lo que queramos. Ha de haber una regla que nos diga en qué circunstancias debemos realizar esas inferencias y construir las nuevas creencias. Además, debería mostrarnos cuales de las creencias que tenemos son de nuestro interés para llevar a cabo la nueva inferencia.

Entonces dice Pollock que debe haber unas tres reglas. Una regla que nos diga que debemos adoptar nuevas creencias si tenemos argumentos de las que inferirlas y tenemos interés en saber de ellas. Otra regla que cancela que lleguemos a estas nuevas creencias si existen argumentos que derroten a los que teníamos. Y, por último, una regla que los reincorpore si se encuentran argumentos que derroten a los que actúan como derrotadores de los argumentos que apoyaban a las nuevas creencias.

Para formalizar estas reglas Pollock hace una serie de propuestas. Para la primera propone dos opciones. La primera opción es simple. Si tenemos unas proposiciones P_1, P_2, P_n que sean razones para creer Q, te interesa Q y no hay argumentos que derroten esas razones, debes creer Q. La segunda opción incluye el hecho de que creamos de forma justificada, por lo que se dirá que, si crees de forma justificada en P_1, P_2, P_n las cuales son razones para creer en Q, no hay ningún argumento que derrote estas razones y Q te interesa entonces deberías creer Q. Pero esta segunda opción al introducir la justificación ya llevaría a indagar en qué justifica P_1, P_2, P_n y así sucesivamente llevando a una regresión infinita. Esto se puede evitar si las reglas son locales, es decir, no es necesario formar primero otras creencias porque respondemos de forma directa.

Los humanos no rastrean cada uno de los argumentos que le llevan a creer en algo, solo asume que las creencias están justificadas hasta que algún argumento diga lo contrario. Un sistema artificial que esté programado para razonar debe funcionar de forma similar, porque si no llegaría al mismo problema que hemos visto con anterioridad en distintas circunstancias, la búsqueda sería inmensa debido a que su memoria se abarrotaría. Es por esto que Pollock cree que la regla que utilizamos para formar creencias debe ser como la primera opción.

Para la segunda regla que se refiere a la derrota del argumento debido a al descubrimiento de derrotadores para este, primero sugiere una opción que será rechazada debido a que vuele a incurrir en la necesidad del seguimiento de las creencias y de los argumentos que las sostienen. Eso como vimos al analizar la segunda opción de la primera regla, no es práctico, no hacemos eso. Esta opción propuesta por Pollock dice que si creemos Q en base a una línea de argumento que ha sido obtenida por la regla **R** partiendo de (P_1,P_2,…,P_n) como razón prima facie para Q, si adoptas una razón R que lleva a ~Q o a ~[(P_1,P_2,…,P_n)→Q], entonces, debemos dejar de creer en Q. Otra opción consiste en que, si creemos que creemos Q sobre la base de un argumento, en el cual hay una línea que se obtiene del uso de una razón prima facie P para creer en Q, y creemos en un derrotador, entonces, debemos dejar de creer en Q. Pero esto implica que la derrota tenga un proceso de orden superior en los razonamientos y los humanos solemos hacer la derrota de forma automática. Si tengo una razón R que derrota a P automáticamente dejo de creer en Q.

Esto es uno de los problemas que trata de solucionar Pollock. En un principio dio por imposible crear reglas de derrota que funcionaran de forma automática, pero en (Pollock,1987) cuando trata el problema de construir un marco general del razonamiento retractable creando una teoría de argumento retractable basada en algunos supuestos intenta crear reglas de este tipo.

En el siguiente apartado vamos a pasar a ver cómo funciona el marco general de razonamiento retractable creado por Pollock basándonos en (Pollock,1987) y (Pollock,1995). Recibe el nombre de OSCAR y según Pollock es una representación del

funcionamiento del razonamiento retractable que, además, podría ser usado para crear un sistema inteligente que razonara de esta forma. Pollock va variando OSCAR hasta llegar a la creación de "asignación de estado parcial" con la que logra solventar la mayoría de los problemas con los que se encuentra.

2.2.2 OSCAR

Pollock crea un sistema llamado OSCAR donde primero toma unos supuestos simplificadores y luego se eliminarán esas simplificaciones. Primero volverá a tomar argumentos lineales únicamente, se interesará por todo para poder sacar el máximo de combinaciones posibles, aunque, habrá que tener cuidado con los esquemas de razón que se les dará para que no haya un exceso y se sature, y todos los argumentos van a tener el mismo valor, por lo que al llegar a argumentos contradictorios se llegará a una derrota colectiva.

Siguiendo estos supuestos va a pasar un sistema de reglas que deberá ser modificado posteriormente. Se debe crear un sistema de razonamiento que no acabe derivando al infinito llegando a una combinación combinatoria. Un razonamiento ha de llevarse a cabo en un periodo de tiempo no muy elevado, por lo que han de elevarse búsquedas enormes en el razonamiento. Algunos sistemas en IA como el de Jon Doyle (1979) lleva al sistema a poder seguir los argumentos, almacenarlos y tener acceso a ellos para poder derrotarlo y reincorporarlos. Entonces, la memoria se llena demasiado, puesto que, tiene que guardar las creencias y los argumentos que llevan a esta, que además suelen ser cada vez más largos. Esto implicaría hacer una búsqueda inmensamente imposible. Esto, como vimos en la introducción, es lo que tratan de hacer los sistemas expertos de tipo explicativo. Estos sistemas cuando llegan a unas conclusiones han de poder explicarte los pasos del razonamiento que han dado y mostrarte los datos que han usado. Si para el razonamiento tuviera que llevarse a cabo una cantidad inmensa de pasos y los datos fueran tan elevados esto no se podría llevar a cabo. Para que esto no ocurra, se pueden almacenar las creencias recién obtenidas de las creencias que ya se contienen, así se partiría directamente de las nuevas creencias para llevar a cabo el nuevo razonamiento que se deba realizar. Los humanos no retoman cada una de las creencias que contienen, toman las más actualizadas. Por lo tanto, OSCAR tendría, por un lado, un conjunto donde estarían almacenadas todas las creencias que se han adquirido, y por otro, un conjunto que contenga las creencias más recientes que se han adoptado. Cuando se lleva a cabo un nuevo razonamiento, las creencias que se encontraban en el conjunto donde se almacenan las últimas creencias adoptadas pasan al conjunto donde se contienen todas las creencias y nuevas creencias llenan de nuevo el conjunto de las creencias recién adoptadas.

Pero el problema de no almacenar los argumentos radica en cómo llevar a cabo la derrota. Puede ser una solución el almacenamiento del último paso del argumento. Pero también tenemos que realizar un seguimiento de si la razón es concluyente o retractable haciendo que tengamos que hacer que la derrota trabaje de cierta manera. Si S cree Q sobre X quiere decir que cree Q sobre la base de la razón retractable X. Por otro lado, Si S cree Q con X quiere decir que cree Q sobre la razón concluyente X. También hay que seguir las bases que llevan a la derrota de las razones. Si X es un conjunto de proposiciones que constituye una razón prima facie para creer en Q, y ^X es el conjunto de todos los miembros (P_1,P_2,…,P_n) de X. Al ser derrotada por la subvaloración, S cree ahora ~(^X→Q), quedando la base almacenada en inicio, el conjunto inicial de premisas. La reincorporación se llevará a cabo cuando S encuentre algo para retractarse de ~(^X→Q).

Rebatir la derrota es más complicado. Si tenemos una razón prima facie {P} para creer Q y una razón prima facie {R} para creer S, al llegar a una derrota colectiva no es tan fácil como negar ambas conclusiones, ya que ha de ser posible que se reincorpore una

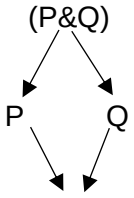
de los dos si una es derrotada. Es decir, si se encuentra un derrotador para {R} como razón para S, entonces debiera de poder creer Q sobre la base de la razón prima facie {P}. Pero solo buscamos creencias más recientes. Puesto que el argumento de que {P} es una razón para creer Q ya no pertenece a el conjunto en el que se encuentran las creencias más recientes, no podemos inferirlo directamente. Por lo tanto, debemos almacenar las proposiciones de la derrota colectiva y el conjunto de los últimos pasos de los argumentos. En este caso {<Q, {P}>, <S,{R}>}. Luego hay que retomar los últimos pasos del razonamiento que son desechables y refutarlos. Pollock define que en un caso donde las razones son todas conjuntos unitarios, q es un antecedente no derrotable de una creencia p si y solo si hay una secuencia de creencias $< p_1,\ldots,p_n>$ $(i>1)$ donde $p=p_n$ y $q=p_1$, para la cual $i<n,< p_{i+1}, \{p_i\}> \epsilon_{conset}$. Para que un par <r,{s}> sea el ancestro no derrotable más inmediato de p, r tiene que ser un ancestro no derrotable de p y <r{s}> ϵconset.

Los razonamientos lineales pueden expresarse mediante estructuras secuenciales o sobre grafos de inferencia según Pollock (1995). La estructura secuencial expresa el razonamiento de forma secuencial codificando las secuencias de inferencia. Tomemos el ejemplo que pone Pollock en este mismo libro:

1. (P&Q)
2. P por eliminación de la conjunción de 1
3. Q por eliminación de la conjunción de 1
4. Q&P por introducción de la conjunción de 2 y 3

1. (P&Q)
2. Q por eliminación de la conjunción de 1
3. P por eliminación de la conjunción de 1
4. Q&P por introducción de la conjunción de 2 y 3

Mediante Grafos de inferencias se elabora una representación gráfica que codifica las relaciones de dependencia. Pollock lo muestra así:



(Pollock, 1995, p.87)

Los razonamientos suposicionales pueden codificarse en grafos de inferencia como el que hemos observado en el ejemplo. Cada nodo en el grafo de inferencia se representa con <X,p> siendo X una suposición y p una proposición que es una conclusión. El grafo avanza mediante reglas de inferencia que son descritas por Pollock (1995):

Para añadir premisas, si tenemos un grafo G y una premisa p, para una suposición X podemos añadir un nuevo nodo <X,p> a G creando así un nuevo grafo de inferencia. Si tenemos en G un conjunto X de suposiciones donde p está contenido podemos formar un nuevo nodo <X,p< en G creando un nuevo grafo de inferencia. Si G tiene un nodo <X U {p}, q> entonces podemos añadir un nuevo nodo formado por <X (p Ↄq)> con un ancestro inmediato <X U {p}, q>. Esto último es una condicionalización. Para hacer un razonamiento por caso en un grafo G, este tiene que tener varios nodos <X (pvq)>, <X

U {p}, r> y <X U {q}, r>entonces se añade un nodo <X,r> con los nodos anteriores como ancestros inmediatos.

En los casos más generales donde los conjuntos de razones no son unitarios esto es más complicado que cuando si lo son. Las razones vienen en esquemas como <X,p> donde X es un conjunto de formas de proposición y p una forma de proposición. Una forma proposicional es, por ejemplo, “x me parece rojo.”, “x es rojo.”, “Aquello que veo en el campo parece una cabra.”, “En el campo hay una cabra”. Una forma de proposición p se divide en dos funciones, $todo_p$ y $partes_p$. $Todo_p$ genera la proposición p y $partes_p$ genera una asignación de constituyentes proposicionales a las variables de p. Una asignación es un conjunto de pares <x,a> donde x es una variable metalingüística que ocurre al formularse formas proposicionales y a es el objeto que la asignación pone a la variable.

Como recurso Pollock dice que si <p, X> ϵ_{conset} entonces X es un conjunto de ancestros no derrotables de p. Si X es un conjunto de ancestros no derrotables de p, q ϵ X, y <q, Y> $\epsilon_{conset.,}$ entonces (X-{q}) U Y es un conjunto de ancestros no derrotables de q. Si q es alguno de los miembros del conjunto de ancestros no derrotables de p entonces es un ancestro no derrotable de p. Si q es un ancestro no derrotable de un miembro de X significa que es también un ancestro no derrotable de ese conjunto X. Para que X sea un ancestro inmediato derrotable de una proposición p, X tiene que ser un conjunto de pares ordenados, el dominio de X ha de ser un conjunto de ancestros no derrotables de p y $X \subseteq_{onset}$. Para que X sea un ancestro inmediato de $(p_1,\ldots,p_n)$ tiene que haber $X_1,\ldots,X_n$ tal que para cada i, X_i sea un ancestro derrotable inmediato para p y X sea igual a X_1 U … U X_n. Esto quiere decir que el ancestro derrotable más inmediato de p es el conjunto de nodos inferiores de un argumento que conduce a p a lo alto. Cuando dos cadenas de razonamiento llevan a conclusiones contradictorias se toman los ancestros no derrotables y estos intentan rebatirse uno al otro, cuando esto pasa, se introducen los dos en el conjunto “refutación”.

Cada vez que se adoptan nuevas creencias, OSCAR repasa las creencias que ya tenía. Se usan las reglas del razonamiento retractable llevando a nuevas creencias y a retractar viejas creencias. Para controlar el surgimiento y la retractación de creencias se utilizan bandera de adopción y bandera de retractación, la cuales, en un inicio estarán a 0 pero cada vez que se adopte o se retracte una creencia se pasara a 1. Si al finalizar el proceso está a 0, se borran los conjuntos de retractación y de adopción para poder seguir procesando nuevas entradas. Si la bandera de adopción está en 1 estas creencias se insertan en el conjunto donde están todas las creencias y en el conjunto de nuevas creencias adoptadas. Si la bandera de retractación está en 1 se eliminan las creencias incluidas en esta del conjunto donde se encuentran todas las creencias, si está en el conjunto de nuevas creencias adoptadas también se elimina y se incluye en el conjunto de creencias retractadas, aunque no se eliminan de conset, ya que, como dice Morado (2004) las creencias cuando son derrotadas no se eliminan, no se tiran a la basura, sino que quedan bloqueadas o desconectadas. Que se haya dejado de tener razones para creer en algo no significa que tengamos que olvidarlo, la no creencia en algo también es conocimiento y de igual forma que se ha derrotado la creencia puede derrotarse el argumento derrotador de esta, pero para poder reincorporarla debemos de tenerla almacenada. De hecho, si no guardáramos las creencias retractadas, cada vez que se nos apareciera tendríamos que volver a hacer todo el proceso de razonamiento hasta derrotarla, cuando si la tenemos almacenada ya no se da pie a eso, al menos que se estén adoptando nuevas creencias.

En los razonamientos suposicionales hay que tener cuidado con la derrota, no basta con tener un nodo <Γ,p> y un nodo que contenga un derrotador para <Γ, p>, solo se derrota si la suposición X está contenida en la suposición Y del otro nodo. En el caso del rebatimiento un nodo tiene que surgir de una razón prima facie basado en la suposición X, y otro nodo sustenta ~q basándose en la suposición Y donde $X \subseteq Y$. En

el socavamiento tiene que haber un nodo que surja de una razón prima facie, sustente q en una suposición Y donde p_1,…, p_n son proposiciones que se sustentan en sus ancestros inmediatos, y otro nodo que sustenta que (p_1 & … & p_n) no llevan a q y ambas tienen la misma fuerza. Un nodo puede ser derrotado por surgir de la inferencia de un nodo derrotado o por un nodo no derrotado. Siguiendo a (Pollock, 1995), primero propuso para asignar el estado de las derrotas en computación que se siguiera lo siguiente:

Para que un nodo no esté derrotado, ni él, ni ninguno de sus ancestros han de estar derrotados por ningún nodo del grafo de inferencia. El nodo no está derrotado si sus ancestros inmediatos no lo están y todos los nodos que lo derrotan están derrotados. El nodo está derrotado si un ancestro inmediato es derrotado o hay un nodo en el grafo de inferencia que no está derrotado y derrota al nodo en concreto.

Pero esto no funciona con los casos de derrota colectiva o con argumentos auto-derrotables. Por ejemplo, el caso donde Jones dice que Smith no es confiable y Smith dice que Jones no es confiable. Si tomamos P como razón prima facie para Q y Q como razón concluyente para ~(R→S) y tomamos R como una razón prima facie para S y siendo S una razón concluyente para ~(P-→Q), entonces Oscar entra en un bucle infinito al recibir la entrada {P,R}. Es por eso que realiza una modificación:

Los nodos D-iniciales no están derrotados, los nodos autoderrotables están derrotados totalmente, un nodo que no es un nodo totalmente autoderrotado, si sus ancestros inmediatos no están derrotados y los nodos que lo derrotan están derrotados no se encuentra derrotado. Si existe un ancestro inmediato totalmente derrotado u otro nodo no derrotado que lo derrota el nodo es totalmente derrotado. En otro caso el nodo será provisionalmente derrotado.

Un nodo provisionalmente derrotado puede llegar a derrotar a sus derrotadores y los argumentos que se autoderrotan son excluidos por ser “defectuosos”. Pero esta noción también llevó a problemas, por lo qué, creó la noción de los estados de asignación parcial:

Cuando hacemos una asignación σ de los estados derrotado y no derrotado a un subconjunto de nodos de un grafo de inferencia G estamos haciendo una asignación de estados parcial si y solo si σ asigna “no derrotado” a todos los nodos D-iniciales, σ asigna “no derrotado” a un nodo α tal que a sus derrotadores se les ha asignado “derrotado” y a sus ancestros inmediatos “no derrotado”, σ asigna “derrotado” a un nodo α que tiene un ancestro inmediato al que se le ha asignado “derrotado” o hay un nodo β al que se le asigna “no derrotado” que derrota al nodo α. Para ser un estado de asignación parcial el nodo no puede estar propiamente contenido en ningún otro estado de asignación parcial.

La asignación de estados parcial logra solventar los problemas con los argumentos autoderrotables y la derrota colectiva.

Ahora queda definir cuando un argumento está justificado en OSCAR. Por un lado, nos encontramos la noción de justificación epistémica y por otro de garantización ideal. La garantización ideal es la noción que tomó Pollock al principio del apartado, la cual, se basa en un conjunto infinito de argumentos. Aquí el razonador tendría la capacidad de crear y analizar infinitos argumentos. La justificación epistémica comprende unos límites y se computa en un tiempo limitado. Nos muestra si una proposición está justificada en el momento o no.

La noción de justificación epistémica dice que un par S= <X,σ> formado por una razón y una conclusión está justificado en grado δ en el estadio i si y solo si existe un nodo α

con una fuerza igual o mayor a ese grado que pertenezca a G_i, no esté derrotado con respecto G_i y además sustente S.

La noción de garantización ideal dice que un par S=<X,σ> formado por una razón y una conclusión está idealmente garantizado en un grado δ relativo a un conjunto P de premisas si y solo si existe un nodo α con una fuerza mayor o igual a la de ese grado que pertenece a un grafo de inferencia G, no está derrotado con respecto a éste y además sustenta S.

OSCAR en un principio no es totalmente realista ya que se basa en suposiciones que han sido simplificadas y además tiene algunos errores relacionados con estas. Por ejemplo, si A es una razón prima facie para creer R y R es una razón concluyente para creer Q, B es una razón prima facie para creer P y P es una razón concluyente para creer Q y C es una razón prima para S , la cual, es una razón concluyente para ~Q. Si tomamos una entrada {A,C} la refutación de la derrota pasa porque las creencias son solo {A,C} y la refutación es [{<R,A>,<S,C>}]. Si agregamos la entrada B, buscamos añadir B a las creencias y añadir para refutar {<P,B>, <S,C>}. B se agrega a las creencias pero en lugar de añadir {<P,B>, <S,C>} se añade <P,B> y se añade Q con P. Esto se debe a que en la refutación no se menciona Q, pero, los ancestros inmediatos de Q son los que han hecho la refutación. Pollock propone una manera ad hoc de manejar este problema mientras no se agreguen condicionales a OSCAR. Si S es una razón concluyente para ~Q, quiere decir que S implica ~Q. De aquí se sigue que si se da Q implica ~S. Como no puede usar un condicional se añadirá que siempre que una proposición D sea una razón concluyente para H, ~H será también una razón concluyente para ~D. Además aparecían como hemos visto al encontrarse con una derrota colectiva o con un argumento autoderrotable. Pero con la noción de asignación de estado parcial se lograron resolver estos problemas.

Pollock muestra cómo se crean argumentos, cómo funciona la derrota y cuándo un argumento está justificado epistémicamente y garantizado idealmente. Esto no es fácil. Hemos visto cómo iban surgiendo problemas, por lo que, fue modificando algunas definiciones hasta alcanzar la noción clave de asignación parcial de un estado. Hemos visto como para el razonamiento retractable la noción de derrota y su asignación. Entonces el razonamiento retractable surge por la existencia de razones prima facie que constituyen razones para creer algo hasta que aparezca un derrotador para ella. Mientras tanto estamos justificados para creer en ellas. Para que el razonamiento sea posible es necesario conocer los argumentos que sustentan a las creencias, pero como tenemos una capacidad limitada nos quedamos con los llamados ancestros inmediatos. En los sistemas artificiales, aunque las capacidades son distintas se limita también el almacenamiento de argumentos puesto que la búsqueda podría derivar al infinito. Pollock crea OSCAR representando en un gráfico la estructura de un estado cognitivo ideal. Crea grafos de inferencia donde mediante el uso de reglas de combinación se van integrando nuevos argumentos, nuevos nodos. Se crea una red donde cada nodo constituye una razón prima facie para el segundo.

OSCAR presenta algunos problemas como por ejemplo los mencionados en (Bringsjord; Govindarajulu, 2018). Uno de los problemas mencionados es la falta de expresividad de OSCAR. OSCAR no maneja operadores intencionales, parece que Pollock iba a tratar el tema de la intencionalidad, pero posiblemente no le dio tiempo llegando en 2009 el culmen de su vida. El segundo problema es que es cierto que OSCAR es capar de tratar acertijos y paradojas de un nivel sencillo, pero no es capar de manejarlos a un nivel más complejo.

En el siguiente apartado vamos a ver la crítica realizada por parte de Wolfgang Sponh en (Sponh, 2014) a la teoría del razonamiento retractable de Pollock llevando a cabo una comparación entre ésta y la teoría de la clasificación del propio Sponh. También hablaremos brevemente de la crítica que se puede llevar a cabo desde el externalismo al internismo de Pollock donde las creencias han de justificarse en otras creencias.

CRÍTICA A LA TEORÍA DE RAZONAMIENTO RETRACTABLE DE POLLOCK

3.1 John Spohn: Teoría de la clasificación vs Teoría del razonamiento retractable

En (Sponh, 2014) se lleva a cabo una crítica a la teoría del razonamiento retractable de Pollock a través de su comparación con la teoría de la clasificación del propio Sponh. Sponh critica la falta de carácter normativo de la teoría computacional de Pollock.

En primer lugar, vamos a introducirnos brevemente en la teoría de la clasificación de John Sponh para poder pasar a la comparación de ambas teorías. La teoría de la clasificación trata el problema de cómo cambiar las creencias después de la llegada de nueva información. Spohn critica a Pollock por tomar como base las percepciones, para él la recepción no debe limitarse a ésta. Como vimos en el apartado (2.1) Pollock reconoce que se obtiene conocimiento a través de distintas formas, no solo de la percepción, pero todas estas acaban dependiendo de la percepción, es por eso y porque es la forma en la que manejamos más información, que Pollock toma la percepción como la base de la que parte el razonamiento. Spohn acepta que la información en un principio se recibe a través de la percepción, pero defiende que sería beneficioso para tratar el cambio de las creencias tomar un límite de entradas relativo a un dominio restringido. Reconoce que en última instancia van a proceder de la percepción, pero, defiende que sería algo pesado tener que ver siempre todo el proceso desde la percepción hasta el límite relativo.

La teoría de la clasificación trata de caracterizar los estados doxásticos mirando hacia dos objetivos. El primero es que han de contener creencias simples que sean verdaderas o falsas. El segundo es que se puedan establecer leyes dinámicas generales y completas para estas creencias.

De cara al primer objetivo, debemos de tomar los contenidos de las creencias que puedan ser verdaderos o falsos, por lo tanto, queda eliminada la teoría de la probabilidad, ya que, una probabilidad de x para una proposición no puede ser ni verdadero ni falso. Sponh pone de ejemplo “la paradoja de la lotería” para demostrar que la teoría de la probabilidad no es buena para abordar estos temas ya que esta se discute cándidamente, sin encontrar solución.

Este primer objetivo parece conseguirlo, por ejemplo, la lógica doxástica. Pero lo realmente complicado es conseguir tratar el segundo objetivo. Muchas teorías fallan a la hora de intentar establecer unas leyes generales para la creación de creencias, entre ellas la lógica doxástica. Sponh (1988) resuelve el problema con la teoría de las funciones de clasificación. Ésta trabaja con rangos de incredulidad. Cuando el rango es 0 la proposición es verdadera o indefinida, ni verdadera ni falsa. Conforme el valor va aumentando mayor grado de incredulidad, por lo que, menos ha de creerse. Para que una proposición deba ser creída, el grado de incredulidad de su negación ha de ser >0.

Esto conlleva la ley de la negación y la ley de la disyunción. La ley de la negación consiste en que una proposición o su negación debe recibir rango 0. La ley de la disyunción dice que si tienes A o B deben tener el mismo rango que A y B. Aunque Sponh (2014) dice que lo realmente importante es la definición de rangos condicionales. El rango de B dado A es el rango de A y B menos el rango de A. La noción de independencia doxástica diría que A y B son independientes si y solo si el rango de B no se ve afectado al condicionarlo a; A o ~A.

Con estas reglas que funcionan como la condicionalización por la que se trasladan las probabilidades condicionadas a la información recibida en la teoría probabilística se resuelve el problema de la creación de reglas dinámicas generales para las creencias, logrando cumplir también con el segundo objetivo.

Después de este breve vistazo a la teoría de la clasificación vamos a pasar a ver las diferencias que establece Sponh desde el punto de vista computacional siguiendo lo dicho en (Spohn, 2014).

Spohn dice que ambas teorías compiten. Tomando la noción de garantía ideal de Pollock, Pollock establece una especie de subida escalonada donde un argumento va ascendiendo desde el nivel 0 al no ser derrotado hasta no poder ser derrotado por nada. Pero la teoría de la creación de creencias da por supuesta esta garantía idealizada y trata de decir cómo se comportan tales creencias.

Esto es lo que critica Sponh. La teoría del razonamiento retractable de Pollock es una teoría computacional, es un modelo de computación humana que pretende mostrar cómo ha de ser la racionalidad, siendo, por lo tanto, puramente normativa. Sin embargo, la teoría de la clasificación de Spohn es regulativa. Trata la estructura dinámica de la garantía ideal sin importar la accesibilidad computacional. Pollock (1995) ve necesario para corregir una teoría de la racionalidad que se aplique como base para un agente racional autónomo debiéndose construir un sistema de IA que implemente la teoría. Defiende esto porque dice que hay muchas teorías que no son implementables.

Spohn piensa que las teorías computacionales no tienen estándares normativos a los que apelar, es por eso que cree que la teoría computacional de Pollock no puede aportar nada sobre la estructura de la garantía ideal. Los problemas normativos residen en Pollock en las reglas de inferencia y las reglas de combinación para integrar argumentos en los grafos de inferencia que vimos al hablar de OSCAR. Debido a pluralidad de formas de inducción existente le parece difícil sistematizarlo en reglas específicas. Spohn critica la forma de las reglas de Pollock, pero, Pollock defiende que las reglas han de ser procedimentales, el razonamiento ha de actuar siguiendo estas reglas. Sin embargo, Spohn piensa que lo verdaderamente importante es encontrar la regla general para pensar. Como dice Spohn (2014), lo que importa no es cómo debe montarse en bicicleta, sino, cómo crear una bicicleta. Spohn reclama más información sobre el comportamiento, un carácter normativo adecuado.

Pollock construye una teoría de argumentos, inferencias y razones, pero Spohn ni siquiera usa estos en su teoría para la revisión de creencias. Parece que las nociones de argumento o de inferencia no pueden substraerse del marco computacional, pero la noción de razón si es vista por Spohn de una forma no computacional. Este dice que A es una razón para B si y solo si A fortalece la creencia en B. El grado de creencia en B es más alto dado A que dado ~A. Las razones no solo sirven para impulsar las inferencias presentes hacia creencias recientes, sino, para impulsar el cambio doxástico. Esto quiere mostrar como realmente no es que la revisión comience donde acaba el razonamiento retractable, sino que, más bien pertenecen al mismo proceso.

Hay razones concluyentes que nunca pueden ser derrotadas, son las únicas razones invencibles. Otras razones son derrotables por otros argumentos y pueden ser derrotadas por la negación de la conclusión o por el ataque a la conexión entre las premisas y la conclusión. Si A es positivamente relevante para B y C es negativamente relevante para B y por lo tanto, una razón para ~B, entonces, C es un derrotador que refuta B. Si A es positivamente relevante para B y C hace que A ya no sea positivamente relevante para B entonces C es un derrotador de socavamiento. Esto dice Spohn (2014) que parece mostrar que realmente él y Pollock están tratando el mismo tema. Pero Spohn critica que el tiene una teoría y unas explicaciones para guiar las afirmaciones acerca de las razones, mientras que Pollock solo tiene un apoyo intuitivo de sus afirmaciones.

Spohn defiende la necesidad de tratar estos temas desde una perspectiva normativa. No ha de tomar la forma de reglas de inferencias, sino que ha de enunciarse en los términos de razón que él presenta. Considera las reglas de inferencia de Pollock como restricciones a priori de los estados doxásticos. Ve deficiencias normativas en la teoría del razonamiento retractable de Pollock, piensa que una teoría computacional aporta poco realmente al campo y propone como alternativa su teoría regulativa de la clasificación.

3.2 Objeciones del exterismo

La concepción externista de Armstrong según (Grimaltos; Iranzo, 2009) del conocimiento afirma que lo que convierte una creencia en conocimiento es la relación entre el sujeto y el mundo. El internismo dice que solo a lo que el sujeto puede tener acceso cognitivo o tiene puede actuar como justificador. El hecho de que el cielo esté nublado no puede justificar por si mismo mi creencia de que va llover, solo mi creencia o percepción de que está nublado, en tanto que estados mentales a los que tengo acceso por introspección o reflexión pueden hacerlo. Como vimos al principio de este trabajo, según Pollock, tenemos unas creencias epistemológicamente básicas que constituyen la base del conocimiento y a cada intuición se le asocia una creencia. Se genera entonces una teoría piramidal del conocimiento donde las creencias se van justificando unas con otras y van ascendiendo de nivel, convirtiéndose estas en razones para justificar a otra creencia. El problema de la idea de que una creencia está justificada si la justifica otra creencia lleva a que solo se justifica una creencia si somos conscientes de la existencia de una creencia que la justifica y para que una creencia justifique a otra esta debe tener a su vez razones y dice el externalismo que en algún momento sería necesario que existiera alguna evidencia que no fuera una creencia. Cuando tengo una razón prima facie para creer que veo x rojo porque hay una luz roja alumbrando x no porque sea rojo, justifico esta creencia en que hay una luz roja alumbrando x y las luces rojas suelen hacer que objetos se vean rojos aunque no lo sean, pero en última instancia tengo que apelar a la luz roja en sí. Para el externista la justificación vendría de que veo un objeto y hay una luz roja que lo alumbra. La justificación vendría dada de la relación del sujeto y el mundo a través de la percepción. Además, para que la verdad no se aleje de la justificación, aunque ambos acepten que hay creencias que están justificadas que no son verdaderas, es necesario remitirse al mundo, a algo externo a nosotros. Podemos tomar el ejemplo que pone Arsmtrong (1963) donde digo que tengo dolor y más tarde me pregunto por qué nadie prestó atención a mi grito y muecas de dolor. La persona tenía la creencia de que tenía dolor y además creía que había gritado y que estaba haciendo gestos de dolor, pero, acudiendo al mundo, resulta que no había gritado, sino que había hablado bajito y que además no tenía ningún gesto de dolor. Al decírselo las personas a su alrededor que lo habían observado corrige la creencia, es decir, personas externas que habían atendido a lo externos le mostraron su error y evaluando lo que otros han percibido se corrige la creencia, se ve que es errónea y se corrige. Esto se debe a que no tenemos consciencia de todo, como dijimos con anterioridad al hablar de la negación de la teoría de la incorregibilidad en Armstrong (1963), una cosa son los informes sobre algo y otra cosa es ese algo.

CONCLUSIONES

El enfoque sobre el razonamiento retractable de Pollock consiste en enumerar un conjunto de reglas que son computables con el objetivo de describir la forma en que un agente cognitivo ideal construye un conjunto de creencias partiendo de una base de datos muy básica. El razonamiento es algo no estático que va cambiando según recibimos nuevas entradas, nueva información. Al no tener un conocimiento completo y cerrado la información que vamos adquiriendo hace que tengamos que llevar a cabo una revisión de aquello que creemos y revisemos si con la nueva información seguimos creyendo lo mismo, lo modificamos o lo desechamos aparte de añadir nuevas creencias. Define las nociones de razones prima facie y derrotadores, los cuales, son los principales causantes del razonamiento retractable. Aquello que tengo en este momento como razón para una creencia puede dejar de serlo al ser derrotado, es por eso que se crea la asignación de estado parcial que asigna a un argumento el estado de "derrotado" o "no derrotado" pudiéndose revisar y cambiar su estado. Para que una creencia esté garantizada en relación con una base epistémica debe estar apoyada por un argumento que haya resultado invicto, es decir, que no haya podido ser derrotado por ningún otro argumento, y que pertenezca a la base epistémica. Pollock elabora una teoría de la garantía y luego crea un marco representativo donde se representa la estructura de forma gráfica mediante grafos de inferencia. Mediante reglas de combinación se irán añadiendo nodos al grafo de inferencia. Pollock tuvo que enfrentarse a diferentes problemas, por lo que, tuvo que ir modificando su teoría para poder solventarlos. Pollock, como se dice en (Bringsjord; Govindarajulu, 2018) parecía intentar crear un intelecto artificial donde se comprobará su teoría, ya que, para este, como dijimos con anterioridad, es importante que las teorías puedan aplicarse a un agente racional autónomo para comprobarla.

Esto que hace Pollock muestra, como se dice en (Bringsjord; Govindarajulu, 2018), como la IA puede nutrirse de la filosofía y a través del uso de las técnicas y herramientas de estas llegar a tratar temas encontrando algunas soluciones que pueden ser comprobadas a posteriori al intentar implementarlas en un programa computacional. Aunque la IA no sea filosofía puede ayudarse de ésta, al fin y al cabo, para la IA es crucial el estudio del razonamiento y la observación del comportamiento racional, y que mejor que ayudarse entonces de ramas como la Filosofía de la Mente, la Epistemología, la Filosofía del Lenguaje o la Lógica. De igual forma, la Filosofía puede razonar sobre temas de IA como, por ejemplo, la IA "Fuerte" o la IA "Débil". La IA "Débil" busca crear máquinas que procesen información que parezcan tener el repertorio mental de los humanos. La IA "Fuerte" trata de crear personas artificiales que realmente tengan el repertorio mental de los humanos. John Searle, por ejemplo, trató de derrocar la postura de la IA mediante el experimento mental de la habitación china perteneciente a (Searle,1980). Pollock escribió "*A Blueprint for How to Build a Person"* (1995) pero OSCAR no tiene las capacidades de los seres humanos. OSCAR es capaz de llevar a cabo un razonamiento retractable bastante sofisticado, aunque tenga algunos problemas como comentamos anteriormente. El trabajo de Pollock fue esencial en IA para el tratamiento del razonamiento no monotónico y aunque OSCAR quedó estancado tras el fallecimiento del filósofo parece según (Bringsjord; Govindarajulu, 2018) que Kevin O'Neill ha recuperado OSCAR y se están llevando a cabo experimentos con este.

BIBLIOGRAFÍA

Armstrong, D. M. (1963). Is Introspective Knowledge Incorrigible? *The Philosophical Review*, Vol.*72*. No. 4, pp. 417–432.

Bringsjord, S y Govindarajulu, N. S. (2018) "Artificial Intelligence" *The Stanford Encyclopedia of Philosophy (Winter 2019 Edition)*, Edward N. Zalta (ed.),

https://plato.stanford.edu/archives/win2019/entries/artificial-intelligence/

Carnota, R. J. (1995,2005,2013) *Lógica. Enciclopedia Iberoamericana de Filosofía.* Ed. Trotta. CSIC. Vol. 7 pp. 143-181.

Delgado, M. (1996) *La Inteligencia Artificial. Realidad de un Mito Moderno.* Discurso de apertura Universidad de Granada curso académico 1996-1997.

DeVries, W. (2011), "Wilfrid Sellars", *The Stanford Encyclopedia of Philosophy (Winter 2019 Edition),* Edward N. Zalta (ed.),

https://plato.stanford.edu/archives/win2019/entries/sellars/

Gettier, E. L. (1963) Is Justified True Belief Knowledge? Analysis, Vol. 23, No. 6, pp. 121-123 https://www.jstor.org/stable/3326922

Grimaltos, T. y Iranzo, V. (2009) "Externismo/Internismo en la justificación epistémica" *Cuestiones de Teoría del Conocimiento.* Quesada D.(ed.) Madrid, Tecnos, pp. 33-76

Koons, R. (2005) "Defeasible Reasoning" *The Stanford Encyclopedya of Philosophy*, Edward N. Zalta (ed.), *https://plato.stanford.edu/archives/sum2022/entries/reasoning-defeasible/*

Mc Dermott, D. y Doyle, J. (1980) *Nonmonotonic Logic I, Artificial Intelligence* Vol. 13, pp. 41–72.

Doyle, J. (1979) "A truth maintenance system" *Artificial Intelligence.* Vol.12, No. 3, pp. 231-272, *https://doi.org/10.1016/0004-3702(79)90008-0*

Morado, R. (2004) *Problemas filosóficos de la Lógica no monotónica. Enciclopedia Iberoamericana de Filosofía.* Ed. Trotta. CSIC. Vol.27, pp. 313-344.

Pollock, J. L. (1975) *Knowledge ad Justification*. Princeton University Press. Princeton, New Jersey.

Pollock, J. L. (1976) *Subjuntive Reasoning*. Philosophical studies series in philosophy Vol. 8

Pollock, J. L. (1987) "Defeasible Reasoning" *Cognitive Science* Vol. 11, pp. 481-518.

Pollock, J. L. (1995*) Cognitive Carpentry: A Blueprint for How to Build a Person*, MIT Press

RussellL, S. J. y Norving, P. (2004) *Inteligencia Artificial, Un enfoque Moderno.* Pearson Educación, S.A., Madrid (2004)

Searle, J. (1980) The Begavioral and Brain Sciences. Vol 3 (1980) pp. 417-424.

Spohn, W. (2002) *A Brief Comparison Of Pollock's Defeasible Reasoning And Ranking Functions.* First publ. In : Synthese 131 (2002), pp. 39-56 ,

http://dx.doi.org/10.1023/A:1015004212541

Turing, *A.* (1950) *Computing Machines and Intelligence. Mind a quaterly review, Vol. 59. 236, pp. 433-460.*