

LA RACIONALIDAD CIENTIFICA COMO UN PROBLEMA DE LA EVOLUCIÓN DE NORMAS CULTURALES

Sergio F. Martínez

Instituto Investigaciones Filosóficas, UNAM

1. Introducción. El planteamiento del problema de la racionalidad. En la introducción al libro Ian Hacking 1983 nos dice que la famosa “crisis de la racionalidad” que marcó a la filosofía de la ciencia de la segunda mitad del siglo XX fue una consecuencia de retomar una visión histórica de la ciencia desarrollada en el XIX y que había sido totalmente dejada de lado en la primera mitad del XX. La visión estática de la ciencia es ejemplificada por Hacking con las ideas de Carnap y Popper, los dos principales filósofos de la ciencia de buena parte del siglo XX. A pesar de sus diferencias, ambos filósofos consideraban que la física era el paradigma de qué debía de considerarse conocimiento y racionalidad. Ellos tenían diferencias respecto a cómo elaborar esta idea. Popper pensaba que la mejor manera era hacer ver cómo la física, y la ciencia en general, eran el resultado de seguir un cierto método (el método de conjeturas y refutaciones). Carnap pensaba que la mejor manera de hacer ver cómo se desplegaba la racionalidad humana en la ciencia era tratando de entender la estructura lógica del lenguaje científico. Compartían, sin embargo, el supuesto de que había una diferencia epistemológica fundamental entre dos tipos de actividades científicas, aquellas que tienen que ver con el llamado *contexto de descubrimiento*, y aquellas que tienen que ver con el llamado *contexto de justificación*. El contexto de descubrimiento se refiere a todos aquellos procesos que permiten explicar cómo se llegó de hecho a un descubrimiento, por ejemplo los condicionantes sociológicos, psicológicos o económicos que juegan un papel en la promoción o inhibición de una línea de investigación o de una propuesta teórica. El contexto de justificación se refiere a los procesos relacionados con la evaluación epistémica de los productos finales, las teorías o hipótesis. El problema de decidir si una hipótesis ha sido satisfactoriamente corroborada por pruebas, o en qué grado está apoyada por un cuerpo de evidencia son típicas preguntas propias del contexto de justificación. Para Carnap y Popper, como para la gran mayoría de filósofos de la ciencia de ese tiempo, la filosofía de la ciencia estudiaba el contexto de justificación de la ciencia, y ese contexto podía articularse de manera independiente de la forma como históricamente la teoría científica en cuestión había llegado a configurarse.

La irrupción del historicismo en la filosofía de la ciencia dejó claro que la generación, evaluación y aceptación de teorías no parecía poder dividirse claramente en algo que sucedía en un contexto de descubrimiento, por un lado, y en un contexto de justificación, por el otro. La justificación que los científicos daban de sus teorías muchas veces parecía arrastrar temas y maneras de ver las cosas que eran fácilmente identificables como parte de su situación histórica o personal particular. Los historiadores hicieron ver también que las decisiones de los científicos respecto a qué teoría seguir no parecían estar guiadas por un método definido ni por una “lógica de la confirmación”. Todo esto generó una crisis.

Si se asumen los supuestos sobre los que se asienta esta “crisis de la racionalidad” el diagnóstico y tratamiento posterior de la crisis es predecible: un cuestionamiento de la pertinencia epistemológica de la distinción tajante entre descubrimiento y justificación lleva de manera natural a un cuestionamiento de la racionalidad misma de la ciencia. La racionalidad de la ciencia siempre se ha asociado a la manera como, a diferencia de otros tipos de conocimiento, la labor científica se sustenta sobre evidencia y razonamiento que no depende de cuestiones subjetivas. En la medida en que se cuestiona la diferencia entre ambos contextos, se cuestiona la objetividad de las categorías epistémicas (un concepto de evidencia objetiva y de razonamiento basado en evidencia en particular) sobre las que se ha apoyado la supuesta racionalidad de la ciencia.

Así, siguiendo este cuestionamiento, en los estudios empíricos de la ciencia se ha tendido a negar que haya algo distintivo racional que distinga a la ciencia de otro tipo de creencias. Muchos filósofos de la ciencia han respondido aceptando que no hay un tipo de racionalidad distintivo de la ciencia, la racionalidad científica no es más que una racionalidad instrumental que busca acoplar medios con fines. Lo distintivo de la ciencia, se piensa, reside en el tipo de fines que persigue, fines que se considera son epistémicos. Los fines epistémicos son lo que le da unidad y coherencia racional a la empresa científica¹. Los diferentes filósofos difieren respecto a cuáles son esos fines epistémicos. Hay quienes piensan que el fin es la resolución de problemas; otros, que es la búsqueda de la verdad o el conocimiento de la estructura causal del mundo, pero hoy en día entre los filósofos de la ciencia es muy común pensar que es en términos de fines cómo debe entenderse la unidad de la ciencia, y su estructura epistemológica. Ronald Giere, Larry Laudan y Philip Kitcher son algunos de los autores contemporáneos que siguen este tipo de planteamiento. Esta concepción instrumentalista de la racionalidad de agentes individuales generalmente se implementa como un modelo de la racionalidad científica a través del supuesto de que existen mecanismos del tipo de la “mano invisible” que explican cómo esta búsqueda de acoplamiento entre medios y fines individuales lleva a la coordinación espontánea de la acción social y, por lo tanto, a un modelo de cómo la ciencia se estructura socialmente.²

Por otro lado, están los filósofos historicistas, como Kuhn, Feyerabend y Hacking, que piensan que la racionalidad tiene ya poco de interesante, y que en la medida que no podemos pensar la ciencia como relacionada por fines últimos tampoco podemos identificar una clase de buenas razones privilegiada que serviría de base para distinguir entre teorías. Para alguien como Feyerabend o Hacking, hay muchos estilos de articular lo que se consideran buenas razones, y esto es todo. No hay una noción normativa de racionalidad que sirva de guía para articular una pretendida unidad de la ciencia y, por lo menos implícitamente, se rechaza la idea que un mecanismo de “mano invisible” permita explicar mucho de cómo se llegan a identificar y a evaluar lo que se consideran buenas y malas razones.

Ahora bien, si se asume que la ciencia es esencialmente un conjunto de teorías que se aceptan o rechazan a partir de un análisis de la evidencia disponible, entonces el problema lleva a una de las dos posibles respuestas esbozadas arriba. O la racionalidad se

¹ La mayoría de estos filósofos piensa que no es posible identificar un conjunto de fines de la ciencia que sean válidos a través de toda la historia de la ciencia, pero piensan que en una cierta época es posible identificarlos con relativa facilidad.

² En (Martínez 2003), examino y critico esta concepción de la racionalidad más a fondo.

asocia con una estructura de medios y fines, y se privilegian ciertos fines, o bien se llega a sostener algún tipo de relativismo que considera que la racionalidad de la ciencia en última instancia sólo puede caracterizarse como una descripción de maneras de proceder históricamente identificables, pero no epistémicamente privilegiadas. Pero este dilema es un artefacto de supuestos implícitos en el planteamiento del problema. Se presupone que la racionalidad de la ciencia tiene que darse en la relación entre evidencia y teorías. Sin embargo, si como ya Neurath lo recalca, la ciencia como una empresa racional no puede entenderse haciendo abstracción de todo ese complejo de actividades humanas que se estructuran en prácticas científicas y en instituciones de diverso tipo, entonces el problema de la racionalidad de la ciencia puede replantearse de maneras que evaden la dicotomía anteriormente mencionada.

2. Hacia un replanteamiento del problema de la racionalidad. Aquí quiero explorar la siguiente propuesta: Un planteamiento apropiado del problema de la racionalidad científica requiere entender en general la relación de conformación mutua que existe, por un lado, entre la realidad histórica, moral y social (lo que incluiría dimensiones antropológicas, psicológicas, etc.) de los agentes humanos, articuladas en normas e instituciones, y por el otro, la generación, transformación, estabilización y abandono de normas (estándares, valores, etc.) de diferente tipo, a lo largo de la historia de la cultura. *El objetivo central del trabajo es sugerir una manera de replantear el problema de la racionalidad partiendo de propuestas respecto a modelos de la cognición que evitan el tipo de supuestos que nos hace caer en la dicotomía tradicional entre modelos historicistas e instrumentalistas.* Es claro que esta formulación del problema tiene sentido si partimos del supuesto de que una explicación filosóficamente interesante de la ciencia no puede reducirse a una mera explicación de la estructura y dinámica de teorías, ni a una explicación de las normas de aceptación de teorías dada cierta evidencia. Es decir, estamos partiendo de la idea de que el problema de la racionalidad en la ciencia tiene que responderse en el marco de una explicación general de cómo diferentes tipos de actividades y normas se articulan en prácticas e instituciones. Esto por supuesto va a requerir una cierta caracterización de qué entendemos por "práctica" y por "instituciones". A continuación explico brevemente qué entiendo por ambos términos.

Una práctica es un complejo de actividades (y, por lo tanto, de normas, reglas, valores estándares y tecnología) de diferente tipo que tienen una estructura estable y relativamente autónoma, con capacidad de reproducir esa estructura con variantes a través de diferentes tipos de procesos y en diferentes instituciones, pero que son identificables como ejemplares de la misma práctica por períodos relativamente largos de tiempo.³ Una práctica científica es una práctica que involucra normas y estándares epistémicos, tecnológicos y éticos que son propios de la cultura científica.

Muchas veces se habla como si las normas fueran siempre reducibles o formulables en términos de valores. Pero prefiero entender el concepto de norma en un sentido restringido (y, por lo tanto, mantener la distinción entre normas y valores). Por ejemplo, hay normas relacionadas con convenciones sociales o con tradiciones de investigación (como los modales en la mesa o técnicas de observación en un telescopio) que no tienen por qué verse como normas provenientes de valores sociales. Los valores

³ Por supuesto que algo más tiene que decirse respecto al tipo de autonomía en cuestión, pero para nuestros fines esto no es necesario.

sociales pueden generar o promover normas, pero también la ciencia y la tecnología pueden hacerlo. Esta distinción puede parecer artificial, pero nos permite formular la idea que las normas de una práctica responden a criterios de corrección que, por lo menos en parte, provienen de otra práctica, y no directamente de valores sociales. Esto nos permite decir que la normatividad pertinente para entender y utilizar criterios de corrección específicos está *implícita y distribuida*, en toda una geografía o una arquitectura de prácticas que se norman mutuamente en diferentes aspectos establecidos a través de la historia de las prácticas humanas. En esta geografía normativa juegan un papel importante las propiedades materiales de objetos y aparatos que son parte constitutiva de una práctica pero que ciertamente no queremos identificar con valores. Las propiedades de conducción eléctrica o el peso específico muchas veces juegan un papel como parte constitutiva de la estructura normativa, sin estrictamente ser normas o regulaciones de los agentes, el tipo de normas o regulaciones que podríamos asociar con valores en el sentido tradicional. Si uno va a reparar una instalación eléctrica debe de seguir ciertas normas de seguridad, por ejemplo. Estas normas son diferentes dependiendo del voltaje de la instalación y dependiendo de muchos otros factores materiales (por ejemplo el nivel de humedad usual en el ambiente).

El concepto de institución como es usado muchas veces en economía y otras ciencias sociales se traslapa en buena medida con lo que llamo una práctica; se usa muchas veces para referirse a aquellas convenciones, rutinas o procedimientos que articulan la vida social, y por otro lado, se refieren a las organizaciones a las que los seres humanos pertenecen y en las que llevan a cabo su actividad de agentes. El primer sentido de institución se refiere a lo que estoy llamando prácticas, cuando hablo de instituciones voy a hablar más bien en el segundo sentido. Así, podemos decir que las prácticas organizan las fuentes de autoridad (las normas, valores, etc.) en términos de, o en vista de las instituciones a las cuales pertenecemos como agentes.

3. **Cognición y representación.** Ahora bien, un aspecto muy importante de cualquier explicación de la actividad científica en términos de prácticas es el tipo de modelo de la cognición humana que implícita o explícitamente entra en la explicación. Hasta hace muy poco era usual entender la cognición como haciendo referencia a la manera como se adquiere y procesa la información, y se pensaba que, en particular, no debería referirse a conocimiento procedimental, el tipo de conocimiento que nos permite llevar a cabo tareas o actividades como andar en bicicleta o llevar a cabo una medición experimental. Esta manera estándar de entender la cognición está ligada a dos supuestos. En primer lugar, al supuesto de que es posible hacer una distinción clara entre el contenido representacional del conocimiento, por un lado, y su articulación en tecnología e instituciones, por el otro, (una distinción que, por supuesto, está relacionada con la distinción entre contexto de descubrimiento y contexto de justificación). En segundo lugar, con el supuesto, asociado al proyecto clásico de la inteligencia artificial, y en buena parte de la psicología cognitiva hasta la fecha, de que la cognición puede caracterizarse en términos del modelo CDI (Creencia-Deseo-Intención). Según este modelo, la cognición consiste en un conjunto de fines (deseos) y un conjunto de planes o “recetas para la acción” que residen en una memoria o librería de planes. El mecanismo (o agente) selecciona un plan de acuerdo a una evaluación que hace de su potencial para satisfacer un determinado fin, y luego lo

activa para llevarlo a cabo. En otras palabras, la cognición tiene lugar en “estados internos” que cambian de acuerdo a los resultados de las acciones.

Para poder llegar al tipo de conclusiones que me interesan respecto a las implicaciones de modelos de la cognición para discusiones sobre el concepto de práctica y su papel en discusiones sobre la racionalidad en la ciencia es necesario establecer una distinción entre dos conceptos de *representación*. Por un lado, caracterizo un tipo de representación asociado con los modelos tradicionales de la cognición que llamaremos *correspondentista*, este es el tipo de concepto de representación asociado con el proyecto clásico de Inteligencia Artificial y que es también ampliamente utilizado (o presupuesto la mayoría de las veces) en filosofía de la mente y la filosofía de la ciencia centrada en teorías. Por otro lado, vamos a caracterizar un tipo de representación que llamaremos *interaccionista*. Este segundo tipo de representación busca identificar los rasgos cruciales del tipo de representación que juega un papel en proyectos en varias áreas de las ciencias cognitivas que buscan caracterizar a la cognición como “corporalizada” o “anclada”⁴. Lo que nos interesa es capturar la distinción que consideramos crucial en ambos tipos de proyecto.

Por un lado, la representación se entiende como una correspondencia entre un conjunto de insumos y un conjunto de resultados y, por lo tanto, en última instancia, se compromete con la idea de que toda representación codifica información sobre un estado cognitivo que se transforma en un estado cognitivo posterior a través de mecanismos cognitivos internos que supuestamente monitorean nuestra interacción con el mundo. Por otro lado, en la concepción interaccionista, muchas representaciones tienen que entenderse como emergentes, a partir de la acción y la interacción de agentes. Para ejemplificar estas ideas, adelante presentamos el modelo interaccionista de R. Brooks. Es claro que este segundo concepto de representación ha sido defendido como parte de la tradición pragmatista en filosofía (James y Baldwin, en particular). Pero en este trabajo no trataremos de explorar esas conexiones. A continuación resumo los rasgos distintivos de ambos conceptos de representación.

Una manera usual de formular la idea de que las representaciones son un conjunto definido de correspondencias es la llamada "semántica informacional". La idea básica es que una cierta mancha en nuestro campo visual representa una mosca, por ejemplo, en la medida en que la mancha lleva a nosotros información acerca de la mosca. Por supuesto, tienen que formularse restricciones al tipo de información que representa algo. Podemos

⁴ Por ejemplo, Wheeler 2005, Keijzer 2001, Hendriks-Jansen 1996. Hendriks-Jansen habla de “emergencia interaccional” como un aspecto crucial del tipo de modelo de la cognición situada que él explora en su libro. Este tipo de emergencia es un aspecto crucial del tipo de representación que estamos llamando interaccionista. En este trabajo vamos a asumir que hay emergencia de este tipo que juega un papel importante en la caracterización de los procesos cognitivos. No vamos a entrar en una discusión sobre los detalles de este tipo de emergencia. Para esta discusión puede consultarse el libro de Hendriks-Jansen. Este segundo tipo de representación juega también un papel importante en el llamado “nuevo conductismo”, véase por ejemplo (Staddon 2001). En este trabajo me voy a centrar en la discusión del concepto de representación implícito en el trabajo de R. Brooks en robótica, y en particular en la propuesta que desarrolla en los ensayos recopilados en su libro (Brooks 1999). Como lo menciono adelante, y como muchos críticos han hecho ver, Brooks dice muchas cosas sobre representación que no parecen ser consistentes, y que, en particular, no son consistentes con la propuesta de lo que aquí caracterizo como “representación interactiva”. Pero creo que la mejor manera de entender su propuesta es considerar que se centra en el reconocimiento de un concepto de representación interactivo que en particular no requiere de un procesador central (véase la discusión adelante).

restringir los conjuntos de información que son representaciones a aquellos conjuntos que codifican correlaciones causales, o casos en los que la información proviene de una ley de la naturaleza (casos en los que se describe una relación nomológica)⁵. Hay discusiones sobre el tipo de restricciones que deben implementarse, pero lo que es claro es que tienen que ponerse restricciones, no todo conjunto de información (inputs o datos) puede considerarse una representación; la idea detrás es que hay una cierta normatividad de la representación que debe establecerse para determinar las restricciones de lo que puede ser o no una representación. Cada instancia de una ley causal podría ser parte de una representación, pero, por lo general, no lo es. Sólo ciertas instancias se consideran dignas de tomarse como representaciones, cuáles sí se consideran como tales depende de criterios respecto de lo que se considera conocimiento en disciplinas o teorías específicas. En general, es un serio problema para la concepción correspondentista de las representaciones el determinar el tipo de restricciones que distinguen representaciones de conjuntos de información sobre un sistema.

Hay varias y buenas razones para pensar que esta visión tradicional de la cognición es problemática y debe ser abandonada⁶. Un problema de fondo se conoce en la inteligencia artificial como el problema del “anclaje de los símbolos” (symbol grounding problem). El problema es cómo anclar el significado de usos particulares de símbolos (tokens) que no tienen significado en otra cosa que no sean otros símbolos que no tienen significado (Harnad 1990). Hay muchas maneras de tratar de resolver este problema, pero parece ser que cualquier solución requiere de alguna forma abandonar la idea de que la cognición puede caracterizarse en términos de símbolos que tienen significado por su anclaje en otros símbolos sin significado. Por lo que de diferentes maneras se ha sugerido que el problema puede resolverse sólo si se reconoce que el uso de símbolos tiene que verse como emergiendo de representaciones implícitas en interacciones o expectativas de interacción, que van teniendo significado a través de un proceso de aprendizaje. Así, este problema desaparece en la medida en que dejamos de lado la idea de que la cognición es mero procesamiento de símbolos.

Otro problema (o más bien, tipo de problemas) muy famoso se le conoce como el *problema del marco* y sus variantes. El problema surge de la dificultad de entender cómo es posible que podamos aprender y manejar cognitivamente la variedad de conceptos acerca de nuestro entorno que de hecho manejamos. Sabemos que los árboles no caminan, que las hierbas no vuelan, que los peces no viven en la tierra, sabemos que un libro no cambia el número de oraciones que contiene de un día para otro (a menos que le arranquemos una página) o que un camión es mas pesado que una bicicleta, etc., etc. El problema es, ¿cómo podemos aprender tal cantidad de cosas y usar los conceptos relacionados de una manera que requiere identificar el tipo de contexto al que se refiere el “usualmente” en los usos correctos de los diferentes conceptos? Si las representaciones que utilizamos para razonar son codificaciones de información, no parece haber manera de dar cuenta de cómo podemos aprender tal variedad de cosas y sobre todo cómo podemos relacionarlas de manera racional. En este último caso, parece que tenemos que apelar a criterios *extra-racionales*, “intuiciones” o “conocimiento innato” para explicar

⁵ Ver por ejemplo, Fodor 1990 y Birkhard & Terveen 1995.

⁶ Para una discusión detallada de estas razones desde perspectivas diferentes ver Hendriks-Jansen 1996 y Birkhard & Terveen 1995.

esa capacidad, o bien, como Kant lo explicaría, en términos de que tales representaciones son condiciones *trascendentales* de posibilidad de la experiencia y del conocimiento humano.

El problema del marco fue descubierto y formulado por primera vez con cierta claridad en relación al problema de construir un robot que tuviera la capacidad de identificar y modificar su conducta a partir de las consecuencias de sus acciones. Es fácil programar un robot para que identifique la consecuencia de que si empuja un vaso sobre una mesa, éste va a caerse, pero no es para nada obvio cómo programarlo para que el robot infiera que el vaso no cambia de tamaño, o que no cambia de color. O que el vaso sigue siendo el mismo si lo cambia de cuarto, etc.⁷ El problema puede reformularse como *el problema de identificar los factores relevantes requeridos para poder evaluar las consecuencias de una acción*. Cuáles son los factores relevantes depende de manera crucial de contextos que es muy difícil clasificar e identificar a partir del tipo de correlaciones que permite la concepción correspondentista de la cognición. En general, el tamaño o la forma del vaso no cambian, pero si el vaso es de un material especial, o si la temperatura es muy diferente de la usual, entonces ciertamente que la forma puede cambiar. Así, es claro que una versión del problema del marco es el problema de cómo identificamos los marcos de creencias relevantes para entender las consecuencias de nuestras acciones y, por lo tanto, poder planear nuestra relación con el entorno. En la medida en que esta identificación de los contextos relevantes parece requerir de un proceso de manipulación de representaciones internas muy complejo, tan complejo que fácilmente se vuelve computacionalmente intratable, la idea de que las representaciones pueden entenderse como codificaciones de información que caracterizan correspondencias deja de ser satisfactoria.

Ahora bien, diferentes propuestas para resolver el problema del marco requieren tomarse en serio la idea de que, de una u otra manera, muchas representaciones son implícitas. Una representación implícita de alguna forma apunta o involucra aspectos del entorno que no están explícitos como información (codificada en un estado del sistema, por ejemplo). Es decir, una representación implícita de alguna forma apunta o describe posibilidades que no pueden derivarse de información codificada en la representación. Una representación teatral implícitamente refiere a muchas cosas que no están previstas en la obra de teatro. Una obra de Shakespeare no incluía una serie de interpretaciones que podemos hacer hoy día sobre la condición humana en el siglo XXI. En ese sentido, podemos decir que la obra no representa explícitamente una situación que tiene lugar hoy en día, pero implícitamente sí.

Podemos decir que un sistema de axiomas implícitamente caracteriza una teoría. Pero este no es el sentido en el que queremos usar el término. Un sistema de axiomas potencialmente es una representación explícita de todas sus consecuencias lógicas, pero éstas están bien definidas y determinadas independientemente de lo que suceda, y en ese

⁷ Tal vez convendría enmarcar esta discusión recordando la respuesta de Kant a este problema. Kant afirmaba que la unidad del objeto se establece sobre la base de la unidad del sujeto, i.e., en la medida en la que el sujeto de conocimiento conforma, a priori, una unidad sintética de la apercepción es posible la unidad y uniformidad de la experiencia. Una pregunta respecto del ejemplo anterior es ¿en qué medida podría hablarse de una unidad análoga a la que Kant se refiere en un robot como condición necesaria para que éste pudiera conformar la unidad y uniformidad de su experiencia? A grandes rasgos la respuesta es “emergencia interactiva” y teoría de la evolución (ver Wimsatt y Griesemer en prensa y Hendriks-Jansen 1996, capítulo 10 y siguientes). Algunas ideas en esta dirección las desarrollo adelante.

sentido no es una representación implícita en el sentido que nos interesa aquí desarrollar. El sentido de implícito que nos interesa desarrollar tiene que ver con el tipo de incertidumbre que está esencialmente asociada con procesos en donde aspectos contingentes juegan un papel crucial en la determinación de las posibilidades futuras de desarrollo.

Una representación es implícita si es una indicación de una posible interacción I con consecuencias definidas D en un contexto C. Es decir, una representación implícita podemos verla como un diagrama de lo que es posible dados ciertos supuestos respecto a lo que puede suceder (que tienen por lo general el tipo de sistematicidad asociado con reglas, normas o leyes); nos dice, por ejemplo, que si llevamos a cabo la acción A en el contexto C_i entonces podemos prever ciertas consecuencias definidas D_i . O que si en un contexto C_j no tiene lugar esa consecuencia, entonces podemos concluir que no es del tipo D_i . *La representación es implícita en la medida en que este tipo de representación no requiere que haya una representación explícita de cuáles son las propiedades relevantes que hacen que el contexto C_i sea o no de un cierto tipo. Esto es parte de lo que se toma como situación por defecto (default).*

A diferencia del caso de un sistema axiomático, las propiedades que distinguen un tipo de contexto de otro no están implícitas en la estructura lógica de sus consecuencias. Están implícitos en el diagrama que nos dice que en un contexto determinado tiene lugar una determinada consecuencia dado un cierto tipo de acción (en condiciones default). Las condiciones se presuponen en la expectativa racional de lo que es la situación usual, normal o por *default*. Este tipo de expectativa racional es una manera de darle cuerpo (o si se prefiere, sistematizar) el papel de lo contingente en posibles interacciones futuras. Si la temperatura del medio ambiente se mantiene dentro de cierto rango, como lo prevé una determinada expectativa, entonces la búsqueda de conejos por un zorro tiene ciertas consecuencias previstas, pero si la temperatura es mucho más alta o baja que lo normal, ya no son esperables esas consecuencias (en este caso capturar una presa) sino otras ligadas al nuevo tipo de contexto (la muerte por ejemplo). Pero tanto las condiciones como las expectativas o mapas de posibles consecuencias y su evaluación respecto a futuras acciones sólo están dadas implícitamente en un ambiente regulado por lo que es normalmente el caso. En particular, y de manera muy compleja en contextos sociales, por normas asociadas con prácticas o actividades que estructuran a su vez los contextos que se distinguen por medio de acciones diferentes. En otras palabras, las normas están implícitas en la estructura de lo que se considera posible.

4. Modelos alternativos de la cognición y sus implicaciones. El conexionismo es quizás el principal modelo alternativo al modelo tradicional basado en la idea que la cognición consiste en el procesamiento interno de símbolos. Según este modelo, las representaciones que constituyen conocimiento pueden verse como distribuidas en toda una red: el contenido representacional del conocimiento no puede separarse de su implementación en la red en donde está distribuido. Por lo que es difícil adoptar la distinción tradicional entre operaciones internas y externas a un sistema, así como la distinción entre contenido representacional y la manera en la que está implementado en la red. No obstante, por lo menos en sus versiones más usuales, el conexionismo sigue basándose en representaciones correspondentistas. Esas representaciones se entienden

como distribuidas en redes, pero siguen estando caracterizadas por información acerca de parámetros que constituyen la representación. Las representaciones conexionistas tienen varias de las propiedades que nos interesa resaltar en modelos no estándar de la cognición, pero parece ser que las representaciones conexionistas no pueden dejar de verse como codificaciones de información, si bien distribuidas en toda una red, y, por lo tanto, serían correspondentistas. Por razones que espero queden claras más adelante, es crucial que las representaciones no sean correspondentistas para que tengan la capacidad de evolucionar sin confrontar alguna versión del problema del marco.

Rodney Brooks inició hace ya más de dos décadas un proyecto que él caracterizó durante mucho tiempo como basado en un rechazo al papel de las representaciones en la cognición. Uno de sus artículos más famosos se llama “Intelligence without representation” (incluido en Brooks 1999). Varias críticas a las ideas de Brooks han estado dirigidas a mostrar que es simplemente absurdo pensar que la cognición puede entenderse sin el concepto de representación.⁸ Sin embargo, como él mismo reconoce más tarde (en el prefacio a Brooks 1995) el punto no es que la cognición puede entenderse sin representaciones, sino que puede entenderse sin representaciones correspondentistas⁹.

Es más, me parece que una de las contribuciones más importantes de Brooks es hacer ver *que una representación no correspondentista no requiere de que haya un sistema central en el cual se procesen las representaciones*. A esta conclusión llega Brooks a partir de la convicción de que la única manera en la que sería posible llegar a desarrollar realmente inteligencia artificial sería por medio de la construcción de robots que pudieran llegar a ser viables en determinados ambientes, *sin tratar de prever los tipos de cosas que tendrían que identificarse y el tipo de funciones que tendrían que ser programadas en el robot*. La diferencia crucial con programas tradicionales de inteligencia artificial es que en lugar de buscar una descomposición de un problema complejo en funciones, de lo que se trataría es que el robot, *a través de un proceso de aprendizaje*, llegara a descomponer un problema en “actividades”.¹⁰ Para Brooks una actividad es un patrón de interacciones con el mundo, por ejemplo, “explorar lugares interesantes, o “evitar zonas con muchos obstáculos”. Las acciones para Brooks no tienen que preprogramarse como subrutinas de un programa central, y en ese caso *las decisiones no se tomarían sobre la base de un sistema interno de representaciones del mundo*, o sea que los rumbos de acción ya estaban previamente determinados para cada

⁸ Ver la introducción a su colección Brooks 1999.

⁹ Brooks no habla de tipos de representación y dice cosas que por lo menos aparentan ser contradictorias respecto a qué concepto de representación tiene en mente. Lo que hago a continuación es una reconstrucción de su proyecto de una manera que me parece fiel a su objetivo central y que deja claro cuales son sus preocupaciones centrales.

¹⁰ Cabe hacerse la siguiente pregunta, ¿no se necesitaría de un procesador central o una unidad sintética en algún sentido para llevar a cabo esta descomposición de un problema en "actividades"? ¿no lo es, incluso, para cualquier proceso de aprendizaje? Kant respondería, sin duda, afirmativamente, una condición para la descomposición o análisis es la síntesis previa, el enlace de lo múltiple en una unidad. Nosotros, en cambio, respondemos que lo que hace falta no es una unidad sintética, sino una noción de "agencia" que emerge de las interacciones con el entorno. La integración de experiencias que constituye agencia es una cuestión de grados, el resultado de un proceso evolutivo. Las representaciones del mundo se van integrando como representaciones de agentes particulares conforme las posibles consecuencias se van delineando implícitamente y asociando con otras acciones a futuro.

contexto y el robot sólo tenía que aplicar la instrucción correspondiente de acuerdo con la situación en la que se encontrara; en cambio en el modelo de Brooks cabe la posibilidad de que el robot desarrolle nuevas rutas de acción sobre la base de lo aprehendido. En el fondo, se trata de la idea de que las representaciones del mundo en el modelo correspondentista eran fijas, y en el modelo de Brooks tales representaciones son dinámicas o interactivas. En la arquitectura de la cognición propuesta por Brooks las acciones resultan del comportamiento que a su vez se produce como resultado directo de la transformación de insumos. Esta transformación de insumos es una representación implícita (en el sentido anteriormente especificado) de las posibles relaciones del agente con el medio ambiente, i.e., que haya una adaptación de cierto tipo de respuesta ante una situación que plantee un abanico de posibilidades. Ello implica que la base de la adaptación de ciertas representaciones o de la decisión del curso a seguir resida en el robot y no prefigurada en una especie de manual de instrucciones finito. Ahora bien, lo interesante de estas ideas es que no es tan importante que el robot aplique correctamente una acción en contexto específico, sino que pueda enfrentarse a otros contextos muy diversos con base en las aplicaciones pasadas. Pero ello implica, a nuestro modo de ver, no sólo que el robot pueda aprender de las aplicaciones exitosas, sino más importante aún, que pueda aprender de las que no lo fueron, pues tanto unas como las otras conforman el criterio de aplicación de las representaciones, entendidas en un sentido dinámico e interactivo. Las actividades que llegan a establecerse como viables en un determinado ambiente se superimponen en lo que Brooks llama una *arquitectura de subsunción*. El comportamiento observado es el resultado de diferentes actividades autónomas que se superimponen de acuerdo a jerarquías localmente aprendidas y dependientes del despliegue de actividades en un determinado ambiente.

El núcleo de la idea de Brooks no es que podemos dar cuenta de la cognición sin representaciones, ni siquiera que las representaciones en cuestión no son representaciones internas. Más bien, lo más distintivo y prometedor de su propuesta tiene que ver con la idea que no hay explicación de la cognición que no incluya una caracterización de su evolución como un proceso no teleológico. Incluso en el caso de los robots, la cognición tiene que empezar a construirse desde abajo, desde lo más simple, *sin control central*.

5. Implicaciones de modelos de cognición no correspondentistas para la discusión sobre racionalidad. La lección crucial de la arquitectura de la cognición propuesta por Brooks es que la manera tradicional de entender el objetivo de la Inteligencia Artificial clásica, basada en la tesis de que cognición es procesamiento de representaciones que codifican información está equivocada porque se requiere presuponer la existencia de algún tipo de agencia central que procese la información que constituye la correspondencia. Brooks trata de mostrar por qué este supuesto tiene que ser abandonado. Sus argumentos no están basados en consideraciones metafísicas, sino en lo que él considera un hecho: que una arquitectura cognitiva que incluya este tipo de sistema no es un punto de partida viable para construir agentes autónomos. Pero si debemos abandonar esta idea en un modelo de la cognición para que podamos construir robots, ¿por qué tendríamos que mantenerla para entender la cognición humana? Hoy en día hay varias propuestas que se toman en serio esta posibilidad. Por un lado hay propuestas en las que el comportamiento inteligente se entiende como un producto de un organismo como un todo (Gallese, 2003 y 2005, Gallagher 2005, Wheeler 2005, Keijzer 2001) y propuestas en las que se trata de modelar la

mente como situada en un entorno que incluye diferentes organismos o “cuerpos” (Clark 2003). Veamos ahora cómo este tipo de tendencias en las ciencias cognitivas a alejarse de la idea de la cognición como un procesamiento de representaciones correspondentistas tiene implicaciones para la manera en la que entendemos a la ciencia como evolución cultural.

La importancia que han adquirido los modelos evolucionistas en la generación de explicaciones en diferentes áreas de la ciencia (en las ciencias sociales, la economía y la psicología, por ejemplo) también se ha hecho sentir en la filosofía de la ciencia. Hay una gran variedad de modelos evolucionistas que tratan de servir de base para explicar la estructura y dinámica de la ciencia. Hay modelos que tratan de explicar el progreso científico como un tipo de evolución basado en la caracterización de un tipo de mecanismo de selección natural distintivo de la cultura científica, por ejemplo, la manera como el crédito por nuevas ideas sirve de alicata para el desarrollo de investigaciones¹¹. Algunas de las explicaciones evolucionistas simplemente dan cuenta de la actividad científica en la medida que lo que pretenden explicar es el comportamiento social en general, y consideran que la ciencia es un caso especial de ese tipo de comportamiento. Parten de un ejemplo paradigmático de cómo los modelos evolucionistas pueden explicar el comportamiento socialmente estructurado, y entonces la solución se extrapola a otros sistemas de organización social, la ciencia, por ejemplo. La sociobiología de E. Wilson es un ejemplo de ese tipo de enfoque (Wilson, 1975). En esta propuesta la evolución orgánica se extiende para explicar el comportamiento.

Esto no es la manera más usual de utilizar modelos evolucionistas en la explicación del desarrollo social ni de formular explicaciones evolucionistas del desarrollo científico. El libro de Daniel Dennett 1995, ejemplifica otra estrategia con muchos seguidores hoy día. Se parte del supuesto que la manera como el mecanismo de la selección natural da lugar a adaptaciones es simplemente un caso particular de un mecanismo explicativo relevante en la explicación de prácticamente cualquier proceso natural o social; una vez que ciertas cosas que pueden ser copiadas fielmente (“replicadas”) existen, la adaptación por selección es una consecuencia del hecho que la replicación es diferencial, y que algunas cosas van a replicarse más rápidamente que otras. En el caso de la cultura los replicadores, en analogía con el concepto de gene en biología, se les denomina “memes”. En el marco de este enfoque hay varios trabajos que tratan de decirnos qué es la cultura y qué es la ciencia en particular. Una objeción importante a este tipo de enfoque es que los memes no tienen la fidelidad necesaria como para apoyar una explicación evolucionista. Dawkins da el siguiente ejemplo¹². Le mostramos a un niño un dibujo de un barco chino y le pedimos que lo dibuje. El dibujo se le enseñamos a un segundo niño y le pedimos que lo dibuje. El resultado se lo damos a un tercer niño, y así sucesivamente. El resultado de este experimento pensado es claro, después de unos veinte dibujos el resultado va a ser tan diferente del primero que no va a poder establecerse una relación entre ellos. Dawkins piensa que esta objeción puede responderse si consideramos un segundo experimento pensado. En lugar de enseñarle al primer niño un dibujo, le enseñamos a construir un barco con la técnica de doblado de papel conocida como Origami. Cuando el segundo niño ha aprendido la técnica le

¹¹ Hull 1988.

¹² En el prefacio Blackmore 1999.

pedimos que se la enseñe a un segundo niño, y así sucesivamente. Dawkins piensa que el resultado es predecible. Si bien es posible que un niño en la cadena olvide algún paso otro niño puede darse cuenta del salto (o error) y termine con un barco muy parecido al primero. Estoy de acuerdo con Dawkins en pensar que en el segundo caso, pero no en el primero, procesos de corrección entren en juego y que la replicación tienda a ser estable por esta razón. Pero Dawkins no dice para nada cómo puede explicarse el origen de esta “auto-normalización”, y en esta medida él solo muestra un problema para la memética, no una solución. El problema es cómo explicar (en ciertos casos) la estabilidad de la transmisión de las instrucciones de replicación.

Creo que la respuesta a esta pregunta tiene que pasar por una discusión explícita de la manera como se está modelando la cognición humana. Usualmente los proponentes de la memética tienden a hablar de acuerdo con el paradigma dominante en las ciencias cognitivas respecto a la naturaleza de la cognición, a saber, que la cognición consiste esencialmente en el procesamiento de representaciones (internas) correspondentistas. Desde esta perspectiva tiene mucho sentido hablar de memes como habla Blackmore en su libro y como hablan en general los proponentes de la memética: la cultura se entiende como un conjunto de representaciones internas que se transmiten con diverso éxito a través de individuos humanos. Pero las razones que dan Brooks y los otros proponentes de cognición anclada (o corporalizada) dejan claro que este tipo de proyecto parte de un supuesto que va en contra de desarrollos empíricos en Inteligencia Artificial y en psicología cognitiva.¹³ El mismo ejemplo de Dawkins sugiere que este modelo de la cognición tiene que abandonarse en la medida que no tiene manera de explicar el origen de la normatividad asociada con esa “auto-normalización”.

El tipo de proceso auto-normalizador indispensable para explicar la estabilización de las instrucciones a través de generaciones puede entenderse fácilmente si en nuestro modelo evolucionista tomamos en cuenta las normas que conforman una práctica (y su relación con otras prácticas). El niño que aprende Origami no sólo aprende mecánicamente a doblar un papel, sino que aprende una cierta técnica, y por lo tanto ciertas normas asociadas con fines y con criterios de corrección que por lo menos en parte provienen de otras prácticas. Por ejemplo, el hecho de que un doblez no encaje nítidamente con un doblez anterior sugiere por lo menos que hay algo que no está bien. Esta explicación, sin embargo, requiere que por lo menos algunas representaciones no sean correspondentistas. Se requiere que las representaciones que constituyen la cultura estén implícitas en las prácticas y técnicas que explican la “auto-normalización”, y por lo tanto, se requiere que las representaciones en cuestión no sean “memes” sino, en todo caso, lo que podemos llamar “anclas culturales” con capacidad de replicación a través de las prácticas (técnicas, etc.) en las que están corporalizadas implícitamente. Las anclas no se distinguen por la información que portan, sino por su papel en la estructura y función de actividades componentes de prácticas con capacidad de modificación y replicación.

Este tipo de normas permiten explicar la estabilización de las instrucciones, simplemente como una consecuencia de los mecanismos que estabilizan una práctica humana y científica en particular.¹⁴ Pero entonces una explicación evolucionista de la

¹³ Además de las referencias en la nota 5 hay una literatura importante en psicología del desarrollo que apunta en esta dirección. Ver por ejemplo Bickhard 2001.

¹⁴ En Martínez 2003 se sugiere una manera en particular en la que esta estabilización tiene lugar. Wimsatt y Griesemer en varios trabajos (en particular en Wimsatt y Griesemer en prensa) han hecho ver la

estabilización de las instrucciones requeriría recursos que la memética deja de lado. Requeriría poner atención en los diferentes tipos de normatividad que juegan un papel en la estabilización de recursos conceptuales y prácticos, i.e. aquellos recursos que juegan un papel en la evolución de las prácticas, y no simplemente centrar la explicación en un modelo de “memes” que en última instancia parecen siempre entenderse como representaciones internas de algún tipo, o en todo caso se asume que son representaciones correspondentistas. Este tipo de supuesto respecto a lo que es la cultura no es exclusivo de la memética. Boyd y Richerson (1985), por ejemplo, han desarrollado un modelo evolucionista muy conocido de la cultura. Es un modelo diferente en la medida que la explicación no descansa tanto en un mecanismo de selección que actúa sobre memes, sino en el poder explicativo de modelos de poblacionales. Sin embargo, coincide con los modelos de la memética en tener como uno de sus supuestos centrales, y explícitos, que “*la cultura es información guardada en los cerebros humanos*”. Decir que la cultura es “información guardada en los cerebros humanos” asume que la cognición tiene lugar a través de procesos basados en representaciones correspondentistas. Un supuesto que hemos visto es problemático.

6. **Conclusión.** En este trabajo he sugerido que si pensamos la cognición al tenor de un modelo interaccionista, del tipo que Brooks propone, entonces una respuesta al problema de cómo se estabilizan los productos culturales, y en particular las instrucciones de las que habla Dawkins en su ejemplo de evolución cultural, tiene que verse de una manera muy diferente. Un modelo evolucionista de la cultura tiene que ir de la mano de un modelo evolucionista de la cognición, o más bien, *un modelo de la evolución de la cultura tiene que ser al mismo tiempo un modelo de la cognición interaccionista*. La técnica de Origami es estable porque esa técnica involucra el despliegue y alineación de normas orientadas por un fin y criterios de corrección que se apoyan en otras prácticas (por ejemplo prácticas de uso de papel que promueven el valor de la precisión en un doblado, prácticas de manipulación de objetos que promueven valorar un resultado “elegante”, prácticas didácticas que promueven el desarrollo neuro-motor de los niños, etc.). Una explicación de la estabilidad de normas en el proceso de replicación tiene que ir más allá de lo que involucra la mera copia mecánica de un prototipo. Lo que se quiere copiar es algo que tiene que entenderse como un contorno normativo que involucra cosas materiales.

Quizás sea importante recalcar que esto no debe entenderse como que las normas que guían las inferencias o que nos dicen cuales inferencias o explicaciones son permitidas o plausibles, pueden entenderse como derivadas meramente de fines (el tipo de propuesta que sustenta la idea de la racionalidad como mera racionalidad instrumental). No tenemos porqué pensar que a las prácticas subyace una estructura de fines que nos permite entender la estructura de las prácticas en término de esos fines. Los fines sin la arquitectura o geografía de la red de prácticas en las que se distribuye la normatividad no dicen mucho sobre la manera como las prácticas evolucionan. Esta

importancia del desarrollo en un modelo evolucionista de la cultura. El tipo de propuesta que hacemos aquí encaja bastante bien con esta idea. Si las unidades de evolución son representaciones no correspondentistas, entonces no sólo es fácil entender cómo tiene lugar la evolución de la cognición como parte de la evolución de la cultura, sino que permite entender la manera como el desarrollo biológico y el desarrollo de prácticas e instituciones juega un papel en la evolución cultural.

geografía de prácticas está en gran medida implícita en patrones contingentes de interacción que se representan implícitamente en los indicadores de futuras acciones (bajo ciertas condiciones). Los fines son parte del ambiente normativo en el que se apoya la estabilidad de las normas, no son algo ontológicamente primario o epistemológicamente privilegiado. Qué es un fin cambia dependiendo de la perspectiva de uso o las expectativas representadas en actividades, prácticas e instituciones—incluso, podría decirse que muchos de los fines son implícitos en mismo el sentido de *implícito* que se ha manejado en este artículo.

Desde esta perspectiva explicar la evolución de las prácticas científicas (y en particular la evolución de las normas implícitas y explícitas que estructuran esas prácticas) es una manera de responder a las preocupaciones que motivaban las discusiones clásicas de la racionalidad científica. Así, si bien podemos estar de acuerdo con Feyerabend y Hacking en que el problema tradicional de la racionalidad (en la medida que se formula como el problema de la base racional que podemos tener para aceptar o rechazar teorías) es un problema que no tiene mucho de interés hoy en día, las preocupaciones de fondo siguen siendo importantes, y pueden retomarse una vez que reconocemos que el problema de la racionalidad debe de entenderse como el problema de dar cuenta de la evolución de normas implícitas (y a veces explícitas) en prácticas.

REFERENCIAS BIBLIOGRÁFICAS

Blackmore, S., 1999, *The Meme Machine*, OUP, Oxford and New York.

Boyd, R., and Richerson, P. J. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Brooks, Rodney A., 1999, *Cambrian Intelligence: The Early History of the New AI*, MIT Press, A Bradford Book, Cambridge, MA.

Bickhard Mark, 2001, “Why Children Don’t have to Solve the Frame Problem: Cognitive Representations Are not Encodings”, en *Developmental Review* 21, pp. 224-262.

Bickhard, Mark H. & Loren Terveen, 1995, *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution*, Advances in Psychology, Vol. 109.

Clark, Andy, 2003, *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*, OUP, Oxford.

Conein B. y L. Thévenot, Ed., 1997, *Cognition et information en société*, L’Ecole des Hautes Edúdes en Sciences Sociales, Paris.

Dawkins R. prefacio a Blackmore 1999.

- Dennett, Daniel C., 1995 *Darwin's dangerous idea evolution and the meanings of life*, Simon & Schuster, New York.
- Fodor, Jerry A., 1990, *A theory of content and other essays*, MIT Press, Cambridge, MA.
- Gallagher, S., 2005, *How the Body Shapes the Mind*, OUP, Oxford.
- Gallese, V., 2003, "Motor Ontology: The representational reality of goals, actions, and selves." *Philosophical Psychology*, vol 13, 3, 365-388.
- Gallese, V., 2005, "Embodied Simulation: from neurons to phenomenal experience". *Phenomenology and the Cognitive Sciences*, 4:23-48
- Hacking, Ian, (1983/1996) *Representar e intervenir*, Paidós-UNAM, México, 1996, (edición original 1983).
- Harnad, Stevan, 1990, "The symbol grounding Problem", en *Physica D*, Vol. 42, pp.335-346
- Hendriks-Jansen, Horst, 1996, *Catching ourselves in the act situated activity, interactive emergence, evolution, and human thought*, MIT Press, Cambridge, MA.
- Hull, David, 1988, *Science as a process: An evolutionary account of the social and conceptual development of science*, University of Chicago Press, Chicago.
- Keijzer, F., 2001, *Representation and Behavior*, MIT Press, Cambridge, MA,
- Martínez, Sergio F., 2003, *Geografía de las practicas científicas: racionalidad, heurística y normatividad*, Ed. UNAM – Instituto de Investigaciones Filológicas, México.
- Staddon, J. E. R., 2001, *The New Behaviorism: Mind, Mechanism and Society*, Psychology Press, Philadelphia.
- Thévenot, Laurent, 1997 "Un gouvernement pare les normes. Pratiques et politiques des formats d'information", en *Cognition et information en société*, comp. B. Conein y L. Thévenot, Ed. L'Ecole des Hautes Edúdes en Sciences Sociales, Paris, 1997.
- Wheeler, M., 2005, *Reconstructing the Cognitive World*, MIT Press, Cambridge, MA.
- Wilson E. 1975, *Sociobiology, the new synthesis*, Harvard UP

Wimsatt W. & Griesemer J. R., En prensa, "Reproducing Entrenchments to Scaffold Culture: The Central Role of Development in Cultural Evolution" por publicarse en R. Sansome and R. Brandon eds. *Philosophy of Evolutionary Developmental Biology*, MIT Press.