Jonatan García Campos

EPISTEMOLOGÍA Y PSICOLOGÍA COGNITIVA

UN ACERCAMIENTO AL ESTUDIO DE LA JUSTIFICACIÓN

Centro de Estudios
Filosóficos, Políticos y Sociales
Vicente Lombardo Toledano

Jonatan García Campos es doctor en Filosofía de la Ciencia por la UNAM. Investigador de tiempo completo en el CEFPSVLT, sus principales líneas de investigación son la epistemología contemporánea y las ciencias cognitivas, especialmente aquellos estudios sobre razonamiento, estructura de la mente y evolución. Regularmente imparte cursos en la Facultad de Filosofía y Letras de la UNAM.

EPISTEMOLOGÍA Y PSICOLOGÍA COGNITIVA

CENTRO DE ESTUDIOS FILOSÓFICOS, POLÍTICOS
Y SOCIALES VICENTE LOMBARDO TOLEDANO

DIRECCIÓN GENERAL
Marcela Lombardo Otero
SECRETARÍA ACADÉMICA
Raúl Gutiérrez Lombardo
COORDINACIÓN DE INVESTIGACIÓN
Aura Ponce de León
COORDINACIÓN DE SERVICIOS BIBLIOTECARIOS
Javier Arias Velázquez
COORDINACIÓN DE PUBLICACIONES Y DIFUSIÓN
Fernando Zambrana

Primera edición 2009
© CENTRO DE ESTUDIOS FILOSÓFICOS, POLÍTICOS
Y SOCIALES VICENTE LOMBARDO TOLEDANO

Calle V. Lombardo Toledano num. 51
Exhda. de Guadalupe Chimalistac
México, D.F. c.p., 01050
tel: 5661 46 79, fax: 5661 17 87
lombardo@servidor.unam.mx
http://www.centrolombardo.edu.mx

COLECCIÓN ESLABONES EN EL DESARROLLO DE LA CIENCIA

ISBN 978-607-466-013-5

La edición y el cuidado de este libro estuvieron a cargo
de las coordinaciones de investigación y de publicaciones del CEFPSVLT.

Ilustración de la portada: Luis Islas Aranda

JONATAN GARCÍA CAMPOS

EPISTEMOLOGÍA Y PSICOLOGÍA COGNITIVA

UN ACERCAMIENTO AL ESTUDIO DE LA JUSTIFICACIÓN

Centro de Estudios
Filosóficos, Políticos y Sociales
Vicente Lombardo Toledano

A mis padres

ÍNDICE

Introducción 1

I. De la noción de justificación
al estudio de la razón 11

I. 1. Fundacionismo y coherentismo 12

I. 2. Internismo y externismo 21

I. 3. La teoría confiabilista de la justificación 27

I. 4. Críticas al confiabilismo 31

I. 5. Dos ejes de debate en la teoría
contemporánea de la justificación 34

I. 6. Conclusiones 38

II. La tradición de heurística
y sesgo y la racionalidad 43

II. 1. Tres modos de acercarse
al estudio de la razón 44

II. 2. La inquietante evidencia empírica
acerca de cómo razona el ser humano 47

II. 2. 1. Las tareas de selección 47

II. 2. 2. La falacia de la conjunción 49

II. 2. 3. El rechazo a la "razón de base" 50

II. 2. 4. Toma de decisiones 52

II. 2. 5. Ajuste y anclaje 52

II. 3. La interpretación pesimista de la evidencia empírica 54

II. 4. ¿Puede demostrarse empíricamente que los seres humanos son irracionales? 57

II. 5. Dos modos de entender la interpretación pesimista de la racionalidad humana 62

II. 6. La visión estándar de la racionalidad 64

II. 7. Conclusiones 66

III. La psicología evolucionista y la racionalidad 73

III. 1. Ideas centrales alrededor de la psicología evolucionista 74

III. 2. El programa de investigación de la psicología evolucionista 77

III. 3. La hipótesis frecuentista 79

III. 4. La hipótesis de la detección del tramposo 84

III. 5. La conclusión optimista de la psicología evolucionista 89

III. 6. La psicología evolucionista y la interpretación pesimista de la racionalidad humana 91

III. 7. Conclusiones 95

IV. La interpretación pesimista de la evidencia empírica y las visiones de la racionalidad 103

IV. 1. Una respuesta pragmática a la interpretación pesimista 104

IV. 2. Objeciones relacionadas con la interpretación y aplicación de la visión estándar de la racionalidad 107

IV. 2. 1. El rechazo de interpretaciones no frecuentistas de la teoría de la probabilidad 107

IV. 2. 2. El problema de la derivación 112

IV. 3. El rechazo de la visión estándar de la racionalidad 116

IV. 3. 1. El acercamiento deontológico de la racionalidad 116

IV. 3. 2. La limitación de recursos de los sujetos cognoscentes 121

IV. 4. La dependencia de recursos y la interpretación pesimista de la racionalidad humana 125

IV. 5. Conclusiones 128

V. La teoría dual de sistemas 137

V. 1. La teoría dual de sistemas 138

V. 2. La teoría dual de sistemas y la interpretación pesimista de la racionalidad humana 149

V. 3. Algunas dificultades presentes en la teoría dual de sistemas 157

V. 4. La teoría dual de sistemas y la racionalidad 161

V. 4. 1. Racionalidad personal y racionalidad impersonal 161

V. 4. 2. Racionalidad evolucionista y racionalidad normativa 166

V. 5. Conclusiones 168

VI. La teoría dual de sistemas y las nociones de justificación y de racionalidad 173

VI. 1. La noción de justificación y la teoría dual de sistemas 174

VI. 1. 1. La teoría dual de sistemas y una noción no excluyente de justificación 176

VI. 1. 2. La teoría dual de sistemas y el internismo 182

VI. 1. 3. La teoría dual de sistemas y el externismo 186

VI. 2. Racionalidad y justificación desde la teoría dual de sistemas 191

VI. 2. 1. Externismo y la distinción entre racionalidad 1 y racionalidad 2 194

VI. 2. 2. Externismo y la distinción entre racionalidad evolucionista y racionalidad normativa 199

VI. 3. Una noción externista cognitivista de la justificación 202

VI. 4. Conclusiones 208

VII. Conclusiones. Racionalidad, razonamiento y la noción de justificación epistémica 211

Notas 219

Bibliografía 241

INTRODUCCIÓN

Un supuesto común en la epistemología es que hay un vínculo importante entre las nociones de justificación y de racionalidad. Usualmente, la tercera condición de la definición tradicional de conocimiento [1] se enuncia como: *S* está justificado en creer que *p* o *S* cree racionalmente que *p*. A pesar de la aparentemente obvia relación entre la noción de justificación y racionalidad, nunca se ha explicado claramente en qué consiste dicha relación. Esto se debe, en gran medida, a que no existe consenso de qué se debe entender por racionalidad, ni tampoco existe una idea clara de cuál es la noción de justificación que explica cuándo una creencia está justificada.

La noción de justificación ha sido un foco de discusión en la epistemología de los últimos años. En gran parte, esta discusión fue impulsada por el desafío que representaron los argumentos tipo Gettier a la noción tradicional de conocimiento [2]. Han surgido, por un lado, teorías fundacionistas de la justificación y teorías coherentistas y, por otro, teorías internistas y teorías externistas. Al combinar estas opciones se tienen posturas internistas fundacionistas y coherentistas, junto con posturas externistas fundacionistas y, en menor medida, coherentistas. Si a la lista anterior se agregan las teorías híbridas de la justificación, esto es, teorías que intentan combinar distintos elementos de teorías aparentemente rivales, entonces se tiene un panorama completo del debate contemporáneo alrededor de las diversas teorías de la justificación (García Campos, 2003). Sin embargo, no es claro cuál de todas estas teorías es la más adecuada.

Con relación a la racionalidad, se tienen al menos dos acercamientos distintos en pugna. El acercamiento deontológico sostiene que lo que es constitutivo del buen razonamiento es el razonar *únicamente* conforme a los estándares normativos adecuados (Larry y Moore, 2008). Una manera particular de entender este

acercamiento es sosteniendo lo que Stein (1996) denomina la "visión estándar de la racionalidad". Esta visión de la racionalidad sostiene que un sujeto, una acción, una creencia, es racional si satisface los estándares normativos provenientes de los principios de la lógica y la teoría de la probabilidad. Sin embargo se ha puesto en duda que la visión estándar de la racionalidad ofresca un buen significado de que algo sea racional [3]. Una de las motivaciones para apoyar lo anterior es que los sujetos cognoscentes se encuentran inmersos en una gran variedad de limitaciones de recursos —por ejemplo, de tiempo, energía, poder computacional, memoria e información— por lo que una adecuada caracterización de las normas de razonamiento debe fijar su atención en las limitaciones de sistemas cognitivos particulares (Cherniak, 1986). Este tipo de consideraciones, junto con otras, da lugar a un segundo acercamiento a la racionalidad, a saber, el acercamiento consecuencialista. Para éste, la racionalidad consiste en razonar de manera tal que probablemente se alcancen ciertas metas o resultados. Si bien existen diversas versiones del acercamiento consecuencialista de la racionalidad, *grosso modo* los defensores de esta visión mantienen que razonar únicamente de acuerdo con algún conjunto de reglas —por ejemplo, de acuerdo con la reglas de la lógica y la matemática— no es constitutivo del buen razonamiento.

Al observar una variedad de nociones de justificación y dos acercamientos distintos a la racionalidad, cabe preguntarse, ¿cómo se vinculan entre sí las distintas nociones de justificación con los acercamientos a la racionalidad que aquí he esbozado? En este trabajo exploraré distintas maneras de entender la relación entre la noción de justificación epistémica y la racionalidad a través de diversas teorías del razonamiento. En especial, mi interés particular es indagar si las teorías del razonamiento desarrolladas en la psicología cognitiva de los últimos años pueden indicar cuál de las nociones de justificación en pugna es la correcta. Pero antes de exponer cuál es la tesis que defenderé, es necesario argumentar brevemente por qué considero que el estudio del razonamiento puede decir algo acerca de las nociones de justificación y de racionalidad.

¿Por qué una teoría del razonamiento puede ayudar a entender la relación entre la noción de justificación y la de racionalidad? En el primer capítulo de este libro muestro cómo el debate en torno

a las nociones más importantes dentro de la justificación, esto es, las nociones internista y externista, se puede entender como un debate acerca de cuáles son las intuiciones epistémicas que debe recoger la noción de justificación [4]. Lo interesante de estas intuiciones es que en ellas el razonamiento y la racionalidad juegan un papel importante. Por sólo poner un ejemplo, una de las críticas a la noción externista de la justificación consiste en señalar que dicha noción no recoge la intuición de que un sujeto está justificado, si puede dar y ofrecer razones a favor de su creencia (Bonjour, 1980). El "dar y ofrecer razones", según esta crítica, está íntimamente ligado al razonamiento y entonces parece plausible pensar que el estudio del razonamiento humano puede iluminar el debate en torno a la noción de justificación.

El proponer que distintas teorías del razonamiento ayudan a entender cuál es la noción de justificación más adecuada conduce a ciertos problemas. En primer lugar, existen diversas teorías del razonamiento en la psicología cognitiva actual, por lo que habría que preguntarse cuál o cuáles serán las teorías que ayudarán a cumplir el cometido de este libro. En segundo lugar, las distintas teorías cognitivas responden a distintos problemas teóricos, muchos de los cuales no parecen compartir todas las teorías del razonamiento [5]. En tercer lugar, algunas veces los defensores de las diversas teorías del razonamiento tampoco comparten los mismos presupuestos metodológicos y/o epistemológicos. Por último, como mostraré en este libro, las diversas teorías del razonamiento asumen —explícita o implícitamente— distintos acercamientos a la racionalidad [6]. En suma, dentro de la psicología cognitiva del razonamiento, distintas teorías difieren entre sí en torno al tipo de problemas que deben resolver, sus presupuestos metodológicos y, en ocasiones, los supuestos normativos con los que están comprometidos [7]. De las diversas teorías del razonamiento que actualmente se discuten destacan la tradición de heurística y sesgo, la psicología evolucionista y, más recientemente, la teoría dual de sistemas. En este libro reviso estas tres teorías del razonamiento desarrolladas por la psicología cognitiva contemporánea.

De este modo, en este trabajo indagaré qué puede decir la psicología cognitiva acerca del debate en torno a la noción de justificación a través de la tradición de heurística y sesgo, de la

psicología evolucionista y de la teoría dual de sistemas; también analizaré, con la ayuda de estas teorías psicológicas, la relación entre diversas nociones de justificación y distintos acercamientos a la racionalidad. Estos son los dos objetivos centrales de este libro. Un proyecto como el que llevo a cabo claramente se encuentra comprometido con un tipo de epistemología naturalizada, que de modo muy general sostiene que los estudios empíricos sobre la cognición tienen consecuencias importantes en la epistemología [8]. Muchas veces se ha argumentado (Goldman, 1993) que los resultados de la psicología cognitiva, y en general de las ciencias cognitivas, son importantes para la epistemología en tanto que pueden dirigir las perspectivas del conocimiento humano y de lo que es una creencia racional al investigar los procesos cognitivos en diferentes dominios, como la memoria, la percepción o el razonamiento. Considero que las teorías del razonamiento esclarecen la naturaleza de los procesos y mecanismos con los que los seres humanos razonan, por lo que indagaré hasta qué punto el conocimiento brindado por la psicología cognitiva puede ayudar a entender las nociones de racionalidad y de justificación.

Se ha señalado que la epistemología contemporánea tiene varias tareas centrales, entre las que destacan el análisis del conocimiento, la evaluación de los procesos de formación de creencias y el ofrecer una respuesta al escéptico (Stich, 1990, Williams, 2001). El énfasis que diferentes filósofos le han dado a cada uno de estas tareas ha hecho pensar que en realidad se trata de distintos proyectos epistemológicos. Una vez que he señalado el supuesto naturalista del que parto, se verá que la separación de estos proyectos es menos clara de lo que comúnmente se ha pensado. A lo largo de este trabajo mostraré cómo el estudio y análisis de términos epistémicos, como el de la noción de justificación, puede enriquecerse de los resultados provenientes de la psicología cognitiva, por lo que la tarea de analizar los conceptos epistemológicos centrales es compatible y puede encontrar apoyo en los estudios de los procesos de formación de creencias hechos en disciplinas empíricas, como la psicología del razonamiento.

Una vez que he mencionado cuáles son los propósitos generales de este trabajo y el tipo de proyecto epistemológico del cual parto, es preciso apuntar la propuesta particular que defiendo. Argumentaré que hay razones para pensar que la teoría que de

mejor manera explica el razonamiento humano es la teoría dual de sistemas [9]. Si lo anterior es correcto, al hacer uso de la teoría dual se puede apoyar *un tipo de teoría externista* de la justificación. No obstante, sostendré que la teoría dual de sistemas no puede apoyar cualquier tipo de externismo, pero sí puede apoyar a un externismo *i.)* no veritista (aquel que acepta que no es con relación a la verdad con lo que se evalúan los procesos de razonamiento), y *ii.)* comprometido con al menos dos estándares normativos con los que se puede evaluar al razonamiento. Veamos.

Grosso modo, la teoría dual sostiene que existen dos sistemas de razonamiento que se interrelacionan, el primero de estos sistemas (S1) es rápido, innato, evolutivamente antiguo, inconsciente y es compartido con otros animales, mientras que existe otro sistema de razonamiento (S2) lento, reflexivo, consciente, controlado, dependiente del lenguaje y exclusivo de los seres humanos (Evans y Over, 1996, Stanovich y West, 2000, Sloman, 1996 [10]). Dadas las características de estos sistemas y el tipo de debate entre las nociones internista y externista de la justificación, argumentaré que habría tres maneras de "conectar" la teoría dual con el debate en teorías de la justificación:

1. Según las características de cada uno de los sistemas propuestos por los defensores de la teoría dual, es posible argumentar que S1 ofrecería creencias justificadas de modo externista, mientras que S2 ofrecería respuestas justificadas de modo internista (Eraña, en proceso b). Si se acepta esta postura, las dos nociones de justificación podrían complementarse porque evalúan creencias que tienen distinto origen.

2. Otra propuesta que ha sido defendida (Eraña, en proceso a) es que la noción de justificación implica necesariamente el punto de vista reflexivo, en el que los sujetos son epistemológicamente responsables y pueden guiar su razonamiento siguiendo las reglas de razonamiento adecuadas. En tanto sólo el sistema S2 da lugar al uso de reglas y al control consciente sobre los procesos de formación de creencias, Eraña (en proceso a) sostiene que las creencias producto de S2 estarían justificadas según criterios internistas. Por el contrario, las creencias causadas por S1 no podrían estar justificadas.

3. Sin embargo, como defenderé en este trabajo, el internismo por sí mismo es incompatible con la teoría dual. Si lo anterior es

correcto, las propuestas 1) y 2) esbozadas anteriormente deben ser descartas. En su lugar, sostendré que los dos sistemas de creencias pueden responder únicamente a la noción externista de la justificación [11], más precisamente argumentaré que los productos de S1 y S2 ofrecen creencias justificadas bajo criterios externistas. Esta es la tercera propuesta de cómo se puede relacionar a la teoría dual de sistemas con la noción de justificación.

Ahora bien, los defensores de la teoría dual se comprometen con una visión normativa particular. Según estos psicólogos deben ser dos los criterios que evalúan el razonamiento humano. Una pareja de psicólogos (Stanovich y West, 2003) han sostenido que la teoría dual del razonamiento está emparentada con dos modos distintos de entender la racionalidad, a saber, la "racionalidad normativa" y la "racionalidad evolucionista". El primer tipo de racionalidad es aquel que evalúa un proceso de razonamiento a partir de qué tan eficiente es éste para alcanzar metas y deseos individuales. La racionalidad evolucionista, por su parte, es aquella que evalúa un proceso de razonamiento a partir de su eficacia para lograr metas como la supervivencia y la reproducción. Lo que habría que resaltar de estos dos tipos de racionalidad postulados por Stanovich y West (2003) es que evalúan a los procesos de razonamiento a partir de los efectos o consecuencias de los mismos, lo cual supone un compromiso consecuencialista.

Evans y Over (1996), otra pareja de psicólogos defensores de la teoría dual, han postulado dos teorías normativas para evaluar la racionalidad distintas a las propuestas por Stanovich y West. Estas teorías normativas son la "racionalidad personal" y la "racionalidad impersonal [12]". La racionalidad personal evalúa el razonamiento humano a partir de qué tan eficiente es para alcanzar una meta personal. Por su parte, la racionalidad impersonal evalúa el razonamiento a partir de la satisfacción de un estándar normativo fundado en una teoría formal. Bajo la descripción que estos psicólogos hacen de sus dos teorías normativas, es posible argumentar que una de ellas (la racionalidad personal) es de tipo consecuencialista, mientras que la otra (la racionalidad impersonal) es de tipo deontológica.

Las propuestas normativas de estas dos parejas de psicólogos cognitivos muestra que no existe consenso entre los defensores de la teoría dual con relación a las teoría normativas con que se

debe evaluar el razonamiento. Sin embargo, estas propuestas normativas tienen en común *i.)* que debe haber al menos dos teorías normativas con las que se debe evaluar el razonamiento, esto es, no puede haber un único conjunto de criterios para evaluar el razonamiento, y *ii.)* que sus estándares normativos, eso es especialmente claro con respecto al consecuencialismo con el que se comprometen, no es de tipo veritista [13].

Había comentado anteriormente que sostendré que los productos de los dos sistemas de razonamiento están justificados según estándares externistas. Si a esta postura se agrega que, según los defensores de la teoría dual, se debe evaluar el razonamiento desde dos teorías normativas distintas, entonces se tiene que aceptar que la teoría externista que quiera fundarse en los postulados particulares propuestos por los defensores de la teoría dual tendría que aceptar dos tipos de justificación dentro de su propuesta. Además, si las teorías normativas defendidas desde la teoría dual no son veritistas y se desea fundar una noción externista sobre la teoría dual de sistemas en la que se encuentren las propuestas normativas que estos psicólogos proponen, entonces la noción externista de la justificación tendría que ser no veritista. A primera vista, este resultado no parece adecuado, pues en la literatura en torno a la noción de justificación abundan posturas externistas que se comprometen con consecuencialismos veritistas. La lección que esto nos deja es que los compromisos normativos hechos por la teoría dual impiden que posturas como el confiabilismo de Goldman puedan ser compatibles con esta teoría psicológica. Como ejemplo de una teoría no veritista y que esté fundada en la teoría dual de sistemas he esbozado muy brevemente un tipo de teoría externista consecuencialista casi al final de este trabajo. Si bien el objetivo de esta investigación no es proporcionar una noción de justificación original, considero importante trazar las líneas generales del tipo de teoría de la justificación que la teoría dual de sistemas y la psicología cognitiva del razonamiento apuntan.

La estructura de este libro es la siguiente. En el primer capítulo analizo diversas teorías de la justificación epistémica. Existen dos ejes del debate entre las teorías de la justificación: un eje que va del fundacionismo al coherentismo y otro eje que va del internismo al externismo. En tanto que estoy interesado en la noción de

la racionalidad y en ella juega un papel importante la relación que un sujeto tiene con sus creencias, centraré la discusión alrededor del internismo y externismo. Lo que se intenta mostrar en este capítulo es que detrás de las nociones de justificación defendidas por internistas y externistas se encuentran intuiciones epistemológicas distintas, viz., la noción internista de la justificación da cuenta de la intuición de dar y ofrecer razones junto con la responsabilidad epistémica; mientras que la noción externista de la justificación recoge la importancia de los procesos de formación de creencias adecuados o confiables (entre los que destacan el razonamiento inductivo y deductivo). Como las dos intuiciones hacen referencia a la razón, el segundo capítulo comienza con el estudio de ésta.

En el segundo capítulo expongo diversos experimentos sobre el razonamiento provenientes de la tradición de heurística y sesgo. En este capítulo presento algunas ideas de esta tradición psicológica, así como algunos de los supuestos subyacentes a esta teoría del razonamiento. Comienzo haciendo diversas distinciones acerca de cómo es posible acercarse al estudio de la razón y continúo describiendo algunos de los resultados más importantes en la literatura de la psicología cognitiva. Como se verá en este capítulo, existe un grupo importante de teóricos que sostienen que la evidencia empírica proporcionada por estos resultados muestra que la capacidad humana de razonamiento es normativamente problemática (Piatteli-Palmarini, 2005); a esta postura se le denomina la "interpretación pesimista de la racionalidad humana". En el capítulo no sólo intento mostrar el alcance de la interpretación pesimista y el tipo de presupuestos hechos por los defensores de esta tesis, sino también las implicaciones que ésta tiene en el debate en torno a la noción de justificación epistémica. En particular, argumentaré que la interpretación pesimista de la racionalidad, en su versión más fuerte, socava cualquier noción de justificación.

Parte de lo que pretendo en este libro es contrarrestar la visión pesimista que se origina de algunos experimentos hechos en psicología cognitiva, por lo que en el tercer capítulo reviso algunas de las tesis centrales de la psicología evolucionista, entre las cuales se sostiene que los sujetos cognoscentes disponen innatamente de una cantidad importante de módulos de razonamiento, algunos

de los cuales satisfacen los principios normativos apropiados. Si bien la postura evolucionista es ampliamente debatida, se da por supuesta la verdad de sus tesis centrales para indagar si con éstas es posible rechazar la interpretación pesimista de la racionalidad humana. La conclusión de este capítulo es que si bien la postura evolucionista supone el rechazo de una versión de la interpretación pesimista de la racionalidad humana, existe otra manera de entender la interpretación pesimista que la psicología evolucionista no podría contrarrestar. Al final de este capítulo evalúo los alcances de una interpretación optimista de la racionalidad para la noción de justificación.

En el cuarto capítulo reviso diversas objeciones a la interpretación pesimista de la racionalidad humana. Una de ellas es la denominada objeción pragmática, que se centra en factores lingüísticos en contextos experimentales y sostiene que parte de la evidencia usada para apoyar la interpretación pesimista puede ser explicada con referencia a cómo los sujetos interpretan las tareas o pruebas que los psicólogos piden que ejecuten. Argumentaré que la interpretación pragmática misma tiene sus problemas, ya que es incapaz de ofrecer una explicación general a la distinta evidencia actual sobre el razonamiento humano. Otra objeción que reviso y evalúo en este capítulo se centra en la manera en que se interpretan y se aplican las normas de la visión estándar de la racionalidad. Al explorar las anteriores objeciones considero que se debe hacer una revisión de esta visión estándar. Con esto en mente, en este capítulo sugiero la adopción de una explicación consecuencialista de la racionalidad. Junto con el acercamiento consecuencialista, exploro la noción de racionalidad dependiente de recursos, que puede ser integrada al acercamiento consecuencialista de la racionalidad. En el capítulo se concluye que si bien la adopción de un consecuencialismo dependiente de recursos no implica que la interpretación pesimista de racionalidad humana sea falsa, sí hace de la evaluación del razonamiento humano una tarea compleja, en tanto que no habría un único criterio para evaluar el razonamiento de cada uno de lo sujetos situados en diversas circunstancias.

En el quinto capítulo exploro y expongo varias versiones de las teorías duales de razonamiento. Argumentaré que existe un núcleo de afirmaciones que las distintas versiones de la teoría dual

admiten, pero que sus compromisos normativos divergen. En ese capítulo exploro cómo estas divergencias conducen a distintas maneras de entender la relación entre la teoría dual de sistemas y la racionalidad. Revisaré también la respuesta que la teoría dual de sistemas ofrece a los experimentos que han sido interpretados de manera pesimista y sostendré que, de momento, según lo revisado a lo largo del trabajo, la teoría dual de sistemas es la teoría que mejor explica el razonamiento humano.

En un último capítulo exploro las posibles conexiones entre la teoría dual de sistemas y la noción de justificación. Como indiqué anteriormente, una primera manera de relacionar a la teoría dual de sistemas con el debate en teorías de la justificación es sosteniendo que los distintos sistemas de razonamiento postulados por la teoría dual responden a nociones de justificación distintas. Una segunda opción es que sólo los productos de uno de los sistemas de razonamiento (S2) puede brindar creencias justificadas. Una tercera opción consiste en que los criterios externistas de la justificación pueden por sí mismos dar cuenta de los productos de los dos sistemas de razonamiento propuestos por la teoría dual. Esta tercera opción es la que encuentro más adecuada. Argumentaré que cualquier noción de justificación que se funde en la teoría dual de sistemas, aceptando sus propuestas normativas, tendría que defender un externismo no veritista que distinga al menos dos estándares normativos para evaluar el razonamiento humano.

La elaboración de este libro fue posible gracias al apoyo de tres instituciones: la UNAM a través de la Dirección General de Estudios de Posgrado, el Consejo Nacional de Ciencia y Tecnología a través de dos proyectos de investigación dirigidos por el doctor Sergio Martínez ("Filosofía de las Prácticas Científicas" y "Abstracción, Razonamiento y Argumentación") y el Centro de Estudios Filosóficos, Políticos y Sociales Vicente Lombardo Toledano. Agradezco a Salma Saab, Claudia Lorena García, Huang Xiang, Ángeles Eraña y Sergio Martínez sus sugerencias y críticas a una primera versión de este trabajo. Agradezco también los comentarios, paciencia y apoyo de Luis Noriega, Paola Hernández, Susana Ramírez, Raúl Gutiérrez y Fernando Zambrana.

I.
DE LA NOCIÓN DE JUSTIFICACIÓN AL ESTUDIO DE LA RAZÓN

Una de las discusiones más importantes en la epistemología contemporánea se da entre las distintas teorías de la justificación, que buscan explicar qué significa que una creencia esté justificada para un sujeto particular. Esas teorías han propuesto diversas respuestas a esa pregunta, pero a grandes rasgos se puede sostener que hay dos *ejes* que ordenan las teorías de la justificación. Uno de estos ejes lo constituyen el fundacionismo y el coherentismo, el otro eje lo constituyen el internismo y el externismo. El propósito de este capítulo es conocer cuáles son las posturas más relevantes dentro de la teoría de la justificación e indagar si existe una noción de justificación que cuente con las ventajas teóricas para ser considerada como la noción de justificación más adecuada.

A lo largo de este capítulo argumentaré que no existe una única noción de justificación que los epistemólogos acepten de manera consensuada. Por el contrario, sostendré que según el estado actual en teoría de la justificación pueden distinguirse varias nociones de justificación. Concluiré en este capítulo que estas nociones recogen diferentes intuiciones epistémicas de lo que significa estar justificado (Bonjour, 1980, Goldman, 1980). Al final de este capítulo sugiero que el estudio de la razón puede presentarse como una salida al debate actual en teoría de la justificación, puesto que las intuiciones defendidas por cada una de las nociones de justificación en pugna hacen referencia a la racionalidad y al razonamiento.

La estructura de este capítulo es la siguiente. En la primera sección presentaré el debate entre coherentismo y fundacionismo, centrándome en los argumentos que se han dado en contra del coherentismo y las críticas de esta postura al fundacionismo. En la segunda parte de este capítulo se deja de lado el debate entre

el fundacionismo y el coherentismo para centrarse en la discusión entre el internismo y el externismo. Profundizaré en el estudio de una teoría externista particular, el confiabilismo, para posteriormente revisar cuáles son las críticas hechas al confiabilismo y las respuestas que han ofrecido los confiabilistas para contrarrestarlas. Posteriormente mostraré cuáles son las intuiciones epistemológicas que tratan de integrar cada una de la nociones de justificación aquí expuestas y señalaré las coincidencias que existen entre éstas. Al final de este capítulo explicaré por qué el estudio del razonamiento y la racionalidad son pertinentes para el debate contemporáneo en torno a la justificación.

I. 1. FUNDACIONISMO Y COHERENTISMO

El fundacionismo es importante dentro de la epistemología porque representa una opción clásica en la definición de la justificación. Posiblemente es el proyecto epistemológico cartesiano el que ejemplifica de mejor manera una postura fundacionista, pero no sólo este proyecto epistemológico se puede identificar con el fundacionismo; en gran medida propuestas antirracionalistas como el empirismo lógico puede ser entendido como una postura de este tipo. El fundacionismo se puede caracterizar como aquella posición que acepta los dos siguientes principios:

a) Existen dos tipos de creencias, a saber: las creencias básicas, cuya justificación no depende de otras creencias, y las creencias no básicas, cuya justificación depende de su relación con las creencias básicas.
b) La justificación es unidireccional, esto es, la justificación se transmite sólo de las creencias básicas a las no básicas.

Un corolario de estos principios es que la justificación de las creencias no básicas depende siempre de un conjunto finito de relaciones que las conectan con las creencias básicas. Otra idea cercana a las posiciones fundacionistas es que estas teorías se inclinan por ofrecer *definiciones recursivas de la justificación*, esto es, se especifica una clase de elementos que están justificados —i.e., las creencias básicas— y se determina un conjunto de relaciones por las cuales los elementos restantes —las creencias no básicas— podrían estar justificadas [1].

Muchos fundacionistas han considerado que las creencias básicas son incorregibles, indubitables, ciertas, etc. Sin embargo, en los últimos años, los fundacionistas han sostenido que las propiedades esenciales de su posición no tienen necesariamente que comprometerse con posturas infalibilistas con respecto a las creencias básicas, basta con que la justificación de éstas no dependa de otras creencias (Alston, 1976).

Sin profundizar más en la caracterización de la teoría fundacionista de la justificación, es necesario apuntar los principios básicos de la teoría coherentista. Una teoría coherentista de la justificación es aquella que defiende dos ideas centrales que la distinguen del fundacionismo:

a) No existen creencias básicas que sirvan de fundamento a las demás creencias.
b) La justificación es sistémica u holista.

Estos dos principios coherentistas se pueden entender más apropiadamente si se comparan con el fundacionismo. La característica *a)* del coherentismo no supone una distinción entre distintos tipos de creencias, por lo menos conforme a su jerarquía epistémica; en principio, todas las creencias tienen el mismo estatus epistémico, es decir, no hay creencias que por sí mismas no requieran de la justificación de otras creencias. La característica *b)* del coherentismo también es muy importante, pues mientras que para el fundacionismo la justificación es unidireccional, el coherentismo cree que la justificación es holista o sistémica [2], es decir, que las creencias están justificadas al estar inferencialmente relacionadas con otras creencias en un contexto general de un sistema coherente.

Una vez que he presentado las tesis centrales del fundacionismo y el coherentismo parece relativamente claro cómo es que estas dos teorías de la justificación se contraponen. Por ello es necesario indagar si hay un criterio que permita aceptar una de estas dos teorías. Uno de estos criterios podría ser la respuesta que pueden dar estas teorías al denominado "problema del regreso de las justificaciones" (Bonjour, 1985). Dicho argumento se puede formular en su versión más sencilla como sigue:

1. la creencia *p* está justificada si se ofrece una creencia *q* [3] que la fundamente;
2. pero *p* está justificada por *q*, sólo si *q*, a su vez, está justificada por otra creencia;
3. la creencia *q* está justificada sólo si se ofrece una creencia *r* que la justifique;
4. pero *q* está justificada por *r*, sólo si *r* a su vez está justificada por otra creencia
5. En tanto que la justificación de las creencias siempre es lineal.

De ahí se sigue que

la creencia *p* nunca puede estar genuinamente justificada.

A continuación se describen las respuestas fundacionista y coherentista a este argumento.

La respuesta del fundacionista al argumento del regreso de las justificaciones parece obvia: el regreso al infinito se detiene en creencias que no están justificadas por otras creencias, esto es, las creencias básicas. Pero, ¿qué tan legítima resulta la postulación de creencias básicas por parte del fundacionismo? Una crítica importante a la postulación de creencias básicas en teoría de la justificación ha sido desarrollada por Bonjour y parte del supuesto de que la justificación siempre hace referencia a una meta o fin (Bonjour, 1978). Para este filósofo, el conocimiento requiere de la justificación epistémica, y el rasgo característico de este tipo de justificación —a diferencia de la justificación moral— es su relación interna o esencial con la verdad (García Campos, 2003). Una vez que se acepta esta idea, Bonjour afirma que el fundacionista debe sostener que si las creencias básicas están justificadas sin apelar a otras creencias, entonces las creencias básicas deben ser probablemente verdaderas, pero para ello es necesario ofrecer razones para creer que las creencias básicas son probablemente verdaderas, por lo que una creencia básica no podría estar justificada sin apelar a otras creencias. Para ilustrar lo anterior supóngase que la creencia básica *p* de *S* debe tener la propiedad *F* para ser considerada como verdadera, el fundacionista tendría que suponer el siguiente argumento para que la creencia *p* de *S* esté justificada:

1. La creencia de que *p* de *S* tiene la propiedad *F*.
2. Las creencias que tienen la propiedad *F* tienen una alta probabilidad de ser verdaderas.

Por lo tanto,

la creencia de que *p* de *S* tiene alta probabilidad de ser verdadera.

Si el argumento de Bonjour es válido, entonces el fundacionismo no puede dar una respuesta al argumento del regreso de las justificaciones, pues incluso las creencias básicas requerirían de otras creencias para estar justificadas, es decir, no existirían creencias básicas que detengan el regreso al infinito, ya que incluso tales creencias deben su justificación a su relación con otras creencias [4].

El fundacionista, a mi juicio, puede responder de dos maneras distintas a esta crítica hecha por el coherentista. Una primera respuesta es argüir que no siempre es necesario que una persona que ostenta una creencia básica, sepa o esté justificado por creencias extras que muestren que su creencia básica es probablemente verdadera. Estos fundacionistas afirman que es suficiente que las creencias básicas de un individuo estén justificadas, independientemente de si el individuo cree que éstas de hecho lo están (Alston, 1976). Esta respuesta conduce directamente al externismo. Para el externismo, el estatus de las creencias de un sujeto como constituyentes del conocimiento dependen principalmente de las relaciones de esas creencias con procesos cognitivos adecuados [5]. Las creencias básicas, que el fundacionista desea defender, están justificadas por relaciones a las que el sujeto cognoscente pudiera no tener acceso, por lo que las creencias básicas estarían justificadas independientemente de que el sujeto pueda ofrecer razones para tales creencias. Posteriormente revisaré con más detalle un tipo de teoría externista, por lo que de momento no explicaré cuáles son las consecuencias de aceptar esta postura. Una segunda respuesta consiste en afirmar que la justificación de las creencias básicas depende de estados cognitivos más primarios que no requieren de justificación, tales como *intuiciones* o la *aprehensión inmediata*. Las creencias básicas, afirman estos fundacionistas, están justificadas en virtud de la aprehensión de los estados mismos de cosas y, por lo tanto, las creencias básicas no requieren de la

justificación de creencias subsiguientes. A esta postura se le conoce como la teoría de *lo dado.*

El problema general con esta segunda alternativa fundacionista es que nunca se ha explicado de manera adecuada cuál es la naturaleza de la *intuición* o la *aprehensión inmediata* de lo dado. Según Bonjour, habrían dos maneras de entender la naturaleza de una intuición: *i.)* las intuiciones tienen contenido proposicional y por ello son consideraras cognitivas; o *ii.)* las intuiciones no tienen contenido proposicional, esto es, no son tomadas como cognitivas. Pero cada una de estas alternativas nos lleva a un problema, pues si las intuiciones son *i.),* esto es, si tienen contenido proposicional, entonces pueden brindar justificación, sin embargo, ellas mismas requerirían de justificación; mientras que si las intuiciones son *ii.),* es decir, si no tienen contenido proposicional, entonces no necesitan justificación pero serían incapaces de proveerla. Aquí radica para Bonjour (1978) el punto medular del porqué a la teoría de lo dado se le puede considerar un mito filosófico.

Si la respuesta ofrecida por el fundacionismo al problema del regreso de las justificaciones no es adecuado, ¿cuál es la salida que brinda el coherentismo a este problema? La respuesta coherentista al argumento del regreso de las justificaciones consiste en rechazar la concepción lineal o inferencial de la justificación. Para Bonjour, a pesar de su apariencia lineal, la justificación es esencialmente sistémica u holista; las creencias están justificadas al estar inferencialmente relacionadas en un contexto general de un sistema coherente de creencias. Cuando se justifica una creencia particular (dando por supuesto un sistema cognitivo) se puede considerar, según Bonjour (1985), que se está dentro de un nivel local de la justificación, pues se da por sentado el sistema de creencias en general para justificar una creencia particular. Pero también se puede poner en duda la validez del sistema general de las creencias en donde se debe buscar la justificación del sistema de creencias en sí, y en este caso se esta frente a un nivel global de justificación. El argumento del regreso de las justificaciones tiene entre sus premisas el supuesto de que la justificación es lineal; si esta premisa está equivocada, como cree el coherentista, y es modificada por la noción holista de la justificación,

entonces se consigue una salida coherentista al problema del regreso de las justificaciones.

Según los defensores de la teoría coherentista, su respuesta al argumento del regreso de las justificaciones no es circular. La acusación de circularidad surge de la idea de que para el coherentismo la justificación de la creencia *p* —según el argumento del regreso de las justificaciones— proviene de una creencia *q*, *q* de *r* y así sucesivamente hasta que llega el momento que *p* esté en la base de la justificación de la creencia *p* misma. Los coherentistas niegan que su noción de justificación sea circular, pues la justificación de una creencia empírica supone cuatro grandes pasos: *i.)* la inferencia de la creencia particular a partir de sus relaciones con las demás creencias del sistema, *ii.)* la coherencia del sistema total de creencias, *iii.)* la justificación del sistema total de creencias, y *iv.)* la justificación de la creencia particular en virtud de su inserción en el sistema de creencias (Bonjour, 1985). La coherencia de un sistema de creencias implica, además de la consistencia entre los miembros de un conjunto de creencias, las relaciones inferenciales entre las creencias de dicho sistema; la coherencia es una noción gradual y guarda una relación importante con la noción de explicación (Bonjour, 1985).

Existen varias objeciones a la noción coherentista de la justificación, dos de las más importantes son:

a) Si la noción de justificación depende sólo de la coherencia de un sistema de creencias, sería difícil elegir entre distintos sistemas de creencias alternativos e incompatibles, pues resultarían igualmente válidos.
b) La teoría coherentista parece hacer una separación con el mundo empírico del cual provienen las creencias, esto es, se argumenta que la teoría coherentista de la justificación parece estar desprovista de los insumos del mundo.

Bonjour confiesa que la aceptación de creencias provenientes de la experiencia, por ejemplo, creencias observacionales, hace pensar que existen creencias cuya justificación no depende de otras creencias. Si lo anterior fuese correcto, implicaría rechazar una de las ideas centrales del coherentismo: que todas las creencias están igualmente justificadas. Sin embargo, para Bonjour este argumento está equivocado, pues confunde el modo en que una

creencia fue introducida a un sistema de creencias y el modo en que dicha creencia está justificada [6].

> La relación entre una sensación y una creencia no puede ser *lógica*, debido a que las sensaciones no son creencias u otras actitudes proposicionales: la relación es *causal*. Las sensaciones causan algunas creencias y en *este* sentido son las bases o fundamentos de las creencias. Pero la explicación causal de una creencia no muestra cómo o por qué la creencia está justificada (Davidson, 1983, p. 428, la traducción es mía).

¿Cómo puede el coherentista explicar la justificación de una creencia observacional? Según Bonjour (1985), la justificación de una creencia observacional involucra tres aspectos: el primero es que una creencia observacional se origina espontáneamente a través del sentido de la vista; el segundo aspecto es que las condiciones de la observación puedan ser especificadas, por ejemplo, que el sentido de la vista de un sujeto cognoscente funciona normalmente, que haya buena iluminación, etc.; por último, existe el supuesto de que las creencias visuales espontáneas sobre un tipo de objetos de una tamaño medio son confiables, i.e., son altamente probables de ser verdaderas (de algún modo el sujeto debe ser consciente de esto, cree el coherentista). De modo que, la justificación de una creencia observacional se estructura como sigue:

i.) *S* tiene la creencia visual espontánea de que hay una manzana roja sobre la mesa.
ii.) Las creencias visuales acerca de objetos de un tamaño medio, acerca de su color y clasificación son generalmente confiables (esto es, son probablemente verdaderas).
iii.) Existen las condiciones especificadas por las cuales una creencia visual es confiable.

De donde,

la creencias de *S* de que hay una manzana roja sobre la mesa es confiable (o probablemente verdadera).

Este argumento muestra en qué sentido afirman los coherentistas que las creencias visuales espontáneas no están justificadas por una aprehensión inmediata como lo defienden los fundacionistas, sino que, por el contrario, la justificación de estas creencias

tiene, como cualquier otra creencia, un carácter inferencial. Con este esquema y con la distinción entre el origen y la justificación de una creencia se puede responder, según Bonjour (1985), la objeción *b)* hecha al coherentismo. Los coherentistas afirman que las creencias observacionales pueden alterar una parte significativa de un sistema de creencias, ya sea porque las creencias visuales sean inconsistentes con otras creencias o porque la modificación producida por algunas creencias altere ciertas áreas del sistema de creencias en general.

Bonjour va todavía más lejos cuando sostiene que una descripción adecuada del conocimiento empírico debe requerir —y no sólo permitir— los insumos del mundo dentro del sistema cognitivo. Para Bonjour (1976), un sistema cognitivo que busque tener un estatus de conocimiento empírico debe incluir leyes que atribuyan alta confiabilidad a una variedad razonable de creencias espontáneas cognitivas (especialmente a aquellas creencias introspectivas espontáneas que son requeridas para el reconocimiento de otras creencias espontáneas). A este principio o ley lo denomina Bonjour el *requisito de observación*. La idea básica que yace bajo el requisito de observación es que cualquier sistema de creencias empíricas debe, en principio, ser capaz de ser evaluado observacionalmente, ya sea directa o indirectamente, y de esta manera pueda ser confirmado o refutado. Entendido de esta manera, el requisito de observación efectivamente garantiza, afirma Bonjour (1976), que un sistema de creencias que lo satisfaga recibirá los insumos del mundo externo, de ahí que la justificación no dependerá solamente de las relaciones internas dentro de un sistema estático de creencias empíricas. No obstante, el hecho de que se requieran insumos del mundo externo no significa que dichos insumos reciban su justificación debido a su origen; la justificación de las creencias se lleva a cabo siempre con referencia a un sistema coherente de creencias.

Según Bonjour (1976), la respuesta a la objeción *b)* da algunas pistas para responder a la objeción *a)*, según la cual la noción de coherencia estipulada por la teoría coherentista es insuficiente para determinar la elección de distintos sistemas coherentes. Habría dos frentes de defensa en contra de esta objeción:

1. Es posible sostener que la coherencia de un sistema es una cuestión de grados, es decir, un sistema de creencias *M* puede

estar más justificado que *M'*, porque dentro del sistema *M* las relaciones entre las creencias son inferencialmente más fuertes que dentro del sistema de creencias *M'*. Esto implica que los distintos sistemas de creencias no son igualmente válidos, sino que algunos de ellos pueden tener una justificación mayor dependiendo de la coherencia interna entre sus creencias.

2. El requisito de observación hace patente, afirma Bonjour (1985), la probabilidad de que un sistema de creencias que cumpla con tal requisito y que resulta coherente en un momento dado podría dejar de serlo después de subsecuentes insumos que entren en contradicción con áreas importantes dentro de dicho sistema de creencias. Una vez hecho explícito lo anterior, Bonjour (1976) divide a la objeción *a)* en dos partes. La primera parte de la objeción es que en un momento dado existen distintos sistemas de creencias coherentes, pero incompatibles, de los cuales la teoría coherentista de la justificación no puede ofrecer bases para decidir cuál de ellos resulta más justificado. Esta afirmación, según Bonjour, es correcta. Parece concebible que en un momento determinado no haya razones para saber cuál de los sistemas de creencias incompatibles es el más adecuado, y esto no lo puede resolver el coherentismo ni, a juicio de Bonjour, ninguna otra posición epistemológica. La segunda y crucial parte de la objeción sostendría que la posición coherentista es siempre y en todo momento incapaz de señalar cuál de varios sistemas de creencias altamente coherentes pero incompatibles es el más adecuado. Esta crítica, cree Bonjour, es errada. De acuerdo con este filósofo coherentista, la idea del continuo impacto de la observación cierra la posibilidad de que todos los sistemas de creencias que parecen igualmente coherentes en un momento dado continúen en el futuro siendo igualmente coherentes dados los insumos del mundo externo. Esto se debe a que, para Bonjour, los insumos pueden modificar la coherencia de los sistemas de creencias. Así, el relativismo que resulta si se acepta el coherentismo, no es de la magnitud que comúnmente se supone esta teoría promueve.

De 1 y 2 queda resuelta, según Bonjour, la objeción *a)* al coherentismo.

Si lo anteriormente examinado fuese correcto, entonces sería plausible responder a las dos objeciones importantes que se han hecho al coherentismo. Las respuestas a estas objeciones radican,

según Bonjour, en entender el papel que la observación juega en una teoría de la justificación de esta naturaleza. La postura coherentista como la de Bonjour, además, parecería ofrecer una respuesta al problema del regreso de las justificaciones. Sin embargo, no es claro que todas las teorías de la justificación deban resolver este problema, ni que el coherentismo de Bonjour sea una postura libre de críticas. En la siguiente sección presentaré dos teorías de la justificación que pueden ser compatibles o no con el fundacionismo y el coherentismo, esto es, el debate entre el internismo y el externismo. Primero expondré en qué consiste el internismo, puesto que el coherentismo de Bonjour explícitamente se compromete con una posición como ésta. Posteriormente mostraré algunas críticas a esta postura para presentar entonces la teoría externista de la justificación.

I. 2. INTERNISMO Y EXTERNISMO

Una teoría internista de la justificación es aquella teoría que sostiene que los rasgos que justifican a una creencia son un conjunto de propiedades a las cuales el sujeto tiene acceso de manera inmediata. En general, el internista afirma que el agente cognoscitivo debe tener acceso inmediato a las condiciones necesarias y suficientes que justifican su creencia (King, 2000). Si lo anterior es el caso, siempre que un sujeto está justificado es porque dicho sujeto tiene acceso a las condiciones de justificación de sus creencias. Por lo tanto, desde la postura internista, siempre que un sujeto está justificado, dicho sujeto cree justificadamente que lo está. Pappas (2005) distingue entre dos tipos de *acceso* dentro del internismo: *i.) acceso actual* es aquella postura internista que sostiene que siempre que un sujeto *S* cree justificadamente que *p*, entonces *S* es consciente de las razones que justifican a *p*; y *ii.) accesibilidad* es aquella postura internista en donde siempre que un sujeto *S* cree justificadamente que *p*, entonces *S* puede estar consciente por medio de la reflexión de las razones que justifican a *p*. Casi todos los internistas, como bien señala Pappas, han sostenido esta segunda manera de entender acceso interno. A lo largo de este libro mantendré la idea de que la noción internista de justificación supone un tipo de acceso o accesibilidad a los rasgos que justifican las creencias de los sujetos [7].

En el coherentismo de Bonjour lo anterior se hace patente al considerar que es el acceso que el sujeto tiene a su relación entre creencias y la coherencia entre éstas lo que justifica las creencias de los sujetos.

> La persona debe *saber* todas estas cosas, al menos de una manera... aproximada. Debe ser capaz de reconocer creencias que son resultado de los procesos en cuestión... debe saber que tales creencias son confiables, y debe saber en un caso dado que cada condición necesaria para la confiabilidad está satisfecha (Bonjour, 1976, p. 295, la traducción es mía).

El problema al que se enfrenta el coherentismo de Bonjour surge cuando, por una parte, sostiene que el sujeto epistémico debe tener acceso al sistema de creencias para justificar a una creencia particular, comprometiéndose así con el internismo, pero al mismo tiempo, parece que ningún sujeto epistémico puede tener acceso o puede aprehender completamente su sistema de creencias ni a los principios que juegan un papel importante dentro de la justificación coherentista. Dos críticos a la noción de coherencia como criterio que pueda ser usado para definir la noción de justificación o la de algún principio de racionalidad han sido Cherniak (1986) y Goldman (1993). El argumento de estos dos filósofos parte del supuesto coherentista de que la noción de coherencia implica la noción de consistencia, esto es, todo grupo de creencias coherentes debe ser consistente (Bonjour, 1985).

¿Pero cómo pueden los seres humanos evaluar si su sistema de creencias es consistente? Una manera de conocer la consistencia de un conjunto de creencias es a través de las tablas de verdad. El resultado de una tabla de verdad en un conjunto consistente de creencias debe ser aquel en donde al menos uno de los resultados de tal tabla de verdad sea verdadero. Sin embargo, al construir una tabla de verdad para comprobar si un conjunto es consistente se debe hacer referencia al número de proposiciones atómicas que intervienen en un conjunto de creencias, donde resulta que el número de renglones de nuestra tabla de verdad es siempre 2^n. Incluso si el sistema de creencias contara con 138 proposiciones, el número de renglones para revisar la tabla de verdad sería de 2 elevado a la potencia 138 (esto es 3.5×10^{41}). Así, sería imposible

para cualquier ser humano cumplir con el requisito de que para ser racional o tener creencias justificadas se debe suponer un sistema consistente de creencias. De este modo, según Goldman, la consistencia podría ser vista como un ideal epistémico, que trasciende por mucho las capacidades de los sujetos para alcanzarlo (Goldman, 1993, p. 218).

Lo anterior brinda razones para sostener que difícilmente los sujetos cognoscentes podrían recurrir al conjunto total de creencias para poder justificar una creencia particular. De donde, si se acepta que para tener creencias justificadas el sujeto deber tener acceso a su sistema general de creencias, pero es imposible que un sujeto tenga acceso a su sistema total de creencias, entonces se sigue que los sujetos, bajo una perspectiva coherentista no podrían tener creencias justificadas. A mi juicio, las cosas empeoran cuando Bonjour (1985) afirma que una de las tareas que debe cumplir una teoría de la justificación es ofrecer un estándar o criterio bajo el cual se pueda especificar cuándo una creencia está justificada. Asimismo, Bonjour considera que un criterio o estándar por sí mismo no justifica a una creencia, si no se está justificado en creer que dicho estándar o criterio conduce a la verdad. A esta segunda justificación la denomina Bonjour la "metajustificación" de una teoría de la justificación.

> La tarea central de una teoría del conocimiento empírico se divide en dos... la primera es dar una descripción de los estándares de justificación epistémica; y la segunda es proveer lo que denomino una metajustificación para la descripción propuesta, al mostrar que los estándares propuestos conducen adecuadamente a la verdad (Bonjour, 1985, p. 9, la traducción es mía).

El problema con la metajustificación, es que hay razones para pensar que Bonjour no sólo considera que los criterios o estándares deben estar justificados para los epistemólogos, sino para cualquier sujeto cognoscente.

> Si un supuesto sujeto cognoscente dado es epistemológicamente responsable al aceptar creencias en virtud de que éstas satisfacen los estándares de una descripción epistemológica, entonces parece seguirse que una apropiada metajustificación de tales estándares debe, en principio al menos, ser disponible para *él*... tal metajustificación

> podría ser más o menos tácita o implícita (Bonjour, 1985, p. 10, la traducción es mía).

Esto parece indicar que Bonjour considera que un sujeto *S* está justificado en creer que *p* no sólo si la creencia *p* de *S* cumple con las condiciones que exige el estándar para justificar dicha creencia —en el caso de Bonjour la coherencia con el sistema de creencias y otras consideraciones— sino que al mismo tiempo *S* debe estar justificado, al menos implícitamente, en creer que dicho estándar —en este caso la coherencia— muy probablemente conduce a la verdad. Esto supondría que, para Bonjour, ningún estándar puede por sí mismo justificar a una creencia, a menos que el sujeto esté justificado en creer que dichos estándares conducen a la verdad. Este punto pone de manifiesto una de las consecuencias del internismo que he mencionado anteriormente, esto es, para un internista siempre que *S* está justificado, *S* puede creer justificadamente que está justificado.

Dado que el estándar propuesto por Bonjour es la coherencia de una creencia con un sistema de creencias —entre otras condiciones— la pregunta pertinente es: ¿cómo se puede justificar la creencia de que la coherencia lleva a la verdad? Para responder a esta pregunta Bonjour no recurre a la coherencia del sistema de creencias, pues a su juicio esta estrategia sería circular. La justificación aquí debe ser entonces, según Bonjour (1985), *a priori*, i.e., no debe apelar a ninguna premisa empírica. La circularidad desaparece si Bonjour apela a una justificación *a priori* de la coherencia como estándar de la justificación. Sin embargo, su noción de justificación se complica de la siguiente manera:

> La creencia *p* de *S* está epistémicamente justificada si y sólo si hay un estándar correcto D que *p* satisface y *S* está justificado *a priori* en creer que D conduce a la verdad (Goldman, 1989, p. 107, la traducción es mía).

El recurso *apriorista* al que recurre Bonjour, según Goldman (1989), parece problemático por lo siguiente:

a) Parece contraintuitivo, a juicio de muchos filósofos (especialmente los naturalizados), afirmar que, en última instancia, la

justificación empírica tenga que fundarse sobre la base de una justificación no empírica o *a priori*.

b) El recurso *apriorista* de Bonjour pospone dificultades en tanto que no habría motivo para pensar que la justificación *a priori* de los estándares de justificación empíricos no necesitan a su vez de una metajustificación. Bonjour se limita a afirmar que la justificación *a priori* no necesita de una justificación extra o una metajustificación que señale que la justificación *a priori* conduce a la verdad.

Por otra parte, y quizá el argumento más importante, es que la demanda de una justificación *a priori* de los estándares utilizados para justificar las creencias empíricas es demasiado exigente. Parece que muy pocos sujetos cuentan con creencias que, además de estar justificadas vía la coherencia con un sistema de creencias, puedan *a priori* saber que dicha coherencia conduce a la verdad. Los pocos sujetos que podrían satisfacer tales criterios serían aquellos que han tenido algún entrenamiento epistemológico. Bajo esta perspectiva, la postura internista de Bonjour lleva a la idea contraintuitiva de que los únicos sujetos que pueden ser llamados "conocedores" o "creyentes justificados" serían los filósofos. En otras palabras, Bonjour acepta que la justificación de una creencia para un sujeto cognoscente supone que éste debe creer justificadamente *a priori* que los estándares por los cuales su creencia está justificada conducen a la verdad, pero ello implica que muy pocas personas podrían estar realmente justificadas. De este modo, la caracterización coherentista de la justificación parece chocar con ciertas intuiciones, especialmente con la idea de que no sólo los sujetos epistemológicamente sofisticados poseen creencias justificadas. La violación de estas intuiciones ofrece razones suficientes, creen algunos filósofos, para abandonar el coherentismo (Goldman, 1980). Bonjour parece reconocer que de hecho cualquier teoría internista podría llegar a conclusiones como la suya cuando afirma lo siguiente:

> cualquier descripción no externista del conocimiento empírico, que tenga alguna plausibilidad, impondrá estándares de justificación que no pueden ser satisfechos, de manera completa o explícita, por muchos casos de conocimiento de sentido común. Así, bajo esta visión, tales creencias no serán *estrictamente hablando* instancias ade-

> cuadas de justificación y conocimiento (Bonjour, 1985, p. 53, la traducción es mía).

En suma, para un internista, si un sujeto *S* está justificado en creer que *p*, entonces *S* debe tener acceso interno al conjunto de razones que justifican su creencia en *p*. Para un internista como Bonjour, no es suficiente que *S* tenga acceso a las razones con que justifica su creencia que *p*, si *S* no está justificado en creer que dichas razones lo conducen a la verdad. Esta tesis aparece en el coherentismo de Bonjour —en un versión simplificada— como sigue: un sujeto *S* está justificado en creer que *p* si y sólo si *p* es miembro del sistema coherente de creencias de *S*, y *S* cree justificadamente *a priori* que toda creencia que sea miembro de un sistema altamente coherente es altamente probable de ser verdadera. Esta noción de justificación resulta problemática, entre otras cosas, porque casi ningún sujeto cognoscente podría cumplir los requisitos de justificación que estipula Bonjour. La justificación *a priori* de las metacreencias (i.e., aquella creencias de que las creencias que son miembros de sistemas coherentes son altamente probables de ser verdaderas) conducen a una noción bifurcada de la justificación, donde los estándares coherentistas se justifican *a priori*, mientras que las creencias empíricas (de primer orden) se justifican según los estándares coherentistas. Si lo dicho hasta aquí es correcto, se tienen razones para pensar que la postura coherentista de la justificación conduce a los problemas aquí expuestos a partir del compromiso internista de Bonjour. Es en este punto donde es posible exponer en qué consiste una teoría externista de la justificación.

Las teorías externistas de la justificación consideran que algunos factores que hacen posible la justificación de las creencias pueden hacer referencia a aspectos externos al sujeto, es decir, un externista sostiene que algunos de los requisitos de la justificación se encuentran fuera del acceso introspectivo por parte del sujeto epistémico, por ejemplo, aquellos concernientes al ambiente del creyente, así como a su historia y contexto social (Steup, 2005). Una manera particular de entender el externismo es que la justificación de las creencias de *S* depende del estado externo de cosas que hacen que su creencia sea causada por un *proceso cognitivo adecuado* [9]. Según un externista, puede haber propiedades que

justifiquen a ciertas creencias, tales que el creyente no podría jamás descubrirlas por mera reflexión (introspección o memoria). Si lo anterior es el caso, un sujeto puede estar justificado sin creer que de hecho está justificado. Sin ir más lejos en la descripción del externismo, en la siguiente sección expondré con cierto detalle una teoría externista particular, a saber, el confiabilismo. El confiabilismo es una teoría externista de la justificación que en gran medida representa las ideas centrales de este tipo de teorías, por lo que al presentarlo busco ejemplificar no sólo un tipo de teoría externista, sino también entender de mejor manera las críticas que se han hecho al externismo. Para ello me centraré en la versión confiabilista defendida por Alvin Goldman, pues es el filósofo que ha desarrollado y defendido con mayor ahínco una teoría confiabilista de la justificación [10].

I. 3. LA TEORÍA CONFIABILISTA DE LA JUSTIFICACIÓN

Antes de analizar en qué consiste el confiabilismo de Goldman, es necesario apuntar algunos de los presupuestos en los cuales descansa su noción de justificación. En primer lugar, Goldman (1986) distingue dos maneras de entender a las teorías de la justificación. Una manera de concebirlas es como aquellas que brindan un criterio para determinar si dada una creencia un sujeto está justificado o no en sostener tal creencia. En este sentido, las teorías de la justificación brindan un criterio que puedan usar los sujetos para decidir qué creer. Además de esta manera de entender a las teorías de la justificación, también se las puede entender como aquellas que brindan una descripción de las propiedades que tiene una creencia que está justificada. Goldman considera que la noción de justificación que él ofrece va en este segundo sentido, esto es, el confiabilismo no es un criterio que pueda ser usado por un sujeto en cualquier situación en la que desee saber si una creencia particular está justificada, sino una noción de justificación que simplemente explica cuál es el conjunto de propiedades que tienen las creencias justificadas.

Otro rasgo importante del confiabilismo que hay que destacar es que la noción de justificación debe ser definida a partir de nociones no epistemológicas, esto es, debe ser definida a través de términos que no tengan una carga epistemológica, pues real-

mente no se ofrece una definición de justificación si se lleva a cabo con términos que deban a su vez ser definidos. Por último, Goldman sostiene que la noción de justificación que desea defender debe dar cuenta de las intuiciones epistemológicas que tenemos al respecto.

> Sugiero que el significado del término "justificado" (en su uso epistémico) está sujeto a ciertas cosas que presumimos acerca del mundo, sobre si estamos en lo correcto o no (Goldman, 1986, p. 108, la traducción es mía).
>
> A diferencia de acercamientos tradicionales, no trato de prescribir estándares de justificación que difieran de, o mejoren a, nuestros estándares ordinarios. Intento meramente explicar los estándares ordinarios que son, creo yo, muy diferentes de muchas explicaciones clásicas, por ejemplo, la cartesiana (Goldman, 1979, p. 1, la traducción es mía).

Desde esta perspectiva, Goldman se compromete con una definición de justificación que no rebasa las barreras del análisis conceptual. Parece que a este nivel la composición del mundo es completamente irrelevante para la noción de justificación, pues basta con analizar las intuiciones de lo que significa estar justificado para descubrir cuál es la noción adecuada de justificación [11]. La propuesta que en este trabajo defiendo es que el análisis conceptual —y en especial aquel relacionado con la noción de justificación— puede ser enriquecido por los estudios empíricos sobre el razonamiento. El análisis que propongo a los largo de los siguientes capítulos entre la noción de justificación y racionalidad por medio de diversas teorías del razonamiento parte de la idea de que las contribuciones en psicología cognitiva son importantes para la epistemología.

El confiabilismo sostiene, en su versión más simple, que una creencia *p* está justificada para un sujeto *S* si y sólo si la formación de *p* se debe a un proceso (o conjunto de procesos) confiable de formación de creencias en un ambiente normal de formación de creencias (Goldman, 1979). Uno de los rasgos importantes en el confiabilismo es el énfasis en los procesos de formación de creencias o procesos cognitivos. Al confiabilista le interesan los procesos cognitivos tales como los buenos razonamientos, la introspección,

la percepción y la memoria, que parecen generalmente producir mayor número de creencias verdaderas que falsas. De ahí que a los procesos cognitivos que producen mayor número de creencias verdaderas que falsas los denominan procesos cognitivos "confiables". De este modo, para un confiabilista los razonamientos confusos o las simples corazonadas no son procesos cognitivos confiables, mientras que la introspección y la memoria, por mencionar algunos procesos cognitivos, pueden ser considerados como procesos confiables de formación de creencias. De esta manera, la confiabilidad consiste en el grado o la tendencia que tiene un proceso cognitivo de producir mayor número de creencias verdaderas que falsas [12].

Una vez que he señalado qué entienden los confiabilistas por procesos confiables de formación de creencias, es necesario entender qué es un proceso de formación de creencias en general. A grandes rasgos, un proceso cognitivo es un procedimiento que genera un mapeo de ciertos estados "insumos" o *inputs* a otros estados "resultados" o *outputs*. Los resultados de los procesos cognitivos serán siempre creencias, mientras que como insumos se podrán contar no sólo con creencias. Conforme a la naturaleza de los insumos en un sistema de creencias se puede distinguir entre aquellos procesos de formación de creencias cuyos insumos son creencias, como la memoria o el razonamiento deductivo, de aquellos procesos cuyos insumos no son creencias, sino sensaciones o estímulos, como ocurre en el caso de la percepción. Esta distinción es importante porque al confiabilista le interesa señalar cuándo un sistema de formación de creencias es confiable y si, a su vez, la confiabilidad de un proceso se debe a la tendencia a generar mayor número de creencias verdaderas que falsas, se podría argumentar que aquellos procesos de inferencia donde los insumos sean creencias, si éstas son generalmente falsas, producirán más creencias falsas que verdaderas. De este modo, los procesos de formación de creencias que tienen como insumos creencias no podrían ser considerados confiables. Para evitar este tipo de problemas, Goldman (1986) sostiene que en caso de que los insumos de un proceso cognitivo sean creencias, se dirá que los procesos de formación de creencias son "condicionalmente confiables". La confiabilidad de este tipo de procesos es condicional en tanto que estos procesos son *dependientes de las creencias* que

les sirven de insumo. A diferencia de estos procesos hay otros como la percepción, que son *independientes de creencias* en tanto que en sus insumos no aparecen creencias. Lo importante aquí es señalar que para que puedan producir creencias justificadas, los insumos-creencias de los procesos "condicionalmente confiables", deben estar previamente justificados, es decir, los insumos-creencias de este tipo de procesos deben ser producto de un proceso cognitivo confiable. Por ejemplo, en el caso de la memoria —en tanto proceso "dependiente de creencias"— será considerada como un proceso cognitivo confiable si y sólo si entre sus insumos se cuenta con creencias previamente justificadas.

Con ello se llega a una definición de justificación [13] más precisa por parte del confiabilismo:

a) S está justificado en creer que *p* en *t*, si *p* en *t* es resultado de un proceso cognitivo "independiente de creencias".
b) S está justificado en creer que *p* en *t*, si *p* en *t* es resultado de un proceso cognitivo "condicionalmente confiable" y las creencias de este proceso están ellas mismas justificadas.

¿En qué sentido la teoría confiabilista es una teoría externista? Debido a que la confiabilidad de un proceso cognitivo es parte dependiente del ambiente externo en el cual los sujetos cognoscentes están constantemente operando, en la justificación epistémica entran en juego, sostiene Goldman, no sólo los estados internos de un sujeto, como cree el internista, sino también los hechos y sucesos en el mundo. Goldman (1986) sostiene que para el confiabilismo una creencia puede estar justificada aun cuando el sujeto no crea que lo está, ya que lo que realmente justifica a cualquier creencia es que su origen causal se deba a un proceso cognitivo confiable. Es concebible que la creencia *p* de un sujeto *S* esté justificada sin que dicho sujeto lo sepa en un momento determinado, pero que posiblemente tiempo después pueda descubrirlo (lo cual parece compatible con el internismo accesibilista); no obstante, el confiabilista considera que es posible que la creencia *p* de *S* esté justificada y que *S* nunca pueda establecer que su creencia *p* está justificada.

I. 4. CRÍTICAS AL CONFIABILISMO

El argumento que expongo en esta sección en contra del confiabilismo de Goldman ha sido usado para rebatir la noción general de justificación ofrecida por el externismo. El argumento, que considero correcto, toma como punto de partida el rasgo externista de que un sujeto puede estar justificado sin saber que está justificado o creer justificadamente que lo está. *Grosso modo*, se argumenta que la noción de justificación tiene una estrecha relación con la noción de racionalidad y responsabilidad epistémica, por lo que parece contraintuitivo sostener que un sujeto esté justificado y al mismo tiempo no sea capaz de ofrecer razones para apoyar su creencia. Asimismo, se ha argumentado (Bonjour, 1985) que la tesis confiabilista conduce a una visión contraintuitiva del conocimiento, ya que, si se considera que una creencia está justificada sin que el sujeto *S* crea que lo esté, pero dicha creencia es verdadera, el confiabilismo concluiría que *S* sabe sin siquiera tener razones para ello. Con ello se violarían ciertas intuiciones acerca de la justificación y la definición de conocimiento que serían suficientes para mostrar que el confiabilismo no puede ofrecer una explicación satisfactoria de la justificación (Bonjour, 1985, Lehrer, 2000).

Bonjour (1980) ha recurrido a contraejemplos para mostrar cómo la noción externista de la justificación viola las intuiciones que deben estar detrás de cualquier definición de justificación. Estos contraejemplos apelan al supuesto poder de clarividencia cómo un proceso cognitivo, si bien hacer uso del poder de clarividencia resulta controversial, en la presente discusión este aspecto no parece tan importante, en tanto que basta con suponer que si la idea misma del conocimiento por clarividencia fuera posible, entonces una posición como la que Goldman defiende tendría que dar cuenta también de la clarividencia [14]. Uno de estos contraejemplos es el siguiente:

i. María, *M*, cree tener poderes de clarividencia.
ii. *M* no tiene razones para creer que tiene poderes de clarividencia.
iii. Los poderes de clarividencia de *M* son confiables.
iv. *M* cree que *p*: Felipe Calderón está en Palacio Nacional.

v. *M* sostiene *p* apelando a su poder de clarividencia, aun cuando existe evidencia (a través de la radio, la televisión, etc.) de que Calderón está en Los Pinos.
vi. Calderón está de hecho en el Palacio Nacional, la evidencia de lo contrario fue parte de un engaño oficial por una amenaza de terrorismo en su contra.

De donde,

M está justificada en creer que *p*, según el confiabilismo, y si se acepta la noción tradicional del conocimiento *M* sabe que *p*.

En este contraejemplo *M*, de acuerdo con la tesis confiabilista, está justificada en creer que *p*, pues *p* es el producto de un proceso cognitivo confiable, y dado que '*p*' es verdadera, entonces *M* sabe que *p*. Sin embargo, este resultado parece contraintuitivo, ya que *M* es irracional e irresponsable —epistemológicamente hablando— si aun cuando tiene motivos para no creer en su poder de clarividencia *M* confía en él. Esta razón bastaría, según Bonjour (1980), para mostrar que *M* no está justificada y por lo tanto la tesis confiabilista no ofrece una explicación adecuada de la justificación.

Después de analizar distintas versiones de su definición de justificación Goldman modifica su postura para poder dar cuenta del contraejemplo anterior. Su propuesta es:

S está justificado en creer que *p*, si la creencia *p* de *S* es resultado de un proceso confiable de formación de creencias y *no hay algún proceso cognitivo disponible a S*, tal que si hubiera sido usado por *S* además del proceso que de hecho fue usado, habría resultado que *S* no creyera que *p*.

Esta definición da cuenta del contraejemplo anterior, puesto que en él *M* tiene evidencia en contra de la creencia que sostiene, de modo que si hace uso de esa evidencia ella dejaría de creer lo que sostiene. En ese sentido, *M* no fue epistemológicamente responsable, en tanto que no hizo lo que epistemológicamente debió hacer.

No obstante, esta reformulación de la teoría confiabilista de la justificación todavía tiene algunos problemas. Uno de ellos se relaciona con el término "disponible", ¿qué significa que un proceso cognitivo sea "disponible" a un sujeto cognoscente [15]? Pero

dejando esta dificultad de lado, ¿puede esta reformulación del confiabilismo responder realmente a las críticas que señalan que el confiabilismo no da cuenta de la racionalidad y la responsabilidad epistémica que parecen estar detrás del concepto de justificación? Como mostraré a continuación, lejos de responder completamente a las críticas al externismo, es posible reformular el contraejemplo mostrando que hay todavía un sentido en el cual esta nueva versión del confiabilismo de Goldman no puede integrar las intuiciones que exigen los internistas (Bonjour, 1980). Este nuevo contraejemplo puede formularse como sigue:

i. Pedro, *P*, tiene poderes clarividentes.
ii. *P* no posee razones para afirmar que tiene poderes clarividentes, ni tampoco tiene razones para creer que existe la posibilidad de que los seres humanos puedan tener este tipo de poderes.
iii. Un día *P* cree sin ninguna razón aparente que *p*: Felipe Calderón está en Los Pinos.
iv. Los poderes de clarividencia de *P* son confiables.
v. La creencia *p* es producto del poder clarividente de *P*.
vi. '*p*' es verdadera.
De donde,

P está justificado en creer que *p*, según el confiabilismo, y si se acepta la noción tradicional del conocimiento *P* sabe que *p*.

En este contraejemplo, a diferencia del contraejemplo anterior, el sujeto cognoscente *P* no tiene acceso a un proceso cognitivo que señale la evidencia en contra de su creencia *p* como ocurría anteriormente. Por tal motivo, en tanto que la creencia *p* de *P* es producto de un proceso confiable de formación de creencias, y en tanto que *P* no dispone de otro proceso que al usarlo cambie su estado doxástico, entonces, según la reformulación del confiabilismo, *P* está justificado en creer que *p*, y, en tanto que es el caso que *p*, el confiabilista tiene que aceptar que *P* sabe que *p*. Nuevamente, para los críticos al externismo este resultado es contraintuitivo, pues de la irracionalidad de *P* con respecto a sus creencias, se sigue que *P* no puede estar justificado en creer que *p*, por lo que no se puede sostener que *P* sabe que *p*. Detrás de esta crítica, a mi juicio, se asume que la intuición importante detrás de la noción

de justificación es la de racionalidad, que en este caso *P* no parece satisfacer.

> Sin duda soy muy afortunado si mis creencias preceptuales e introspectivas están [producidas de manera confiable]... sin embargo, nada de esto tiene que ver con la cuestión de la justificación epistémica que se relaciona... con la racionalidad y la irracionalidad de las creencias del agente, en lugar de lo que parece ser, desde el punto de vista del creyente, una mera suerte cognitiva (Bonjour, 2003, p. 27, la traducción es mía).

Esta crítica parece apelar a la idea de que la creencia de un sujeto está justificada si la aceptación de dicha creencia por parte del sujeto fue racional y epistemológicamente responsable (Eraña, en proceso a). Además, esta crítica al externismo pone de manifiesto una idea común en epistemología, esto es, que la noción de justificación y la racionalidad se encuentran íntimamente unidas.

I. 5. DOS EJES DE DEBATE EN LA TEORÍA CONTEMPORÁNEA DE LA JUSTIFICACIÓN

De acuerdo con lo revisado en estas secciones, se puede sostener que existen dos ejes de debate que ordenan las distintas noción de justificación. El primer eje, fundacionismo-coherentismo, se puede entender como un cuestionamiento acerca de la naturaleza de las creencias justificadas y la dirección en que las creencias "trasmiten" la justificación. Para el fundacionismo: *i.)* existe un conjunto de creencias denominadas "básicas", cuya justificación no depende de otras creencias y las creencias no básicas, cuya justificación depende de las creencias básicas; y *ii.)* la justificación es unidireccional, esto es, va solamente de las creencias básicas a las creencias no básicas. Por su parte, el coherentismo sostiene que: *i.)* con respecto a la justificación, todas las creencias tienen el mismo estatus epistémico, i.e., no hay una distinción entre creencias básicas y creencias no básicas; y *ii.)* la justificación es holista.

El segundo eje de debate se centra no tanto en la naturaleza de las creencias justificadas sino en la relación entre el sujeto cognoscente y las propiedades o rasgos que hacen que una creencia esté justificada. Para el internismo, la justificación de las creencias de *S* depende de estados internos (reflexión, razonamiento o memo-

ria) a los cuales *S* tiene acceso inmediato. En tanto, para el externismo la justificación de las creencias de *S* depende del estado externo de cosas que hacen que su creencia sea producto de un proceso cognitivo adecuado o confiable [16], y entonces *S* no necesariamente tienen acceso a dichos estados. En otras palabras, para un internista siempre que *S* está justificado en creer que *p*, *S* cree justificadamente que lo está; por el contrario, para un externista *S* puede estar justificado en creer que *p*, aun cuando no crea que está justificado en creer que *p*.

Es posible imaginar un plano cartesiano en el que se establece un eje que va del fundacionismo al coherentismo y otro eje que va del internismo al externismo. Es preciso indicar que en gran medida el segundo eje de discusión se centra en la relación que el sujeto cognoscente tiene con sus creencias justificadas y no en cómo se trasmite la justificación de esas creencias, por lo que en este eje la racionalidad y el razonamiento parecen jugar un papel importante. Como mi propósito es indagar en qué medida las teorías del razonamiento pueden decir algo acerca de la noción de justificación, y pueden ayudar a entender la relación entre las nociones de justificación y de racionalidad, me centraré en el eje de debate entre las posturas internistas y externistas.

En este capítulo he expuesto con cierto detalle las posturas coherentista de Bonjour y confiabilista de Goldman, las cuales representan ejemplos del internismo y externismo respectivamente. En las conclusiones de este capítulo mostraré cómo es que estas dos teorías de la justificación sugieren que el estudio de la razón puede ayudar a formular una noción adecuada de la noción de justificación. Pero antes de exponer tales conclusiones, deseo subrayar que detrás del debate entre el internismo y el externismo se encuentran dos intuiciones de aquello que debe ser incluido en la noción de justificación. Veamos.

Ya que para el externista un sujeto puede estar justificado sin creer justificadamente que lo está, la principal crítica hecha a la noción de justificación externista es que no recoge lo que los seres humanos afirman tener cuando hablan de creencias justificadas, esto es, razones para sostener que se está justificado (Bonjour, 1980). Por su parte, ya que el internista considera que siempre que un sujeto está justificado éste cree [17] justificadamente que lo está, el principal argumento en contra del internismo es que su noción

de justificación haría que muchas de las creencias que los sujetos cognoscentes consideran justificadas en realidad no lo estén, viz., el criterio internista de justificación es demasiado estrecho y deja fuera muchas de las creencias que comúnmente se consideran justificadas (Goldman, 1980). Estas críticas pueden ilustrarse con base en una analogía con la ética, que ha sido usada tanto por los defensores del externismo como del internismo.

Para ejemplificar las críticas al externismo Bonjour (1985) presenta una interesante analogía con la ética con la que intenta señalar el problema al que se enfrenta cualquier posición externista. La perspectiva subjetiva (o la perspectiva de la primera persona [18]) de un sujeto moral puede ser completamente distinta a la perspectiva de un observador externo que tenga acceso a los hechos y consecuencias que están fuera del alcance de la perspectiva subjetiva. Supóngase que se defiende un tipo de externismo en la moral que sostenga que la justificación moral está determinada por una perspectiva externa independiente de la concepción del agente moral. Si se acepta una teoría utilitarista de este tipo, una acción estaría justificada moralmente si dicha acción tiene como resultado las mejores consecuencias objetivas, aun cuando la acción que el sujeto haya anticipado o planeado tuviera la intención de producir consecuencias distintas. Así, la acción de un sujeto *S* podría ser indeseable y extrema bajo su perspectiva, pero si bajo una perspectiva objetiva o desde la tercera persona tuviese consecuencias ventajosas, la acción de *S* estaría moralmente justificada. Sin embargo, parece difícil sostener que dada la irresponsabilidad moral de *S* —quien ha planeado consecuencias reprobables para la acción que comete— su acción esté justificada moralmente. Asimismo, si *S* no tuviera razones para creer que su acción traerá consigo las mejores consecuencias, pero su acción generará las mejores consecuencias *S* sería irresponsable moralmente, de ahí que su acción no estaría, según Bonjour (1985), moralmente justificada. Así como la violación de un deber moral, afirma Bonjour, supone que el sujeto que viola dicho deber es moralmente irresponsable y por ello incapaz de estar justificado moralmente, la violación de un deber epistémico supone que el sujeto que viola dicho deber es epistemológicamente irresponsable y, por lo tanto, no puede estar justificado epistemológicamente. Desde esta posición internista lo importante detrás de la

noción de justificación es la responsabilidad epistémica y la racionalidad, por lo cual, un acercamiento externista no puede brindar una noción adecuada de la justificación

Goldman (1980) también recurre a una analogía con la ética para ilustrar los problemas a los que conduce el internismo. Se puede distinguir en ética, afirma Goldman, entre una bondad subjetiva y una bondad objetiva (que también puede entenderse como una bondad vista desde la primera persona y una bondad vista desde la tercera persona). Se dice que un acto es objetivamente bueno sólo si satisface las condiciones de bondad moral, mientras que un acto es subjetivamente bueno si el agente moral cree que su acto y las consecuencias de éste son objetivamente buenas. El externismo resulta, según Goldman (1980), ser una teoría análoga a la concepción de bondad objetiva, en tanto que considera que la justificación de una creencia debe satisfacer condiciones objetivas de justificación —como son las relaciones externas al sujeto cognoscente— mientras que el internismo parece ser una teoría análoga a la concepción de bondad subjetiva pues en él intervienen las creencias de un sujeto acerca de cuál es el criterio correcto de justificación. Puesto que distintos sujetos consideran correctos distintos criterios, y la bondad subjetiva y el internismo consideran que todo lo que se debe tener en cuenta son las creencias de los sujetos acerca de cuál es el criterio correcto, el problema central con la noción de bondad subjetiva así como con el internismo, afirma Goldman, es que conducen a un relativismo ético o epistemológico, según sea el caso. Para evitar este relativismo, Bonjour (1985) sostiene que los sujetos deben estar justificados *a priori* acerca de los criterios con los cuales justifican sus creencias, esto es, que tales criterios conducen a la verdad. Sin embargo, como se revisó anteriormente, esta propuesta implicaría que casi ningún sujeto estaría genuinamente justificado, pues casi ningún sujeto es consciente de los estándares que justifican sus creencias, como tampoco están justificados en creer que dichos estándares conducen a la verdad. La noción de justificación, desde la perspectiva externista, debe recoger la intuición de que la justificación depende de los criterios externos adecuados y no de lo que cada sujeto crea que es el criterio correcto para justificar una creencia.

Así, detrás del internismo y el externismo se encuentran dos modos distintos de entender la noción de justificación, que a su vez responden a intuiciones epistémicas distintas. Los defensores del internismo consideran que la noción de justificación debe recoger la intuición de la responsabilidad epistémica y la racionalidad, es decir, de dar y ofrecer razones. Por su parte, los defensores del externismo consideran que la genuina intuición detrás de la noción de justificación es aquella que remite finalmente a los procesos y métodos correctos de formación de creencias, como son la percepción visual y los procesos de razonamiento inductivo y deductivo, en tanto que éstos conducen a metas epistemológicas como la verdad. Si bien las dos intuiciones que consideran los defensores del internismo y externismo son distintas, las intuiciones que defienden estas dos posturas comparten una referencia a la racionalidad y el razonamiento: el internismo, al enfatizar el papel de "dar y ofrecer razones" como esencial a la noción de justificación, y el externismo al destacar la idea de que los procesos cognitivos adecuados (como lo es el razonamiento) deben estar contenidos en la noción de justificación.

I. 6. CONCLUSIONES

A lo largo de este capítulo se ha visto que no existe una única noción de justificación epistémica que haya sido aceptada por los filósofos. Existen varias nociones de justificación que intentan dar respuesta a qué significa que una creencia esté justificada. De las diversas teorías de la justificación expuestas aquí, se puede distinguir un primer eje de debate que responde a la preguntas sobre el tipo de creencias que están justificadas y sobre cómo se trasmite la justificación. Dentro de este primer eje de debate se encuentran las teorías fundacionistas y coherentistas. El segundo eje de debate en la teoría de la justificación busca principalmente responder a la pregunta por la relación que tiene un sujeto cognoscente con sus creencias justificadas o con los rasgos que hacen que una creencia esté justificada. Dentro de este segundo eje de debate se encuentran las teorías internistas y externistas. Dado que mi interés central en este libro es conocer los vínculos que hay entre las nociones de justificación y de racionalidad, me centraré en la relación que este segundo eje de debate tiene con la racionalidad.

Pero antes de sugerir cuál es la naturaleza de esta relación es preciso repasar otro punto central de este capítulo.

Detrás del debate entre el internismo y el externismo, he argumentado, se encuentran dos modos distintos de entender la noción de justificación, que a su vez responden a intuiciones epistémicas distintas. Los internistas consideran que la noción de justificación debe recoger la intuición de la responsabilidad epistémica y la racionalidad, es decir, de dar y ofrecer razones. Por su parte, los externistas consideran que la genuina intuición detrás de la noción de justificación tiene que ver con la importancia que juegan los procesos y métodos correctos de formación de creencias, que generalmente remiten a la verdad (o probable verdad) de las creencias en cuestión; dentro de dichos procesos se encuentran los procesos de razonamiento inductivos y deductivos. El debate acerca de la noción de justificación entre internistas y externistas parece atrincherarse en un debate acerca de intuiciones epistemológicas distintas donde, al menos en principio, no habría criterio o estándar para preferir alguna de las intuiciones defendidas. Considero que habría una lección importante en lo revisado en este capítulo: no existe consenso acerca de la intuición o conjunto de intuiciones que debe recoger la noción de justificación epistémica. Dado que no existe un único conjunto de intuiciones que sirva para seleccionar cuál es la noción de justificación más adecuada, en este libro propongo un criterio para seleccionar una noción de justificación *fuera* del debate que hasta ahora se ha llevado a cabo en el análisis epistemológico contemporáneo.

Si bien las dos intuiciones que consideran los defensores del internismo y externismo son distintas, las intuiciones que defienden estas dos posturas comparten una referencia, de manera directa o indirecta, a las nociones de racionalidad y razonamiento. Por parte de los internistas, al sostener que el criterio que está detrás de su noción de justificación es el de racionalidad; mientras que al externismo le interesa los procesos cognitivos de formación de creencias, donde los procesos de razonamiento juega un papel primordial. Lo anterior es el motivo por el cual el estudio de la razón puede ser pertinente para abordar el debate epistemológico acerca de la noción de justificación.

En la literatura epistemológica se ha hecho, de modo explícito o implícito, una conexión "natural" entre la noción de justificación

y racionalidad. Como mencioné en la Introducción, comúnmente se considera que una creencia justificada es lo mismo que una creencia racional. Muchas veces se ha dado por hecho que una creencia es racional si satisface los criterios de una teoría de la racionalidad. Del mismo modo, una creencia está justificada si satisface los criterios determinados por una teoría de la justificación. Por ello, a mi juicio, el debate en teoría de la justificación debe incorporar un estudio más amplio que involucre la racionalidad y el razonamiento. Esta idea parece reforzarse por el hecho de que las nociones de justificación y de racionalidad son fundamentalmente términos normativos, en tanto que evalúan creencias que comparten cierta propiedad o conjunto de propiedades. Debido a que la racionalidad y la noción de justificación son términos normativos, el estudio de los orígenes y fundamentos de la normatividad que abordaré posteriormente puede arrojar luz a su función normativa y a la relación que hay entre racionalidad y justificación.

Otro motivo por el cual parece importante el estudio de la razón se funda en la opinión común de que el debate entre internismo y externismo, y en general entre las distintas teorías de la justificación, ha llegado a una encrucijada entre dos intuiciones epistémicas distintas (King, 2000). Es posible que el estudio del razonamiento y la racionalidad ilumine y enriquezca parte del debate contemporáneo en teorías de la justificación, lo cual ayude a salir del atolladero originado de la defensa de nociones de justificación comprometidas con intuiciones epistemológicas distintas. Si la interpretación que se hace del debate en torno a la noción de justificación en este capítulo es correcta, el considerar que el estudio de la razón puede *iluminar* las discusiones en torno a la noción de justificación trae consigo diversos problemas. Uno de ellos, quizá el más inmediato, es qué se debe entender por razonamiento y racionalidad, y cómo se deben estudiar. Con respecto a esto último, es preciso indicar que asumiendo el supuesto naturalista de que los estudios empíricos deben ser tomados en cuenta dentro de las indagaciones filosóficas, en los siguientes capítulos exploraré distintas teorías del razonamiento defendidas por la psicología cognitiva en los últimos años.

Este acercamiento naturalista no ha sido el más estudiado en epistemología, ya que tanto internistas como externistas han su-

puesto que el esclarecimiento de la noción de la justificación se reduce al mero análisis conceptual. Como he mencionado anteriormente, los diversos teóricos de la justificación revisados en este capítulo suscriben el principio analítico, según el cual, "los problemas filosóficos pueden ser tratados, o reducidos, a problemas acerca de las expresiones lingüísticas del pensamiento" (Cohen, 1986, p. 8, la traducción es mía). Incluso Goldman, quien ha defendido explícitamente una postura naturalista, sostuvo que su propuesta "es hecha en el espíritu de tratar de elucidar la concepción ordinaria de justificación" (1986, p. 109, la traducción es mía). Parte de lo que deseo mostrar en este libro es que el estudio del razonamiento ayuda a *i.*) evaluar diferentes nociones de justificación, así como a *ii.)* entender la relación entre la noción de justificación y racionalidad. A lo largo de este trabajo indagaré las implicaciones que los estudios empíricos sobre cognición tienen en la epistemología. Como ya lo he mencionado, al final del libro defenderé que existe una teoría del razonamiento humano que, por el momento, se encuentra *por encima* de otras teorías psicológicas del razonamiento, y sostendré que dicha teoría apoya una noción de justificación de corte externista.

II.
LA TRADICIÓN DE HEURÍSTICA Y SESGO Y LA RACIONALIDAD

Uno de los propósitos de este trabajo es conocer las implicaciones de las teorías del razonamiento sobre el debate entre las nociones internista y externista de la justificación. En este capítulo indagaré si una teoría psicológica particular, i.e., la tradición de heurística y sesgo, ayuda a entender qué significa que una creencia esté justificada. En el capítulo anterior sostuve que las intuiciones defendidas por los internistas y externistas hacen referencia a la razón. Los internistas hacen lo anterior al defender que la noción de justificación debe recoger la noción de "dar y ofrecer" razones; mientras que los externistas sostienen que la noción de justificación debe recoger la conexión entre las creencias y los procesos cognitivos adecuados en donde el razonamiento juega un papel importante. Si bien las dos intuiciones hacen referencia a la razón, es importante distinguir claramente entre el razonamiento y la racionalidad, y en general entre las teorías del razonamiento y las teorías de la racionalidad. Aquí se hace dicha distinción, presentando y discutiendo parte de la evidencia empírica que, según cierta interpretación, sugiere que los seres humanos son irracionales. No sólo mostraré el alcance de esta interpretación y el tipo de presupuestos que hacen los defensores de dicha postura, sino las implicaciones que tal postura tiene al debate en torno a la noción de justificación epistémica.

La estructura del capítulo es la siguiente. En una primera sección se distinguen diversas maneras de estudiar a la razón. En dicha sección presento una herramienta de análisis para distinguir a las teorías del razonamiento de las teorías de la racionalidad. De entre los primeros trabajos en psicología cognitiva del razonamiento se encuentran los llevados a cabo por la tradición de heurística y sesgo, los cuales a partir de la década de los setenta

han sido el fundamento de posteriores discusiones en psicología del razonamiento. Por tal motivo, después de hacer ciertas distinciones entre las teorías del razonamiento y las teorías de la racionalidad, expongo algunos de los experimentos hechos por la tradición de heurística y sesgo. Comúnmente, la interpretación de muchos de los experimentos hechos por los defensores de esta tradición psicológica es de tipo pesimista, pues se considera que tales experimentos señalan que los seres humanos son sistemáticamente irracionales. Sin embargo, se pueden distinguir distintas versiones de esta interpretación, las cuales expondré. En una de las últimas secciones mostraré que la interpretación de los experimentos está unida a una visión de racionalidad particular. Finalmente revisaré algunas de las implicaciones epistemológicas de estas interpretaciones, centrándome en la racionalidad y las nociones internistas y externistas de la justificación revisadas en el primer capítulo de este libro.

II.1. TRES MODOS DE ACERCARSE AL ESTUDIO DE LA RAZÓN

Grosso modo se puede entender el estudio de la razón como simplemente el estudio del razonamiento humano, esto es, como el estudio de los procesos y mecanismos que llevan de un conjunto de información a una respuesta de un cierto tipo (Damasio, 2003). El razonamiento está unido íntimamente con la toma de decisiones; han creído algunos psicólogos, filósofos y científicos cognitivos que su propósito es la decisión, y decidir significa elegir una acción, una respuesta verbal, o la combinación de lo anterior en un momento determinado. Sin embargo, en muchas ocasiones los seres humanos llevan a cabo acciones o respuestas que no parecen encontrarse en esta descripción del razonamiento, digamos, cuando no se tiene un conocimiento de las posibles consecuencias de una acción o bien la estrategia de inferencia no es consciente. Por ejemplo, cuando un sujeto se aparta bruscamente para esquivar un objeto que cae, parece que la respuesta es automática y rápida, sin esfuerzo de deliberación. Algunas veces se considera que este tipo de respuesta no es parte del razonamiento, aunque hay teorías que han integrado a estas respuestas dentro de procesos de razonamiento que no son explícitos (Evans y Over, 1996). Para englobar las diversas maneras de entender el

razonamiento, entenderé por una teoría del razonamiento una teoría descriptiva que explica cómo es que los seres humanos de hecho razonan y toman decisiones, independientemente de que ello conlleve un tipo de deliberación explícita por parte del sujeto.

Por otra parte, si la noción de razonamiento se puede entender de este modo laxo, hay un modo de acercarse al estudio de la razón que no se reduce al cómo los seres humanos razonan. Las nociones normativas de "racional" o "irracional" conducen a otro modo de estudiar a la razón. El motivo para afirmar lo anterior es que el estudio de la razón toma una carga evaluativa, en tanto que alguien es racional si normalmente razona y actúa como debe hacerlo —esto es, según ciertos *criterios* epistémicos. Este nivel evaluativo del estudio de la razón parte de la manera en cómo los seres humanos razonan, contrastándolo con un conjunto de normas de lo que significa razonar correctamente. Si, por ejemplo, alguien sostiene que "la decisión de *S* es irracional", se está evaluando cómo dado el modo en que *S* tomó la decisión, ésta no cumple con un estándar de lo que es tomar una decisión correctamente. Así, el estudio de la razón puede entenderse como el resultado de contrastar el modo en que se razona con el estándar o conjunto de estándares que determinan qué significa razonar correctamente. A este estudio de la razón lo denominaré el estudio evaluativo de la razón.

¿De dónde se originan los criterios bajo los cuales una teoría evaluativa de la razón determina si tales criterios son satisfechos o no? La respuesta a esta pregunta conduce a un estudio de la razón distinta a las que ya he mencionado, esto es, a una teoría normativa de la racionalidad. Una teoría normativa de la racionalidad será aquella teoría que brinde una explicación de los orígenes y criterios normativos usados para evaluar el razonamiento. Desde esta perspectiva, lo importante no es cómo los sujetos de hecho razonan, o si su razonamiento satisface los criterios de lo que es racional, sino la constitución misma de los criterios de razonamiento adecuados. Bajo esta perspectiva, al menos en cierta medida, es posible distinguir la pregunta, acerca de cuáles son las normas con las que se debe razonar, de las preguntas "¿satisface el razonamiento humano las normas con las cuales se debe razonar?" o, "¿cómo de hecho razona el ser humano?".

En suma, distinguiré, al igual que lo han hecho otros filósofos (Samuels, Stich y Faucher, 2004), tres distintas maneras de estudiar a la razón: los estudios descriptivo, normativo y evaluativo. El primero consiste en describir, generalmente de modo empírico, cómo es que el ser humano de hecho razona, así como descubrir los mecanismos y procesos psicológicos que subyacen en los patrones de razonamiento observados. Este tipo de estudio es llevado a cabo principalmente por psicólogos cognitivos, antropólogos cognitivos y científicos computacionales. Aquí se ubican las distintas teorías del razonamiento defendidas por psicólogos cognitivos que expondré posteriormente. El estudio normativo de la razón intenta dar cuenta, a diferencia del estudio descriptivo, no tanto de cómo los seres humanos razonan, sino en cómo deberían razonar. Una de las metas centrales en este tipo de estudios es descubrir las reglas o principios que indiquen qué es razonar correctamente o racionalmente. Aquí se encuentran las distintas teorías de la racionalidad. Por último, el estudio evaluativo de la razón busca determinar el alcance que tiene el razonamiento humano de acuerdo con un determinado estándar normativo. Dados ciertos criterios normativos, quienes realizan un estudio evaluativo de la razón determinan cuándo el razonamiento cumple con tales criterios. Por ejemplo, el proyecto evaluativo de la razón podría indicar cuándo el ser humano ha actuado racionalmente con respecto a ciertas normas y principios.

Cada uno de los resultados o descubrimientos provenientes de estos tres distintos estudios de la razón se pueden relacionar con los otros estudios. Por ejemplo, cuando se evalúan los procedimientos usados por un individuo se parte ya de un conjunto de normas que se contrastan con una descripción de lo que el individuo realmente hace, cognitivamente hablando [1]. Del mismo modo, se ha argumentado que la manera en que el ser humano procede en el ámbito cognitivo tiene repercusiones en el modo en que se puede evaluar sus procedimiento cognitivos, e incluso, aunque ampliamente debatido, se ha sostenido que las normas pueden adquirir validez o invalidez a partir del modo en que queda descrito el proceder cognitivo de los sujetos [2]. Muchas veces, como se verá a lo largo de este trabajo, los distintos estudios de la razón están tan íntimamente ligados que resulta difícil trazar líneas claras entre éstos. Sin embargo, conservaré está distinción

para entender con mayor claridad algunas de las posiciones que serán presentadas posteriormente. La distinción entre el nivel descriptivo, normativo y evaluativo de la razón servirá como una herramienta de análisis para entender mejor las teorías psicológicas estudiadas en este libro, y dejará en claro que pocas veces los psicólogos cognitivos hacen uso explícito de tal herramienta de análisis.

Uno de los debates en donde confluyen los tres estudios de la razón arriba descritos surge de una serie de resultados empíricos provenientes de la psicología cognitiva del razonamiento a mediados de los setenta. Estos resultados, según muchos psicólogos y filósofos, apoyan una visión pesimista de la racionalidad humana (Piatelli-Palmarini, 2005). Los resultados de estos estudios, según Kahneman, Tversky y otros psicólogos, reportan que bajo ciertas circunstancias

> las personas no parecen seguir el cálculo de probabilidad o la teoría estadística de predicción. En lugar de ello, confían en un número limitado de heurísticas que algunas veces producen juicios razonables y otras veces conducen a severos y sistemáticos errores (Kahneman y Tversky, 1982b, p. 48, la traducción es mía).

Así pues, ¿en qué cosiste realmente esta evidencia que muchos filósofos, psicólogos y otros teóricos consideran la base empírica de muchas preocupaciones acerca de la racionalidad humana?

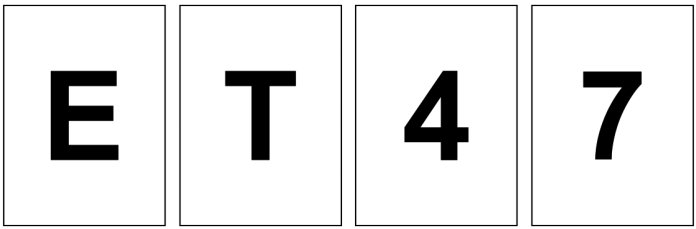
II.2. LA INQUIETANTE EVIDENCIA EMPÍRICA ACERCA DE CÓMO RAZONA EL SER HUMANO

A continuación describiré algunos de los experimentos más importantes que apoyan, a juicio de algunos filósofos (Piatelli-Palmarini, 2005), la tesis de que el ser humano no satisface apropiadamente algunos criterios normativos de razonamiento.

II.2.1. LAS TAREAS DE SELECCIÓN.

Peter Wason (1966) publicó un artículo influyente donde aparece un conjunto de problemas de razonamiento que se conocieron posteriormente como "tareas de selección" o "tareas de selección de tarjetas". Cabe señalar que este problema de razonamiento ha sido uno de los más intensamente investigados dentro de la

psicología del razonamiento (Evans, 1991). En un caso típico de tarea de selección se les presenta a los sujetos cuatro cartas como aparecen a continuación y posteriormente se explica lo que deben realizar.



> Todas las cartas anteriores tienen de un lado un número y del otro una letra del abecedario. Tu trabajo es determinar cuáles de estas cartas tienes que voltear para probar la verdad de la oración: si hay una vocal en un lado de la carta, entonces hay un número par en su otro lado.

Wason y otros psicólogos descubrieron que comúnmente lo sujetos no responden correctamente a este tipo de tareas, pues, aun cuando la respuesta correcta son las cartas E y 7, la mayoría de sujetos sólo consideran que se debe voltear la carta E, o las cartas E y 4. Desde estos primeros estudios de Wason se ha multiplicado el número de experimentos con tareas de selección que varían sustancialmente en sus dimensiones. En distintas versiones de estos experimentos los sujetos han podido realizar de mejor manera algunas tareas de selección. El resultado correcto de los sujetos a pruebas similares a las tareas de selección es desconcertante, pues no parece haber un rasgo o conjunto de rasgos que separen las versiones en los cuales los sujetos realizan correctamente este tipo de tareas, mientras que en otras les resulta muy complicado realizarlas [3]. Se ha sugerido, por ejemplo, que los sujetos hacen un mejor trabajo en problemas más realistas o que tienen mayor familiaridad con los sujetos; sin embargo, hay evidencia que parece poner en duda estas sugerencias (Stich, 1990, Stein, 1996). Los resultados obtenidos en las tareas de selección

han sido la base de distintas teorías que tratan de entender los procesos cognitivos que subyacen a este tipo de razonamiento, aun cuando no hay un consenso acerca de la naturaleza de tales procesos y de los mecanismos que subyacen a tales procesos.

II.2.2. LA FALACIA DE LA CONJUNCIÓN

Otro resultado de investigaciones empíricas que señalan una aparente desviación de los estándares normativos de inferencia se centra en el modo en que los sujetos evalúan la probabilidad de estados o sucesos lógicamente relacionados. Es bien sabido que dentro de la teoría de la probabilidad, la probabilidad de un estado de cosas o suceso compuesto es menor o igual a la probabilidad de los componentes de estados o sucesos. Entre los experimentos mejor conocidos de este tipo se encuentra el problema de la conjunción o falacia de la conjunción. En sus famosos experimentos, Kahneman y Tversky (1982a) presentan a los sujetos la siguiente prueba:

> Linda tiene 31 años, es soltera, abierta y muy brillante. Ella estudió filosofía y como estudiante estuvo profundamente preocupada por asuntos de discriminación y justicia social, y también participó en manifestaciones antinucleares.
>
> De las siguientes oraciones determina cuál es más probable escribiendo un 1 para la más probable y un 8 para la menos probable.
> (a) Linda es maestra de primaria.
> (b) Linda trabaja en una librería y toma clases de yoga.
> (c) Linda es activista en un movimiento feminista.
> (d) Linda es una trabajadora social en el área psiquiátrica.
> (e) Linda es miembro de la Liga de Mujeres a favor del Voto.
> (f) Linda es cajera en un banco.
> (g) Linda es una vendedora de seguros.
> (h) Linda es cajera en un banco y activista en un movimiento feminista.

En un grupo de personas sin ningún tipo de estudios en probabilidad y estadística, 89 por ciento consideraron que el enunciado (h) tiene mayor probabilidad que el enunciado (f), a pesar de que ser "cajera y feminista" es menos probable que ser solamente cajera (Stein, 1996). Esta misma prueba, según Samuels, Stich y

Faucher (2004), se realizó a sujetos entrenados en estadística y el 85 por ciento de sujetos cometieron el mismo error. Resultados de este tipo parecen indicar que independientemente de los conocimientos de probabilidad que adquieren los sujetos, éstos consideran que es más probable un suceso compuesto que cada uno de sus componentes.

II.2.3. EL RECHAZO A LA "RAZÓN DE BASE"

Otro conjunto de estudios consiste en la manera en que los sujetos usan la información ofrecida por la "razón de base" para formar juicios probabilísticos. De acuerdo con una descripción bayesiana, la probabilidad de una hipótesis de un cierto cuerpo de evidencia depende, en parte, de la probabilidad previa (*prior probability*) de cada una de las hipótesis. En una serie de experimentos, Kahneman y Tversky mostraron que los sujetos frecuentemente le dan poca importancia a las probabilidades previas. Uno de estos experimentos es el siguiente (Kahneman y Tversky, 1982a):

> A un primer grupo de sujetos se les comenta que un conjunto de psicólogos entrevistó y administró exámenes de personalidad a 30 ingenieros y 70 abogados, todos ellos exitosos en sus respectivos campos de trabajo. Sobre la base de esta información han sido escritas 30 descripciones de ingenieros y 70 de abogados. Se les proporciona cinco de esas descripciones en un rango de 100 de las descripciones disponibles y su tarea es indicar la probabilidad de que una persona sea un ingeniero sobre una escala de 0 a 100. A otro grupo de sujetos se les presentó el mismo texto excepto que la razón de base fue intercambiada, es decir, los exámenes de personalidad habían sido administrados a 70 ingenieros y 30 abogados.

Algunas de las descripciones proporcionadas a los sujetos estaban diseñadas para ser compatibles con el estereotipo de los ingenieros, pero no con el estereotipo de los abogados, mientras que otras descripciones fueron diseñadas para coincidir con el estereotipo de los abogados pero no con el de los ingenieros; por último, otras descripciones fueron neutrales, esto es, no ofrecían información relevante que pudiera usarse para que los sujetos tomaran una decisión. Dos ejemplos de las descripciones son:

> Jack tiene 45 años, está casado y tiene cuatro hijos, es conservador, cuidadoso y ambicioso. Jack no muestra interés en cuestiones políticas o sociales y gasta la mayor parte de su tiempo libre en *hobbies* como la carpintería, navegación y la resolución de enigmas matemáticos.
>
> Dick tiene 30 años, está casado pero no tiene hijos, es un hombre con muchas habilidades y alta motivación, lo que lo hace una persona exitosa en su campo de trabajo. Dick es muy bien visto por sus colegas.

Como era de esperarse, los sujetos de los dos grupos consideraron que es altamente probable que Jack fuese ingeniero. La diferencia en las descripciones no pareció tener ningún efecto entre los dos grupos de sujetos a los que se les aplicó el experimento. El rechazo de la información ofrecida en la "razón de base" fue aún más sorprendente en el caso de Dick, pues dado que la descripción proporcionada fue totalmente neutral con respecto a su profesión, la única información útil que los sujetos tenían era la "razón de base". Sin embargo, dicha información fue totalmente ignorada, y ambos grupos de sujetos consideraron que la probabilidad de que Dick fuera ingeniero o abogado era de 50 por ciento.

Se considera que el problema del rechazo de la información que proporciona la "razón de base" es un fenómeno que puede acarrear serias consecuencias prácticas. Uno de estos problemas ha sido presentado por Casscells, et. al. (1978) a un grupo de profesores, sus colaboradores y un grupo de estudiantes avanzados de la *Harvard Medical School*. La prueba realizada fue la siguiente:

> Si una prueba para detectar alguna enfermedad cuya prevalencia es 1/1000 tiene tasa de 5 por ciento de obtener un falso positivo, ¿cuál es la probabilidad de que una persona que tenga un resultado positivo de hecho tenga dicha enfermedad, asumiendo que no se sabe nada acerca de los síntomas o signos de dicha persona? (responder al porcentaje.)

Bajo la interpretación más plausible del problema, la respuesta correcta es del 2 por ciento. Pero según los resultados de Casscells, et. al. (1978) sólo 18 por ciento de las personas encuestadas dieron una respuesta cercana al 2 por ciento; 45 por ciento de este grupo de médicos ignoró la "razón de base" y sostuvo que la respuesta era 95 por ciento.

II.2.4. TOMA DE DECISIONES

Hay un conjunto de estudios en psicología cognitiva centrado en la toma de decisiones, muchos de ellos parecen mostrar que la toma de decisiones es normativamente problemática. En un experimento de Kahneman y Tversky (1982a) se solicitó a un grupo de personas lo siguiente:

> Imagina que el gobierno de Estado Unidos prepara una salida inusual a una enfermedad asiática que se espera mate a 600 personas. Hay dos programas alternativos propuestos para combatir esta enfermedad. Presuponiendo que las propuestas tienen el mismo fundamento científico, las consecuencias de los programas son:
> Si se adopta el programa A, 200 personas sobrevivirán.
> Si se adopta el programa B, hay una probabilidad de 1/3 de que las 600 personas pueden sobrevivir, y una probabilidad de 2/3 de que ninguna persona sobreviva.

A un segundo grupo de sujetos les fue aplicado el problema idéntico, sólo que en este caso los programas fueron descritos como sigue:

> Si se adopta el programa C, 400 personas morirán.
> Si se adopta el programa D, hay una probabilidad de que 1/3 que nadie morirá, y una probabilidad de 2/3 de que las 600 personas morirán.

En la primera versión del problema la mayoría de los sujetos escoge el programa A. Pero en la segunda versión la mayoría de los sujetos escoge el programa D, a pesar del hecho de que el resultado descrito en A es idéntico al resultado descrito en C. La implicación desconcertante de esta investigación es que las decisiones tomadas por los sujetos están fuertemente influenciadas por la manera en que las opciones son descritas o "enmarcadas" (Samuels y Stich, 2004).

II.2.5. AJUSTE Y ANCLAJE

Kahneman y Tversky mostraron que los procesos de razonamiento cuantitativos podrían ser grandemente influenciados por los valores que son tomados como punto de partida (Kahneman y Tversky, 1982a). Estos psicólogos denominaron a este fenómeno "anclaje". En uno de estos experimentos se les solicitó a los sujetos que estimaran rápidamente el producto de expresiones numéri-

cas. A un grupo de sujetos se les dio cinco segundos para estimar el producto de:

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

mientras que a un segundo grupo de sujetos les fue proporcionado el mismo tiempo para estimar el producto de:

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

Bajo estas restricciones de tiempo la mayoría de los sujetos pudieron hacer sólo algunos pasos para el cómputo y tuvieron que extrapolar o ajustar sus resultados. Kahneman y Tversky (1982a) pronosticaron que debido a que los ajustes eran insuficientes, el procedimiento conduciría a una subestimación. Del mismo modo, como el resultado del primer paso en la secuencia descendente es más alto que en la secuencia ascendente se pronosticó que los sujetos producirían estimaciones más altas en el primer caso que en el segundo. Ambos pronósticos fueron confirmados. La estimación media para la secuencia descendente fue de 2 250, mientras que la media para la secuencia ascendente fue sólo de 512. Además los dos grupos sistemáticamente subestimaron el valor de las expresiones numéricas siendo la respuesta correcta 40 320.

Se podría pensar que el experimento anterior no es inquietante, ya que les fue impuesto a los sujetos restricciones de tiempo para realizar los cómputos solicitados. Sin embargo, otros ejemplos de "anclaje" resultaron preocupantes aun sin restricciones de tiempo. En otro experimento Kahneman y Tversky (1982a) solicitaron a los sujetos estimar el número de países africanos dentro de la ONU, pero antes de realizar esta tarea se les mostró a los sujetos un número arbitrario que fue determinado por las vueltas de una ruleta en su presencia. Algunos se les mostró el número 65, mientras a otros el número 10, solicitándoles si la estimación era correcta o incorrecta con relación al número indicado en la rueda y que indicaran cuál sería el porcentaje real de países africanos en la ONU. La estimación media fue 45 por ciento para aquellos sujetos cuyo número de anclaje fue 65 y 25 por ciento para aquellos cuyo número de anclaje fue 10. La desconcertante implicación de este experimento es que las estimaciones de los sujetos

pueden ser afectadas sustancialmente por un valor numérico "anclado", aun cuando dichos sujetos estén completamente conscientes de que el número de anclaje ha sido generado por un proceso aleatorio del cual están seguros que es irrelevante para la tarea en cuestión.

II. 3. LA INTERPRETACIÓN PESIMISTA DE LA EVIDENCIA EMPÍRICA

El resultado de los experimentos arriba descritos son para muchos filósofos y estudiosos de la cognición intrínsecamente inquietantes (Piatelli-Palmarini, 2005). Muchos de estos experimentos se han comparado con las ilusiones ópticas, porque los sujetos entienden cuál es el tipo de desviación que cometen en los experimentos y, no obstante, siguen cometiendo los mismos errores o falacias cuando se vuelven a enfrentar a dichos experimentos. De la misma manera en que los sujetos continúan viendo las mismas ilusiones ópticas a pesar de que se les ha explicado el motivo de la ilusión, los sujetos siguen cometiendo los mismo errores en las pruebas de razonamiento (Sloman, 1996). De este modo, en analogía con las ilusiones ópticas, los errores sistemáticos de razonamiento, como la falacia de la conjunción, se les ha denominado *ilusiones cognitivas* (Gigerenzer, 1993).

Kahneman y Tversky inauguran lo que comúnmente se denomina la tradición o el programa de "heurística y sesgo", el cual consiste en estudiar e investigar las ilusiones cognitivas así como sus explicaciones. Haciendo uso de la distinción hecha al inicio de este capítulo, se puede sostener que la tradición de heurística y sesgo es principalmente una teoría descriptiva del razonamiento que intenta describir y explicar la naturaleza de las ilusiones cognitivas [4]. En sus inicios, la tradición de heurística y sesgo se centró en estudiar el razonamiento probabilístico, pero posteriormente su objeto de estudio fue ampliándose a otros tipos de razonamiento. Según esta tradición, los sujetos evalúan la probabilidad de un evento incierto haciendo uso de "principios heurísticos que reducen las tareas complejas de evaluar probabilidades y predecir valores a simples operaciones de juicio" (Kahneman y Tversky, 1982, p. 3, la traducción es mía). A grandes rasgos, Kahneman y Tversky sostienen que dichas heurísticas son muy útiles, pero algunas veces conducen a errores sistemáticos.

La confianza en estas heurísticas conduce comúnmente a sesgos encontrados en juicios intuitivos acerca de la probabilidad y en el razonamiento deductivo. En el libro *Judgment Under Uncertainty* (1982) Kahneman, Tversky y su grupo de investigadores reúnen diversos estudios en donde proponen tres diferentes heurísticas de razonamiento, a saber, la heurística de representatividad, la heurística de disponibilidad y el anclaje y ajuste [5]. En la introducción de dicho libro, en donde Kahneman y Tversky explican brevemente en qué consisten tales heurísticas, se señala que las heurísticas tienen los siguientes rasgos:

a) Las heurísticas son altamente económicas y usualmente efectivas, pero conducen a errores sistemáticos y predecibles (1982a, p. 20).
b) Los sesgos no son atribuibles a efectos motivacionales tales como ilusiones o distorsiones del juicio debido a penalidades o castigos (1982a, p. 18).
c) La confianza en las heurísticas y la prevalencia de sesgos no se restringe a los legos. Los sujetos entrenados en estadística evitan errores elementales, pero sus juicios intuitivos tienden a cometer falacias en problemas de razonamiento más intrincados y menos transparentes (1982a, p. 18).

¿Qué implicaciones tienen los resultados de estos experimentos acerca de la naturaleza de los procesos y mecanismos de razonamiento? ¿Cómo se deben interpretar los anteriores resultados? A mi juicio, se puede distinguir entre dos grandes grupos de reacciones a esta evidencia. Un primer grupo de reacciones relacionadas íntimamente con el programa de heurística y sesgo sostiene que la evidencia empírica señala que los procesos cognitivos usados por los seres humanos se desvían sistemáticamente de las normas de racionalidad apropiadas. Dentro de esta interpretación, denominada la "interpretación pesimistas de la racionalidad humana", cabrían varias versiones que revisaré posteriormente en este capítulo.

Un segundo grupo de reacciones sostiene que los experimentos realmente no muestran que los seres humanos cometen errores sistemáticos de razonamiento. Al igual que con el grupo anterior que defendería la interpretación pesimista de la racionalidad humana, en este grupo cabrían diversas variantes para defender

que los seres humanos no cometen errores sistemáticos de razonamiento. La propuesta más radical dentro de este grupo es aquella que sostiene que los estudios de razonamiento no muestran nada relevante acerca del racionalidad humana. Dentro de las interpretaciones menos radicales se pueden distinguir cuatro maneras distintas de contrarrestar la visión pesimista de la racionalidad humana (Stanovich y West, 2000), a saber: *i.)* que los errores de razonamiento presentes en los experimentos constituyen meros *errores de ejecución* [6], por lo que la aparente desviación de los estándares normativos se debe, por ejemplo, a lapsos temporales o falta de atención; *ii.)* los errores de razonamiento se deben a limitaciones computacionales, constantes e inherentes que evitan la satisfacción adecuada de los estándares normativos; *iii.)* los psicólogos pueden estar aplicando un modelo inadecuado a la tareas psicológica que los sujetos llevan a cabo, y *iv.)* es posible que los sujetos estén haciendo uso de principios normativos adecuados al resolver problemas de razonamiento, pero que interpretan a éstos de manera distinta a lo que los psicólogos hacen de los mismos [7].

Haré uso de la distinción que hacen Stanovich y West (2000) de las distintas reacciones en contra de la interpretación pesimista de la racionalidad humana, puesto que recoge de manera adecuada las distinciones que otros psicólogos y teóricos han hecho (Cohen, 1981, Stein, 1996, Gilovich y Griffin, 2002). Así, en este capítulo revisaré la estrategia *i.)* descrita por Stanovich y West para contrarrestar la visión pesimista de la racionalidad humana, es decir, aquella que sostiene que los errores de razonamiento se reducen a errores de ejecución, producto de lapsos temporales o falta de atención por parte de los sujetos. En el siguiente capítulo revisaré con cuidado la postura defendida por los psicólogos evolucionistas, que en algunos aspectos difiere de modo importante de la interpretación pesimista que se revisa en este capítulo. En el capítulo IV analizaré las maneras *ii.), iii.)* y *iv.)* que, según Stanovich y West, pretenden contrarrestar la tesis pesimista de la racionalidad humana. De momento, es importante señalar que la interpretación pesimista ha sido el punto de partida de muchos de los debates contemporáneos alrededor del estudio del razonamiento. Por lo anterior, en varios capítulos de este trabajo se vuelve continuamente a la interpretación pesimista de la raciona-

lidad humana como una posición en psicología cognitiva que ha motivado el desarrollo de muchas de las teorías contemporáneas del razonamiento. En particular, porque alrededor de este debate es posible entender el fuerte vínculo que existe entre una teoría del razonamiento y el estudio evaluativo de la razón y normativo de la racionalidad. Asimismo, como argumentaré en las conclusiones de este capítulo, la interpretación pesimista de la racionalidad tiene consecuencias importantes para las nociones de justificación revisadas en el primer capítulo de este libro.

II.4. ¿PUEDE DEMOSTRARSE EMPÍRICAMENTE QUE LOS SERES HUMANOS SON IRRACIONALES?

Antes de analizar las distintas interpretaciones hechas de los estudios empíricos antes expuestos, es importante subrayar que a muchos psicólogos cognitivos y filósofos les parece paradójico el resultado de los mismos. Esta "paradoja" ha sido expuesta por Evans y Over (1996), quienes ponen de relieve la inteligencia humana, mostrada entre otras cosas por su capacidad para construir sistemas lingüísticos, para representar y comunicar información, para desarrollar pensamiento abstracto que da lugar a teorías científicas; pero, al mismo tiempo, el desempeño de los seres humanos en pruebas de razonamiento hechas por psicólogos parece sistemáticamente inadecuado. Muchos psicólogos y filósofos han ofrecido argumentos para sostener que del mal desempeño que tienen los sujetos en las pruebas de razonamiento no se sigue que los seres humanos son irracionales o ilógicos. El argumento de estos filósofos y psicólogos muchas veces parte de la idea —obvia para muchos de ellos— de que el ser humano es un ser racional. Una vez que se parte de dicha idea, lo que se busca es mostrar en qué sentido los estudios hechos por los psicólogos de heurística y sesgo, así como la interpretación que hacen de estos estudios, es inadecuada (Evans y Over, 1996). Un filósofo que defiende una postura de este tipo es Jonathan Cohen (1986).

Cohen (1981, 1986) sostiene que la intuición humana es racional y central para la investigación filosófica. El propósito de este filósofo es presentar un esquema teórico en el cual se dé cabida a una noción adecuada de la racionalidad humana junto con los experimentos provenientes de la psicología cognitiva. Con lo

anterior, Cohen pretende resolver la tensión que existe entre la idea de que el ser humano es racional y la interpretación pesimista de la racionalidad humana. La idea central detrás de este esquema teórico se funda, según Cohen (1981), en la distinción entre competencia y ejecución de razonamiento.

La distinción entre ejecución y competencia fue introducida a las ciencias cognitivas por Chomsky a partir de la publicación de su libro *Syntactic Structures* (1957), momento en el cual la interacción entre psicólogos y lingüistas cobró especial importancia. Según Chomsky, los hablantes de un lenguaje producen "intuiciones" o juicios irreflexivos acerca de la gramaticalidad de sus oraciones y acerca de ciertas propiedades y relaciones lingüísticas. Para explicar estas intuiciones y explicar cómo el hablante entiende oraciones en un discurso ordinario, Chomsky propuso que los hablantes de un lenguaje poseen una gramática de ese lenguaje representada internamente. Cuando un hablante produce un juicio se tiene acceso a la información de la gramática representada internamente. Los juicios o intuiciones lingüísticas de un hablante no son, para Chomsky, infalibles, pues la gramática del lenguaje de un hablante no produce por sí misma intuiciones. La producción de intuiciones consiste en un proceso complejo en el cual la gramática representada internamente debe interactuar con una variedad de mecanismos cognitivos incluyendo la percepción, la motivación, la atención, la memoria a corto plazo, etc.

En términos de Chomsky, la reglas y los principios de la gramática representada internamente de un hablante constituyen la *competencia lingüística* del hablante. En contraste, los juicios que un hablante hace acerca de las oraciones, incluyendo las oraciones que de hecho el hablante produce, son parte de la *ejecución lingüística* de un hablante. Un *error en la ejecución* surge cuando algunas de las oraciones que el hablante produce, así como algunos de los juicios que el hablante hace acerca de sus oraciones, no son el reflejo de la competencia lingüística del hablante (Stein, 1996). De este modo, la competencia lingüística hace referencia a las reglas y principios de la gramática internamente representada del individuo, mientras que la ejecución consiste en la producción de oraciones del individuo, las que están determinadas tanto por la competencia lingüística del individuo como por otros factores psicológicos [8]. Una vez descrito en qué consiste la distinción

chomskyana entre ejecución y competencia es necesario ver cómo esta distinción ha sido usada por filósofos y científicos cognitivos interesados en el estudio de la razón.

Uno de los puntos que tienen en común los fenómenos lingüísticos y aquellos relacionados con el razonamiento es que son productos espontáneos y generalmente inconscientes de una clase indefinida de insumos: los hablantes son capaces de entender una clase interminable de oraciones y esbozar inferencias de una clase interminable de premisas. En el ámbito de la lingüística, los individuos son capaces de elaborar juicios espontáneos acerca de una infinita clase de casos, por ejemplo, juicios sobre la gramaticalidad o ambigüedad, mientras que lo mismo pueden hacer los individuos en el ámbito del razonamiento con respecto a juicios acerca de la probabilidad o validez. Dadas estas similitudes, se ha explorado la idea de que los mecanismos que subyacen a la habilidad de razonar son sustancialmente similares a aquellos mecanismos que subyacen a la capacidad de procesar el lenguaje. Si es posible hacer una analogía entre los fenómenos lingüísticos y el razonamiento, y si la teoría de Chomsky es correcta, entonces se puede formular la hipótesis de que los individuos cuentan con reglas y principios de razonamiento internamente representados, al cual se tiene acceso cuando los individuos infieren o hacen juicios acerca de su racionalidad. Al igual que en el caso del lenguaje, este conjunto de reglas y principios no es accesible a la conciencia. Por último, las inferencias, juicios y decisiones de los individuos no son una guía infalible para determinar la validez o legitimidad de una inferencia a partir del conjunto de principios y reglas internamente representadas, pues como ocurre con el lenguaje, los principios y reglas internamente representadas deben interactuar con otros mecanismos cognitivos, por ejemplo, la atención, motivación, memoria de corto plazo, etc.

En suma, quienes defienden la analogía con el lenguaje, el conjunto de reglas y de principios internamente representados de un individuo constituye la *competencia de razonamiento* del individuo. En contraste, los juicios que hace el individuo con respecto a las inferencias, junto con las inferencias que el individuo de hecho lleva a cabo, son parte de la *ejecución de razonamiento* del individuo. Por último, algunas de las inferencias, juicios y decisiones que el individuo hace no reflejan la competencia de

razonamiento del individuo, y en estos casos el individuo comete *errores de ejecución*. Sin profundizar en las diferencias entre la distinción ejecución/competencia en el razonamiento y el lenguaje [9], veamos cómo ha sido usada esta distinción con relación a los experimentos provenientes de la tradición de heurística y sesgo.

Explicada la distinción entre competencia y ejecución de razonamiento que defiende Cohen es posible entender su posición. Para Cohen, los experimentos en psicología cognitiva provenientes de la tradición de heurística y sesgo sólo señalan errores de ejecución y no errores de competencia [10]. Cohen asume que en la mente hay algoritmos que son suficientes para solucionar los problemas de razonamiento que fueron expuestos anteriormente, tanto aquellos relacionados con el razonamiento inductivo (como la tarea de "Linda, la cajera feminista") como el deductivo (como las tareas de selección). Los errores expuestos en los experimentos muestran, según Cohen, lapsos momentáneos de atención, memoria o procesamiento o meras distracciones por parte del sujeto. Asimismo,

> las ilusiones son atribuidas a experimentos no realistas que hacen uso de información empobrecida, a experimentadores actuando como prestidigitadores, y a otros factores que enmascaran la competencia de los sujetos (Gigerenzer, 1993, p. 286, la traducción es mía).

Para Cohen, los seres humanos razonan correctamente acerca de la probabilidad, por lo que no pueden cometer falacias, a menos que sus juicios estén en ese momento "opacados" por la inmadurez, la senilidad, la poca inteligencia, la falta de atención, que estén siendo afectados por factores que inhiban el razonamiento, y que los experimentos sean poco realistas u ofrezcan información insuficiente para solucionarlos. Uno de los ejemplos usados por Cohen para ilustrar su posición parte de algunas versiones menos abstractas de la tarea de selección (Cohen, 1981). Su ejemplo plantea lo siguiente: supóngase que la regla que se debe probar es: "siempre que voy a Guadalajara, lo hago en autobús" y en lugar de presentar tarjetas con números y letras se presentan cartas con nombres de ciudades, en un lado, y nombres de medios de transporte, en su otro lado (las tarjetas podrían ser "voy a Guadalajara", "voy a Monterrey", "viajo en autobús", "viajo en

avión"). Como se señaló anteriormente, este tipo de tareas de selección arroja mejores resultados que la versión original de Wason presentada anteriormente, en la cual sólo aparecen letras y números. De lo anterior Cohen afirma:

> ...el poder del experimentador para generar una ilusión depende aquí de la poca familiaridad y artificialidad de su aparato [experimental]... Así, no se necesita presumir que los sujetos que razonan falazmente acerca del problema de las cuatro cartas carezcan del "programa" deductivo correcto... Los sujetos meramente fallan al reconocer la similitud de esa tarea con aquellas actividades familiares en las que han sido beneficiados al usar el procedimiento deductivo de la contraposición (Cohen, 1981, p. 324, la traducción es mía).

De este modo, según Cohen, los seres humanos tienen las reglas necesarias y suficientes para resolver los problemas de los experimentos hechos por Kahneman y Tversky; no obstante, los individuos cometen errores al no percatarse (en parte porque los problemas que presentan los psicólogos no le son familiares a los sujetos) que deben aplicar las reglas que se encuentran en su competencia de razonamiento. Los supuestos "errores sistemáticos de razonamiento" son así, para Cohen, meros errores de ejecución transitorios y fortuitos.

La respuesta que comúnmente se le ha dado a una visión como la defendida por Cohen es que existe un número importante de experimentos, hechos por distintos psicólogos, en los cuales parece haber la suficiente consistencia para sostener que los seres humanos más que "meros errores" cometen sistemáticas desviaciones de los estándares normativos adecuados (Stanovich y West, 2000). Por ejemplo, se ha puesto en duda que la familiaridad con los problemas de razonamiento por parte de los sujetos, como sugiere Cohen, realmente muestra que los errores de razonamiento en los experimentos sean sólo errores de ejecución. Stein (1996) apoya la tesis anterior al mostrar experimentos donde *a)* los sujetos realizan adecuadamente una tarea de razonamiento que no les es familiar [11]; y *b)* hay experimentos en donde los sujetos no ejecutan de manera apropiada tareas que sí le son familiares [12]. Las opciones *a)* y *b)* sugieren que la familiaridad no necesariamente pone en marcha la aplicación de los principios normativos de

razonamiento. Lo cual, a mi juicio, pondría en duda la postura de Cohen.

En suma, la objeción a la interpretación pesimista revisada aquí señala que toda evidencia proveniente de la psicología cognitiva, entre ella los experimentos alrededor de la tarea de selección, representa meros errores de ejecución. Cohen sostiene que los seres humanos, en la tareas de selección abstractas no aplican el principio de razonamiento adecuado porque no logran percatarse que la tarea en cuestión es similar a aquella que sí pueden resolver y que, de hecho, resolverían correctamente. Sin embargo, la consistencia en los diversos experimentos realizados en psicología del razonamiento, así como la evidencia de que existen experimentos centrados a situaciones "familiares", en donde los sujetos se desvían de los estándares normativos adecuados o experimentos no familiares que son resueltos adecuadamente, señalan que la interpretación de Cohen es incorrecta. Si la evidencia empírica proveniente de la psicología no puede ser interpretada como meros errores de ejecución, ¿debe entonces aceptarse que dicha evidencia señala errores de competencia? Para contestar esta pregunta es importante primero entender de mejor manera en qué consiste la interpretación pesimista de la racionalidad y evaluar hasta qué punto se puede defender esta interpretación.

II. 5. DOS MODOS DE ENTENDER LA INTERPRETACIÓN PESIMISTA DE LA RACIONALIDAD HUMANA

Existen varias maneras de entender la interpretación pesimista de la racionalidad humana, y a continuación presento dos versiones. Una primera versión de la interpretación pesimista se relaciona directamente con la manera en que los investigadores de la tradición de heurística y sesgo han explicado la ejecución de los sujetos en muchos de los problemas de razonamiento. La estrategia explicativa es postular la existencia de heurísticas de razonamiento que los sujetos emplean cuando razonan (Gilovich y Griffin, 2002). Kahneman y Tversky, por ejemplo, consideran que en el caso de la falacia de la conjunción los sujetos cognoscentes fundan su razonamiento en lo que han denominado heurística de representatividad (Kahneman, Slovic y Tversky, 1982). Dada una evidencia específica, por ejemplo, el esbozo de la personalidad, el

conjunto de resultados, como las ocupaciones a las que alguien se puede dedicar, son generalmente ordenadas por el grado en que tales consideraciones sean representativas de la evidencia. Para ilustrar esto, en el caso de la falacia de la conjunción, expuesta anteriormente, una vez mostrada la evidencia específica ofrecida por la personalidad de Linda, los resultados a considerar, en este caso las posibles ocupaciones de Linda, son ordenadas por el grado en que estas ocupaciones son representativas de la personalidad de Linda. Kahneman y Tversky (1982a) sostienen que los seres humanos predicen por representatividad, es decir, que los sujetos seleccionan u ordenan los resultados por el grado en que tales resultados representan los rasgos esenciales de la evidencia. Kahneman y Tversky afirman:

> en muchas situaciones, los resultados producto de la representatividad son de hecho más probables que otros resultados. No obstante, esto no es siempre el caso, porque hay factores (por ejemplo, los resultados de probabilidades previas y la confiabilidad de la evidencia), que afectan la probabilidad de los resultados pero no su representatividad. Debido a que estos factores son ignorados, las predicciones intuitivas violan las reglas estadísticas de predicción en modos sistemáticos y fundamentales (Kahneman y Tversky, 1982a, p. 19, la traducción es mía).

Si las heurísticas realmente explican cómo los seres humanos desarrollan juicios sobre racionalidad, entonces una primera versión de la interpretación pesimista de la racionalidad humana puede entenderse como:

> I. Muchas veces en que los juicios se desvían de las normas apropiadas de racionalidad se explica por el hecho de que los juicios y decisiones hechas por los sujetos se fundan en heurísticas que conducen a errores sistemáticos y severos.

Esta versión de la interpretación pesimista de la racionalidad humana subraya el papel de las heurísticas, las cuales explican el origen de los errores de razonamiento.

Si esta versión de la interpretación pesimista presenta desafíos a los defensores de que los juicios de los seres humanos son sistemáticamente racionales, existe una segunda versión de la interpretación pesimista que resulta aún más radical. Algunos

psicólogos han sugerido que la evidencia de los distintos estudios sobre la racionalidad humana muestra que los sujetos cognoscentes hacen uso de heurísticas porque no cuenta con otros mecanismos disponibles para resolver problemas de razonamiento. Posturas de este tipo parecen apoyarse en afirmaciones como la siguiente:

> Parece que las personas carecen de programas correctos para muchas tareas de juicio, no hemos tenido la oportunidad de desarrollar un intelecto capaz de lidiar conceptualmente con la incertidumbre (Slovic, Fischhoff y Lichtenstein, 1976, p. 174, la traducción es mía).

Una afirmación todavía más fuerte que conduce a una versión radical de la interpretación pesimista es la defendida por Piatelli-Palmarini, quien después de revisar los experimentos de la tradición de heurística y sesgo afirma:

> Se podría muy bien aducir que estamos cegados a las probabilidades... me gustaría sugerir una ley probabilística simple y general: cualquier intuición probabilística hecha por alguien que no haya sido instruido en cálculo de probabilidades tiene más del cincuenta por ciento de probabilidad de estar equivocada (Piatelli-Palmarini, 1994, pp. 131-132, la traducción es mía [13]).

Estas afirmaciones conducen a una segunda versión de la visión pesimista de la racionalidad humana:

> II. Los únicos mecanismos disponibles a los sujetos cognoscentes son heurísticas normativamente problemáticas.

Como es fácil notar en esta versión de la interpretación pesimista, no se hace mención de los juicios de los sujetos, pero se presupone que éstos son normativamente problemáticos porque se fundan en mecanismos que son en sí mismos problemáticos. A diferencia de la versión (I), la versión (II) está dando por hecho que los mecanismos de razonamiento del ser humano dependen siempre de heurísticas normativamente inapropiadas.

II.6. LA VISIÓN ESTÁNDAR DE LA RACIONALIDAD

Una vez expuesta la diferencia entre competencia de razonamiento y ejecución de razonamiento, así como distintas versiones de

la interpretación pesimista de la racionalidad humana, es necesario hacer patente un supuesto que subyace detrás de la interpretación pesimista y de algunas de las posturas que intentan contrarrestar dicha interpretación. En este capítulo he expuesto una teoría descriptiva del razonamiento, i.e., aquella defendida por la tradición de heurística y sesgo. La interpretación del razonamiento humano como "pesimista" presupone que la tradición de heurística y sesgo no sólo describe cómo razonan los seres humanos, sino que califica o evalúa el modo en que lo hacen. Concebir al ser humano de manera pesimista es parte de la teoría evaluativa de la tradición de heurística y sesgo, pues, como he mencionado, la teoría evaluativa de la razón determina si los seres humanos satisfacen los estándares normativos adecuados. Al hacer esta evaluación los psicólogos defensores de la tradición de heurística y sesgo presuponen necesariamente un estándar normativo particular, pues requieren de ciertos criterios para poder evaluar el razonamiento. Aun cuando muchas veces los defensores de la postura pesimista no ofrecen explícitamente una teoría normativa de la racionalidad, ellos asumen lo que Stein ha llamado la "visión estándar de la racionalidad".

> De acuerdo con esta visión, ser racional es razonar en concordancia con los principios de razonamiento que están fundados en las reglas de la lógica, la teoría de la probabilidad u otras teorías formales. Si la visión estándar de razonamiento es correcta, los principios de razonamiento que están fundados en tales reglas son los principios normativos de razonamiento, a saber, son los principios con los que nuestro razonamiento debe estar en concordancia (Stein, 1996, p. 4, la traducción es mía).

Según esta visión estándar de la racionalidad, el criterio apropiado con el cual se debe evaluar el razonamiento humano se deriva de las reglas de la lógica clásica, la teoría de la probabilidad y la teoría de la decisión. De este modo, podría derivarse, por ejemplo, la siguiente regla de razonamiento de la regla de la conjunción de la teoría de la probabilidad

> Principio de conjunción: No se debe asignar un grado menor de probabilidad a la ocurrencia de un suceso A que a la ocurrencia de

> un suceso A y algún otro suceso (distinto) B (Stein, 1996, p. 6, la traducción es mía).

Si se acepta que el principio anterior es correcto, es fácil entender porque los patrones de inferencia discutidos en la falacia de la conjunción, por ejemplo, en el caso de "Linda, la cajera feminista", son normativamente problemáticos. La visión estándar de la racionalidad conduce a que se deben aceptar aquellos y sólo aquellos razonamientos que son sancionados por los principios normativos provenientes de las reglas de la lógica u otras teorías formales (Piatelli-Palmarini, 1994 [14]).

Así, se puede sostener que la tradición de heurística y sesgo es una teoría del razonamiento que explica cómo es que los seres humanos razonan. Esta teoría psicológica también evalúa el proceder racional, teniendo como producto de esa evaluación la interpretación pesimista de la racionalidad humana (en sus dos versiones). Para llevar a cabo dicha evaluación los psicólogos defensores de la tradición de heurística y sesgo parten de una teoría de la racionalidad particular, a saber, la visión estándar de la racionalidad. En las conclusiones de este capítulo revisaré si la interpretación pesimista de la racionalidad humana tiene consecuencias con la manera de entender la noción de justificación.

II. 7. CONCLUSIONES

Partiendo de la idea de que el debate en torno a la noción de justificación conduce al estudio de la razón, en este capítulo se ha hecho una distinción importante entre teorías descriptivas del razonamiento, teorías evaluativas de la razón y teorías normativas de la racionalidad. Esta distinción es una herramienta de análisis que sirve para entender mejor las diferentes teorías psicológicas que reviso en este libro. Aquí he señalado cómo estas distintas maneras de acercase al estudio del razonamiento y la racionalidad se encuentran íntimamente ligadas, pues la tradición de heurística y sesgo, en tanto teoría descriptiva del razonamiento, asume una visión normativa particular, a saber, la visión estándar de la racionalidad. Una vez asumida esta visión de la racionalidad, la tradición de heurística y sesgo determina hasta qué punto los seres humanos satisfacen estos criterios de racionalidad, es decir,

la tradición de heurística y sesgo tiene implicaciones en una teoría evaluativa de la razón. La visión estándar de la racionalidad y la explicación de ciertos experimentos por la tradición de heurística y sesgo conducen a la interpretación pesimista de la racionalidad humana.

En el capítulo he distinguido dos distintas versiones de la interpretación pesimista de la racionalidad humana. La primera versión de esta interpretación la he caracterizado como aquella que afirma que los juicios y razonamientos que se desvían de los estándares normativos adecuados son explicados por heurísticas que conducen a errores sistemáticos de razonamiento. La segunda versión, definitivamente más fuerte que la versión anterior, considera que los únicos mecanismos disponibles a los seres humanos son heurísticas problemáticas. He apuntado, además, distintas maneras de contrarrestar la interpretación pesimista de la racionalidad humana y revisado una de estas estrategias, a saber: aquella que hace una distinción entre competencia y ejecución de razonamiento. Esta distinción, defendida por Cohen (1981, 1986), es usada para sostener que los experimentos de la tradición de heurística y sesgo pueden entenderse como meros errores de ejecución. Según lo visto en este capítulo, esta respuesta a la interpretación pesimista no es adecuada, pues no demuestra que la mayoría de los experimentos en psicología se reduzcan únicamente a errores de ejecución (Stanovich y West, 2000).

En los próximos capítulos discutiré otras teorías del razonamiento y haré explícitas otras posturas alrededor de la teoría evaluativa y normativa de la racionalidad, de modo que, en gran medida, este capítulo servirá como punto de partida para posteriores discusiones. Así pues, es prudente preguntarse: ¿cuáles son las consecuencias epistemológicas de aceptar que los sujetos cognoscentes son sistemáticamente irracionales? y, ¿qué impacto tendría lo anterior sobre la noción de justificación? En estas conclusiones indagaré algunas de las consecuencias que tiene la teoría del razonamiento defendida por la tradición de heurística y sesgo, y la interpretación pesimista de la racionalidad sobre el debate en teoría de la justificación revisado en el capítulo anterior.

La tradición de heurística y sesgo, como se expuso en este capítulo, sostiene que el razonamiento está subsumido en heurísticas que conducen a errores sistemáticos. En la versión más fuerte

de la interpretación pesimista, es decir, la versión (II) se sostiene que los únicos mecanismos disponibles a los sujetos son heurísticas normativamente problemáticas. Una postura como ésta tiene consecuencias negativas tanto para una postura internista como externista, en tanto que ofrece una visión negativa de las capacidades cognitivas humanas. Si se sostiene la versión más fuerte de la interpretación pesimista, se tiene que aceptar que los juicios de los seres humanos se encuentran casi siempre sesgados, y la formación y evaluación de las creencias sería en la mayoría de la casos irracional [15]. Si además se acepta que una creencia que es racional creer es una creencia justificada, casi ninguna creencia estaría justificada. Esta idea choca con los trabajos elaborados en la teoría de la justificación que parten de la idea de que hay creencias que de hecho están justificadas [16]. Precisamente la defensa de una teoría de la justificación particular presupone la idea de que existen creencias que están justificadas y que éstas pueden distinguirse de las creencias que no están justificadas. Veamos un ejemplo de esto.

En el capítulo anterior se estudió con cierto detalle la postura confiabilista de Goldman, que sostiene que la justificación depende de los procesos cognitivos confiables donde el razonamiento deductivo e inductivo juegan un papel primordial. Pero si se aceptan las conclusiones de la tradición de heurística y sesgo, se tendría que aceptar también que los procesos cognitivos relacionados con la deducción y la inducción son procesos que producen errores sistemáticos que señalan desviaciones importantes de los principios normativos adecuados. Dada esta teoría psicológica, incluso los procesos que generalmente se consideran confiables llevan a errores de razonamiento recurrentes. De hecho, la versión (II) de la interpretación pesimista supondría que casi siempre el razonamiento conduce a errores, por lo que simplemente no habría cabida para el confiabilismo y otro tipo de teorías externistas que enfaticen el papel de los procesos de formación de creencias. Tal parece que las diferentes heurísticas propuestas por Kahneman, Tversky y sus seguidores —según esta versión de la interpretación pesimista— conducen a errores, que *en principio* harían difícil entender la idea de un proceso cognitivo confiable, así como imposible separar los procesos cognitivos adecuados de los no adecuados. En otras palabras, no habría manera de enten-

der cómo es que un tipo de proceso cognitivo genera creencias justificadas, si los únicos mecanismos disponibles a los seres humanos son heurísticas normativamente problemáticas.

A mi juicio, el que se acepte la versión (II) de la interpretación pesimista de la racionalidad humana conduzca a un rechazo de cualquier noción de justificación, tiene también otra consecuencia epistemológica inmediata, relacionada con la idea generalizada de que el conocimiento requiere de la noción de justificación. Si la interpretación pesimista en su versión (II) es correcta, entonces no habría creencias justificadas propiamente dichas y de ahí el ser humano no tendría conocimiento, por lo menos, bajo la noción tradicional de conocimiento. Así, la interpretación pesimista de la racionalidad humana, en su versión más fuerte, apoyaría un tipo de escepticismo epistemológico. Por tal motivo, en los siguientes capítulos de este trabajo será importante encontrar apoyo teórico o empírico en contra de esa interpretación pesimista en su versión (II). Entonces, ¿cuáles serían las consecuencias epistemológicas de aceptar la versión más débil de la interpretación pesimista de la racionalidad humana, es decir, la versión (I)?

Los internistas generalmente han querido hacer una distinción entre el modo en que las creencias son producidas o causadas, con la manera en que éstas se justifican (Davidson, 1983). La postura internista de Bonjour deja de lado el modo en que son producidas las creencias, y considera las relaciones lógicas de éstas en un sistema coherente de creencias y la idea de que el sujeto debe (en principio) creer justificadamente que tales relaciones conducen a la verdad. En este sentido, la justificación de las creencias para un internista como Bonjour estaría garantizada por relaciones coherentes en sistemas de creencias, aun cuando los procesos que originaron creencias y los procesos de razonamiento sean algunas veces normativamente inadecuados. Si se acepta la versión débil (I) de la interpretación pesimista, se podría sostener que habría creencias justificadas bajo la noción internista, puesto que es posible que el razonamiento se guíe, algunas veces, por heurísticas que conducen a errores sistemáticos, pero que la justificación depende del acceso a la coherencia de un sistema de creencias y este acceso no esté sesgado por alguna heurística.

Una teoría externista de la justificación hace énfasis en el papel de los procesos de formación de creencias adecuados. La versión

(I) de la interpretación pesimista de la racionalidad humana sostiene que los seres humanos cuentan con heurísticas que conducen a errores sistemáticos de razonamiento. En esta versión cabría la posibilidad de mecanismos y procesos de razonamiento normativamente adecuados. En principio, una postura externista, quizá como la confiabilista de Goldman, podría sostener que las creencias justificadas son producto de aquellos procesos de formación de creencias que no se encuentran sesgados por algunas de las heurísticas que los psicólogos cognitivos han descubierto. Así, es posible sostener que la noción externista de justificación puede ser compatible con la tradición de heurística y sesgo, si la versión de la interpretación pesimista que se sigue de dicha postura psicológica es la versión (I) de la interpretación pesimista de la racionalidad humana.

De este modo, parece que es concebible que las nociones internista y externista de la justificación puedan ser compatibles con la versión (I) de la interpretación pesimista de la racionalidad proveniente de la tradición de heurística y sesgo. Esto es, las nociones internistas y externistas de la justificación son lógicamente compatibles con los postulados de la tradición de heurística y sesgo. Por el contrario, si lo argumentado anteriormente es correcto, cualquier noción de justificación es puesta en duda si se acepta la versión (II) de la interpretación pesimista de la racionalidad humana. Esta versión de la interpretación pesimista pondría en duda también la noción de conocimiento, si se considera que la justificación es una parte elemental de tal noción.

Parte de lo que motiva el estudio de las teorías del razonamiento en este libro es el deseo de encontrar en ellas una salida al debate contemporáneo entre nociones internistas y externistas de la justificación. El estudio de la tradición de heurística y sesgo —en sus aspectos descriptivos, evaluativos y normativos— tiene interesantes puntos de contacto con la epistemología. Por un lado, la versión (II) de la interpretación pesimista de la racionalidad humana apoyaría un tipo de escepticismo epistemológico. Por otro lado, la versión (I) de la interpretación pesimista no conduce a un escepticismo, pero según lo revisado hasta aquí es difícil pensar que la tradición de heurística y sesgo puede ser usada para evaluar las nociones de justificación, puesto que tanto la noción internista de la justificación, como la noción externista de la

justificación muestran compatibilidad con esta teoría del razonamiento. Por tal motivo, en los próximos capítulos exploraré otras teorías del razonamiento que *i.)* puedan eliminar la posibilidad de la versión (II) de la interpretación pesimista de la racionalidad humana, *ii.)* ayuden a entender de mejor manera la relación entre justificación y racionalidad. En los últimos años, el debate alrededor del razonamiento y de la interpretación pesimista de la racionalidad humana ha dado lugar a diversas teorías psicológicas que compiten con la tradición de heurística y sesgo, entre ellas la psicología evolucionista que expondré en el siguiente capítulo.

III.
LA PSICOLOGÍA EVOLUCIONISTA Y LA RACIONALIDAD

> En el futuro veo campo abierto para investigaciones más importantes. La psicología estará asegurada sobre un [nuevo] fundamento...
>
> Charles Darwin, *The Origin of Species*
> (la traducción es mía)

Siguiendo con el hilo conductor que guía este libro, esto es, la idea de indagar la conexión entre las nociones de racionalidad y de justificación a través de las teorías del razonamiento, en este capítulo exploraré la propuesta defendida por la psicología evolucionista. La importancia de estudiar esta postura se debe, en parte, a que los psicólogos evolucionistas defienden una postura que contrarresta la interpretación pesimista de la racionalidad humana.

En los últimos años un grupo de psicólogos evolucionistas, entre ellos Cosmides y Tooby [1], han criticado fuertemente la descripción pesimista del razonamiento humano. La psicología evolucionista es un proyecto interdisciplinario —que adquiere relevancia sobre todo en la década de los noventa— en el que poco a poco se ha empezado a desarrollar de manera consensuada un cuerpo teórico que lo define de manera precisa. Los psicólogos evolucionistas comparten *i.)* una descripción de la estructura de la mente que se ha conocido como la teoría de la modularidad masiva, y *ii.)* el modo en que la investigación psicológica debe proceder; especialmente afirman que consideraciones adaptacionistas deben jugar un papel central en la formación de hipótesis psicológicas. Estas dos tesis centrales se encuentran en la siguiente descripción:

> El objetivo de investigación en la psicología evolucionista es descubrir y entender el diseño de la mente humana. La psicología evolucionista es un *acercamiento* a la psicología, en el que el conocimiento y los principios de la biología evolucionista son usados para investigar la estructura de la mente humana Bajo esta visión, la mente es un conjunto de máquinas procesadoras de información que fueron

> diseñadas por la selección natural para resolver problemas adaptativos a los que nuestros ancestro cazadores-recolectores se enfrentaron (Cosmides y Tooby, 1997, la traducción es mía).

La estructura de este capítulo es la siguiente. Primero exploro las tesis centrales de la psicología evolucionista. Posteriormente discuto cómo dichas tesis han sido aplicadas al estudio del razonamiento humano. Me centraré especialmente en dos hipótesis psicológicas, a saber, la hipótesis de la detección del tramposo o mentiroso, y la hipótesis frecuentista, así como en la evidencia que se ha dado para sostener dichas hipótesis. A pesar de que estas hipótesis son consideradas como afirmaciones psicológicas descriptivas, los psicólogos evolucionistas han sugerido que dichas hipótesis pueden ser usadas para rechazar la interpretación pesimista de la racionalidad humana. Por tal motivo, analizaré si la psicología evolucionista puede contrarrestar las dos versiones de la interpretación pesimista hechas en el capítulo anterior. Por último, reviso las consecuencias epistemológicas, principalmente aquellas que se relacionan con la noción de justificación, que se derivan de las tesis centrales de la psicología evolucionista.

III. 1. IDEAS CENTRALES ALREDEDOR DE LA PSICOLOGÍA EVOLUCIONISTA

Los psicólogos evolucionistas consideran que la teoría que defienden, más que una teoría psicológica particular, es un *acercamiento* general a la psicología, ya que afirman que la psicología es una rama de la biología. De lo anterior, los psicólogos evolucionistas hacen uso de cinco principios provenientes de la biología para entender el diseño de la mente humana: 1) el cerebro es un sistema físico que funciona como una computadora, y en tanto sistema físico está gobernado por las leyes de la física y de la química; 2) los circuitos neuronales de los humanos son producto de la selección natural para resolver problemas que sus ancestros enfrentaron durante la historia evolutiva de la especie [2]; 3) la conciencia es sólo la "punta del iceberg" de lo que realmente ocurre en las mentes, la mayoría de los problemas que parecen fáciles de resolver requieren de un muy complicado circuito neuronal; 4) diferentes circuitos neuronales están especializados para resolver diferentes problemas adaptativos, y 5) para decirlo metafóricamen-

te "los cráneos modernos albergan una mente de la edad de piedra" (Cosmides y Tooby, 1997, la traducción es mía). Algunos de estos principios serán explicados en las siguientes secciones.

Los psicólogos evolucionistas aceptan lo que se ha denominado como la hipótesis de la modularidad masiva [3]. *Grosso modo,* la hipótesis de la modularidad masiva es la afirmación de que la mente humana está compuesta en su mayoría o completamente de mecanismos cognitivos altamente especializados o *módulos*. Aun cuando hay diversas maneras en que esta afirmación puede ser entendida, la mayoría de psicólogos evolucionistas asumen las tres siguientes tesis:

— *Computacionalismo*. La mente humana es un dispositivo o mecanismo de procesamiento de información que puede ser descrito en términos computacionales [4]. Para muchos psicólogos evolucionistas la mente es un computadora elaborada con componentes orgánicos en lugar de "chips" de silicón. El computacionalismo se hace patente en los principios de las psicología evolucionista que defienden Cosmides y Tooby (el principio (1)) revisado anteriormente, así como en otras caracterizaciones.

> Un módulo cognitivo es un dispositivo computacional especificado genéticamente en la mente/cerebro que trabaja sobre sus propios *inputs* relacionados con un dominio cognitivo específico [5] (Sperber, 1994, p. 40, la traducción es mía).

En este sentido, la psicología evolucionista adopta el computacionalismo que prevalece en muchas de las ciencias cognitivas [6].

— *Innatismo*. A diferencia de las posturas empiristas que han dominado la psicología en el siglo veinte, la psicología evolucionista considera que gran parte de la estructura de la mente es innata. En este sentido, la psicología evolucionista se adhiere al innatismo emparentado con Chomsky y sus seguidores [7], y rechaza la propuesta empirista de que la estructura innata de la mente humana se reduce al propósito general de los mecanismos de aprendizaje.

— *Adaptacionismo*. Los psicólogos evolucionistas afirman que en gran medida la estructura cognitiva es el producto de la selección natural. Bajo esta perspectiva, las mentes están compuestas de *adaptaciones* que tuvieron su origen en la selección natural durante

la historia evolutiva de las especies. Las mentes, según la psicología evolucionista, fueron diseñadas por la selección natural para resolver problemas adaptativos, esto es, "problemas evolucionistas recurrentes cuya solución promovía la reproducción" (Cosmides y Tooby, 1994, p. 87, la traducción es mía). Los problemas adaptativos, según Cosmides y Tooby, tienen dos características definitorias:

> Primero, son problemas que se plantearon una y otra vez durante la historia evolutiva de las especies. Segundo, son problemas cuya solución afectó la *reproducción* de un organismo individual (Cosmides y Tooby, 1997, la traducción es mía).

Si bien, según esta perspectiva, la mente humana está diseñada para resolver problemas adaptativos, es claro que puede resolver problemas a los que sus ancestros cazadores-recolectores nunca se enfrentaron, por ejemplo, aprenden matemáticas, manejan automóviles, usan computadoras, etc. Para la psicología evolutiva, la habilidad para resolver otro tipo de problemas es un efecto secundario o un subproducto (*by-product*) de los circuitos neuronales que fueron diseñados para resolver problemas adaptativos (Cosmides y Tooby, 1997). Una manera de ilustrar lo anterior es señalando que una vez que los ancestros humanos se volvieron bípedos tuvieron que desarrollar un buen sentido del equilibrio. De hecho, el ser humano ha desarrollado intrincados mecanismos en el oído interno que permiten alcanzar un excelente sentido del equilibrio; pero este equilibrio permite también hacer otras cosas además de caminar, por ejemplo, patinar o andar en bicicleta. En este sentido, patinar o andar en bicicleta serían meros subproductos de adaptaciones diseñadas para equilibrarse mientras se camina en dos piernas (Cosmides y Tooby, 1997). Este punto será importante para entender mejor dos de las hipótesis más importantes de la psicología evolucionista que revisaré en las siguientes secciones.

Los psicólogos evolucionistas conciben a los módulos como mecanismos o dispositivos computacionales que son "específicos de dominio" en oposición a dispositivos de "dominio general". Decir que una estructura es específica de dominio significa que está dedicada a solucionar una clase de problemas en un dominio restringido [8]. El ejemplo que generalmente ilustra la especificidad

de dominio hace referencia a la idea de que si la estructura cognitiva de la visión es de dominio específico, entonces habría estructuras mentales que son puestas en juego en el dominio del procesamiento visual y no de otros tipos de procesamiento, como el auditivo o el táctil. En contraste, un dispositivo es de dominio general, si es una estructura cognitiva que es capaz de procesar problemas en un rango amplio de diferentes dominios.

En consonancia con su innatismo y adaptacionismo, los psicólogos evolucionistas consideran que los módulos son innatos y que además son adaptaciones producidas por la selección natural. De este modo, para la psicología evolucionista la mente humana está en su mayoría o quizá enteramente compuesta por módulos innatos, entendidos como mecanismos computacionales específicos, producto de la selección natural. A esta noción de modularidad se le conoce como la noción darwiniana de modularidad.

Aun cuando el propósito de esta sección es explicar la noción de módulo defendida por los psicólogos evolucionistas y no un estudio de diferentes posturas acerca de la modularidad, es necesario indicar que la tesis de la modularidad masiva resulta más radical que las primeras descripciones modulares de la cognición, defendidas entre otros por Jerry Fodor. De acuerdo con este filósofo, la estructura modular de la mente está restringida a los "sistemas de insumo" (*input systems*), los cuales son responsables del procesamiento del lenguaje y la percepción (Fodor, 1983). Las características de los módulos son, según Fodor, mecanismos específicos de dominio, cada módulo cuenta con su transductor, sus *outputs* son no conceptuales, los módulos son innatos, asociados con una estructura específica de la mente, mandatorios, rápidos en sus procesamientos, informacionalmente encapsulados e inaccesibles [9], y además son insensibles a las metas cognitivas centrales. Si bien los psicólogos evolucionistas aceptan la tesis fodoriana de que los sistemas de insumos son modulares, también aceptan que las denominadas "capacidades centrales", tales como el razonamiento, pueden tener una estructura modular [10].

III. 2. EL PROGRAMA DE INVESTIGACIÓN DE LA PSICOLOGÍA EVOLUCIONISTA

Una meta central de la psicología evolucionista es construir y probar hipótesis acerca de la visión de modularidad que ellos

consideran constituye la estructura de la mente humana. En la búsqueda de esta meta, los psicólogos evolucionistas, según Samuels, Stich y Faucher (2004), dividen su trabajo en dos etapas. La primer etapa, denominada el "análisis evolucionista", tiene como meta generar hipótesis verosímiles acerca de los módulos darwinianos. Un análisis evolucionista trata de determinar, en la medida de lo posible, los problemas de procesamiento de información que los ancestros humanos tuvieron que enfrentar en lo que se ha llamado "ambiente de adaptación evolutiva" (AAE). El punto más importante consiste en aquellos problemas de adaptación cuya solución exitosa pudo haber contribuido de manera directa o indirecta a un éxito reproductivo. En algunos casos, estos problemas de adaptación fueron planteados por las características físicas del AAE, en otros casos, fueron planteados por restricciones biológicas o por el ambiente social al cual estuvieron sujetos los ancestros humanos. Debido a que son muchos los factores que originaron los problemas de procesamiento de información con los que se enfrentaron los ancestros humanos, el análisis evolucionista suele verse como un ejercicio interdisciplinario. Algunas claves para entender esto depende de distintas líneas de investigación, que pueden ir desde el estudio del clima en el pleistoceno hasta el estudio de las organizaciones sociales en las culturas recolectoras y cazadoras. Una vez que un problema de adaptación ha sido caracterizado, los teóricos pueden formular hipótesis acerca de un módulo que pudo haber hecho un buen trabajo en la solución de un problema en el AAE.

Es importante tener en cuenta que el análisis evolucionista puede ser usado únicamente como una manera de sugerir hipótesis verosímiles acerca de los módulos mentales. Por sí mismo el análisis evolucionista no puede comprobar la verdad de dichas hipótesis. Una vez que el análisis evolucionista ha tenido éxito al sugerir hipótesis verosímiles, la segunda etapa en la investigación de la psicología evolucionista es "probar" las hipótesis indagando si los seres humanos contemporáneos de hecho tienen módulos con las propiedades en cuestión. Aquí, como sucede con el análisis evolucionista, el proyecto es altamente interdisciplinario, y la evidencia puede provenir de estudios experimentales sobre el razonamiento en humanos normales (Gigerenzer y Hug, 1992) o de estudios de desarrollo en la aparición de habilidades cognitivas

o deficiencias cognitivas en seres humanos con ciertas discapacidades (Kamiloff-Smith, 1992). Cuando la evidencia de un número importante de distintas áreas de estudio señalan a la misma dirección, se puede sostener con cierto fundamento la existencia de un módulo tal y como lo sugiere el análisis evolucionista[11].

En las siguientes secciones mostraré cómo estas dos etapas de investigación que distinguen Samuels, Stich y Faucher (2004) ayudan a entender el tipo de hipótesis que brindan los psicólogos evolucionistas para explicar el razonamiento humano. Aun cuando gran parte de la interpretación de los estudios que se presentará son sujetos a debate, un número importante de psicólogos cognitivos —por ejemplo, Cosmides y Tooby, Gigerenzer y su grupo de investigación— han sugerido que ellos muestran que hay algo errado en la interpretación pesimista proveniente de la tradición de heurística y sesgo. Algunos psicólogos evolucionistas sostienen que hay evidencia a favor de la existencia de al menos dos módulos darwinianos sofisticados, a saber, uno diseñado para lidiar con el razonamiento probabilístico cuando la información es presentada en un formato de frecuencia, y el otro diseñado para lidiar con razonamientos acerca de un tramposo o mentiroso en un escenario de intercambio social. La eficacia de estos módulos para resolver de manera adecuada tareas de razonamiento ha hecho que muchos psicólogos evolucionistas defiendan una postura "optimista" acerca de la racionalidad humana [12].

III.3. LA HIPÓTESIS FRECUENTISTA

Los experimentos revisados anteriormente (especialmente aquellos relacionados con los argumentos inductivos como la falacia de la conjunción) indican, según la tradición de heurística y sesgo, que muchas veces los seres humanos no razonan de manera adecuada en problemas donde interviene la probabilidad. Según lo visto anteriormente, los defensores de la tradición de heurística y sesgo afirman que los sujetos cognoscentes hacen uso de heurísticas al tratar este tipo de problemas, que muchas veces conducen a errores sistemáticos de razonamiento. En provocativos artículos Gigerenzer (1994), así como Cosmides y Tooby (1996) han argumentado que desde un punto de vista evolutivo esto es un resultado sorprendente y paradójico. Estos psicólogos parten de

la idea de que elaborar juicios cuando se encuentra frente a información probabilística debió plantear problemas adaptativos para todo tipo de organismos, incluyendo a los ancestros homínidos. Cosmides y Tooby consideran que si un problema adaptativo ha persistido durante un periodo suficiente de tiempo, entonces mecanismos de una complejidad considerable pudieron evolucionar para resolverlo. Sin embargo, para llevar un análisis del razonamiento probabilístico se debe atender al

> engranaje entre el diseño de nuestros mecanismos cognitivos, la estructura de los problemas que hicieron que dichos mecanismos evolucionaran y los ambientes típicos que diseñaron el modo de operar de tales mecanismos (Cosmides y Tooby, 1996, p. 14, la traducción es mía).

En este sentido, se debe tener presente que, según los psicólogos evolucionistas, los seres humanos modernos se enfrentan con información estadística de muy diversa índole, por ejemplo, la probabilidad de que mañana será un día soleado que brinda un informe meteorológico o que el riesgo de padecer diabetes puede reducirse si se lleva una vida activa y se modera el consumo de azúcares. Este tipo de información acerca de la probabilidad de eventos particulares o la misma información expresada en porcentajes no debió estar disponible en el AAE. Lo que seguramente estuvo disponible en el AAE, según los psicólogos evolucionistas, fue el enfrentarse con frecuencias de eventos actuales (por ejemplo, que se tuvo éxito una de las cuatro veces que se cazó junto al lago). Los ancestros homínidos estuvieron inmersos en una gran cantidad de frecuencias observables que pudieron haber usado para mejorar la toma de decisiones. Los ancestros homínidos debieron haber tomado la información de frecuencias como insumo, y una vez que el sistema cognitivo hubo registrado la información acerca de la frecuencia relativa, según estos psicólogos, se debió convertir esta información en otro formato. Si, por ejemplo, el sistema notó que una de las cuatro veces que se cazó junto al lago fue exitosa, pudo haberse inferido la conclusión que hay 0.25 de probabilidad que cazar junto al lago será exitoso.

> Hay ventajas al almacenar y operar bajo representaciones frecuentistas porque preservan información importante que estaría perdida al hacer la conversión a una probabilidad de sucesos simples. Por ejemplo... el numero de sucesos en la que fue fundado un juicio estaría perdido en una conversión. (Cosmides y Tooby, 1996, p. 16, la traducción es mía).

Las distintas consideraciones acerca del ambiente en el cual los sistemas cognitivos evolucionaron conducen a Cosmides y Tooby a proponer la hipótesis de que los ancestros humanos tomaron las frecuencias como insumos, mantuvieron esa información como representaciones frecuentistas y usaron dichas representaciones como bases de datos para razonamientos inductivos efectivos. Así, estos psicólogos denominaron como la *hipótesis frecuentista* a la "hipótesis de que nuestros mecanismos de razonamiento inductivo fueron diseñados para operar con y para producir representaciones en formato de frecuencia [13]" (Cosmides y Tooby, 1996, p. 21, la traducción es mía).

> Algunos de nuestros mecanismos de razonamiento inductivo encarnan aspectos del cálculo de probabilidad, pero fueron diseñados para tomar como insumos información en formato de frecuencia y producir frecuencias como respuestas (Cosmides y Tooby, 1996, p. 3, la traducción es mía).

Debido a que los psicólogos evolucionistas consideran que la mente contiene un número importante de módulos especializados, ellos suponen que puede haber una variedad importante de módulos que estén involucrados en el razonamiento inductivo. Sin embargo, la hipótesis frecuentista no significa que cada mecanismo cognitivo involucrado en una inducción estadística necesariamente opera bajo el principio frecuentista, sólo dice que al menos uno de ellos lo hace. De este modo, si bien el análisis evolucionista hecho por estos psicólogos no excluye la posibilidad de mecanismos inductivos que no estén centrados en frecuencias (Cosmides y Tooby, 1996), sí sugieren que los mecanismos que operan involucrando el principio frecuentista harán un buen trabajo, y entonces las inferencias probabilísticas que produzcan serán generalmente normativamente adecuadas [14]. Esta postura contraviene a la interpretación pesimista de la racionalidad hu-

mana, según la cual las personas simplemente no tienen acceso a las estrategias normativas apropiadas en esta área. Al final de este capítulo analizaré hasta qué grado la aceptación de algunos de los principios centrales de la psicología evolucionista conduce al rechazo de la interpretación pesimista de la racionalidad humana [15].

De la hipótesis frecuentista Cosmides y Tooby derivan un número importante de predicciones: *i.)* la ejecución de una tarea de razonamiento inductivo diferirá, dependiendo de que se le pida a los sujetos juzgar la frecuencia o la probabilidad de un evento singular; *ii.)* la ejecución de tareas que involucren versiones frecuentistas será superior a la versiones no frecuentistas, y *iii.)* entre más puedan ser llevados los sujetos a una forma frecuentista de representación, mucho mejor será su ejecución. Para probar estas predicciones Cosmides y Tooby regresan a algunos de los experimentos de la tradición de heurística y sesgo que fueron usados por los defensores de la interpretación pesimista para demostrar que los sujetos ignoran la "razón de base" y cometen recurrentemente la "falacia de la conjunción". Veamos.

En un primer experimento Cosmides y Tooby replicaron los resultados de Kahneman y Tversky, así como el trabajo de Casscells et al. (1978), revisado anteriormente, en donde los sujetos ignoran la razón de base. Cosmides y Tooby reportaron que este mismo experimento, hecho en Stanford sólo el 12 por ciento de las personas que fueron encuestadas dieron la respuesta correcta, es decir, 2 por ciento; mientras que el 56 por ciento de las personas contestaron que la respuesta era 95 por ciento. De modo que los resultados obtenidos por Cosmides y Tooby fueron similares a los obtenidos anteriormente. Sin embargo, en otro experimento estos psicólogos evolucionistas dieron a 50 estudiantes de Stanford un problema similar en el que la frecuencia relativa, y no los porcentajes, de la probabilidad de un evento particular fueron enfatizados (Cosmides y Tooby, 1996). La versión "frecuentista" del problema es la siguiente:

> Uno de cada 1 000 estadounidenses tienen la enfermedad X. Un *test* médico ha sido desarrollado para detectar qué persona tiene la enfermedad X. Cada vez que el *test* es aplicado a la persona que tiene dicha enfermedad, el test resulta positivo. Sin embargo, algunas veces el *test* ha resultado positivo aun cuando la persona está com-

> pletamente sana, de hecho, de cada 1 000 personas que están completamente sanas 50 de ellas tienen un resultado positivo.
>
> Imagina ahora que tenemos un grupo de 1 000 estadunidenses seleccionados completamente al azar y no se tiene información alguna sobre la condición de salud de cada una de estas personas. Dada la anterior información: ¿cuántas personas que tengan un resultado positivo a partir del *test* de hecho tendrán la enfermedad X? __ de cada__.

Con este nueva versión del experimento los resultados fueron dramáticamente diferentes. 38 de los 50 sujetos, es decir, 76 por ciento de los encuestados dieron con la respuesta correcta.

Como apoyo extra a estas conclusiones, Cosmides y Tooby hacen mención de otros sorprendentes resultados reportados por otros investigadores. Por ejemplo, Fiedler (1988), siguiendo los resultados de Kahneman y Tversky, señaló que el porcentaje de sujetos que cometen la falacia de la conjunción puede ser radicalmente reducida si el problema se presenta en términos frecuentistas. La versión frecuentista del experimento de "Linda, la cajera feminista" de Fiedler es el siguiente:

> Linda tiene 31 años, es soltera, abierta y muy brillante. Ella estudió filosofía y como estudiante estuvo profundamente preocupada por asuntos de discriminación y justicia social, también participó en manifestaciones antinucleares.
>
> Hay 100 personas que cumplen con la descripción arriba presentada. ¿Cuántas de ellas son:
> (a) maestras de primaria?
> (b) feministas?
> (c) feministas y cajeras en un banco?

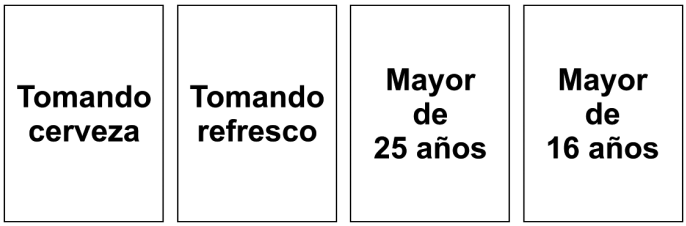
En una réplica de los experimentos de Kahneman y Tversky, usando la formulación original, el 91 por ciento de los sujetos consideró que la opción "feminista y cajera de un banco" es más probable que la opción "solamente cajera". Sin embargo, en la versión frecuentista del mismo experimento sólo el 22 por ciento de los sujetos juzgaron que la opción de "feminista y cajera de un banco" es más probable que la opción "solamente cajera". En otros experimentos, Gigerenzer (1994) planteó a los sujetos que 200 mujeres coincidían con la descripción de "Linda" y les preguntó que estimaran el número de aquellas que fueran cajeras de un

banco, feministas y cajeras de un banco, y solamente feministas. Los resultados, según Gigerenzer, fueron que sólo el 13 por ciento cometieron la falacia de la conjunción.

III.4. LA HIPÓTESIS DE LA DETECCIÓN DEL TRAMPOSO

En el capítulo anterior se expuso una versión de una tarea de selección introducida por Wason, en donde la ejecución de los sujetos es muy pobre. Sin embargo, hay algunas versiones de la tarea de selección donde los sujetos mejoran su ejecución de manera importante. El siguiente es un ejemplo de Griggs y Cox (1982):

> Como una medida severa en contra de los conductores en estado de ebriedad, los oficiales de Massachussets están revocando licencias para el expendio de bebidas alcohólicas. Eres un portero de un bar de Boston, y perderás tu trabajo a menos que refuerces el cumplimiento de la siguiente ley: "si una persona toma cerveza, entonces debe tener más de veinte años de edad".
>
> Las tarjetas de abajo muestran información acerca de cuatro personas sentadas en una mesa del bar donde laboras. Cada tarjeta representa sólo a una persona. De un lado de la tarjeta aparece lo que la persona está tomando y del otro lado aparece la edad de la persona. Indica qué tarjetas tienes que necesariamente voltear para ver si algunas de estas personas está violando la ley.



Desde un punto de vista lógico, este problema es estructuralmente idéntico al problema expuesto en el capítulo anterior, pero su ejecución produce resultados significativamente distintos. Alrededor del 75 por ciento de estudiantes universitarios obtuvieron una respuesta correcta, a diferencia del 25 por ciento de estudiantes universitarios que tienen una respuesta correcta en la versión de la tarea de selección del capítulo anterior (Samuels y Stich,

2004; Evans, 2001). Aunque ha habido cientos de estudios explorando este efecto de contenido en las tareas de selección, los resultados han sido y continúan siendo desconcertantes debido a que no parece haber una propiedad o un conjunto de propiedades compartidas por aquellas versiones de la prueba en las que los sujetos tienen una buena actuación. En diversos y ampliamente discutidos artículos, Cosmides y Tooby (1992) han argüido que un análisis evolucionista ofrece un patrón de explicación para estos por demás desconcertantes resultados.

El punto de partida del análisis evolucionista es la observación de que en el AAE de los ancestros homínidos eran comunes los casos en los que los individuos se encontraban en situaciones de intercambio en las que el beneficio que alguien recibía de otro individuo podía ser significativamente mayor que el costo del donante. En una sociedad de cazadores y recolectores había ocasiones en que un cazador podía tener un día afortunado y tuviera comida en abundancia, mientras que ese mismo día otro cazador no tuviera suerte y estuviera cerca de morir de hambre. Si el cazador exitoso diera algo de comida al cazador que no ha tenido éxito en lugar de "atiborrarse" de comida, esto podría tener un mínimo efecto negativo para el donante, puesto que los nutrientes que pudo haber recibido por el exceso de comida podrían ser de utilidad en el futuro, mientras que para el que recibe la comida el beneficio podría ser mucho mayor, pues quizá, de no recibirla, podría morir de hambre. En muchas ocasiones las personas ayudan incluso a individuos con los que no tienen parentesco, e incluso existe evidencia que sugiere que los primates no humanos hacen frecuentemente lo mismo. Este tipo de altruismo presenta una encrucijada evolucionista, debido a que si el gen que hace que un organismo esté menos dispuesto a ayudar a otros individuos (especialmente a aquellos con los cuales no tenga parentesco) aparece en la población, entonces dichos organismos serán mucho más aptos, y así este gen se propagaría gradualmente en toda la población (Dawkins, 1985).

Una solución a este problema lo propuso Robert Trivers (1971) quien notó que si bien el altruismo es una mala idea desde un punto de vista evolucionista, el *altruismo recíproco* tiene resultados muy distintos [16]. Si un par de cazadores puede contar cada uno con la ayuda del otro cuando uno tiene abundancia de comida,

mientras el otro no ha tenido éxito, entonces dichos cazadores se encontrarán a la larga en una mejor situación que aquellos cazadores que no presenten este comportamiento. Aquellos organismos con los genes o conjunto de genes que los inclinen a tener intercambios recíprocos, o "intercambios sociales", con organismos que no pertenezcan a su parentesco, serán con el paso del tiempo más aptos que los organismos de su misma especie que no tengan estos genes. Pero, por supuesto, los intercambios sociales son vulnerables al engaño. Debido a que el ser humano es capaz de tener relaciones estables de intercambio social, Cosmides y Tooby (1992) conjeturan que los seres humanos cuentan con uno o más módulos darwinianos cuyo trabajo es reconocer los acuerdos de intercambios recíprocos y detectar a los mentirosos o tramposos que aceptan los beneficios de tales acuerdos pero que no están dispuestos a pagar el costo que los mismos conllevan. De este modo, Cosmides y Tooby (1992) a través del análisis evolucionista, presentan como hipótesis la existencia de uno o mas módulos de detección de tramposos. A esta hipótesis se le denomina como la "hipótesis de la detección del tramposo".

Siguiendo el programa de investigación de la psicología evolucionista, si la hipótesis de la detección del tramposo es correcta, entonces se debe encontrar evidencia en los seres humanos contemporáneos que señale la existencia de un módulo para detectar mentirosos, y es precisamente aquí donde la tarea de selección interviene. Para Cosmides y Tooby (1992), algunas versiones de los experimentos en donde los sujetos se enfrentan con tareas de selección interviene un módulo mental diseñado para detectar a los tramposos en situaciones de intercambio social. Obviamente este módulo fue producto de la selección y en ese sentido, según los psicólogos evolucionistas, labora de manera eficiente y adecuada cuando realiza tareas de selección donde se involucra, no así en aquellas tareas de selección que no ponen en marchar a este módulo. Dado que los seres humanos no cuentan con módulos diseñados para resolver estas últimas tareas, cuando las mismas se presentan, resultan más complicadas y sus resultados son menos favorables que en aquellas tareas donde sí interviene el módulo de detección de tramposos. El experimento de selección de tarjetas que presenté en el capítulo anterior ejemplifica una tarea de selección donde no se involucra el módulo de detección

del mentiroso. A diferencia de dicho experimento, el caso del portero en el bar de Boston, presentado al inicio de esta sección, ilustra una tarea de selección donde, según los psicólogos evolucionistas, sí se activa el mecanismo de detección del tramposo.

Cosmides y Tooby ofrecen un importante cuerpo de evidencia en apoyo a la tesis de que la ejecución de tareas de selección donde interviene el módulo de detección del mentiroso mejora considerablemente. A juicio de los psicólogos evolucionistas, la hipótesis de la detección del tramposo explica el claro patrón de experimentos confusos que se habían desarrollado a partir de las tareas de selección antes de que la hipótesis evolucionista fuera formulada. Cuando en 1983 hubo un auge en los experimentos relacionados con las tareas de selección en torno a los trabajos de Wason surgió una gran cantidad de reportes con una amplia variedad de efectos en el contenido, y no existía una teoría que pudiera explicar estos efectos. Una vez que la formulación de la hipótesis de la detección del mentiroso fue formulada, un número importante de experimentos ha sido desarrollado para probar esta hipótesis y rechazar alternativas. Entre los más importantes y persuasivos experimentos se encuentran los trabajos de Gigerenzer y Hug (1992). Al contrastar su posición con la de Cosmides y Tooby, estos psicólogos intentan mostrar en una serie de experimentos que el mero hecho de percibir una regla como parte de un contrato social no es *suficiente* para poner en marcha el mecanismo cognitivo que conduce a una buena actuación en las tareas de selección, pues requiere de la posibilidad de dar pie a una trampa. Para mostrar lo anterior, Gigerenzer y Hug crearon dos diferentes contextos en donde intervienen reglas dentro de contratos sociales. Uno de los contextos del experimento requería que los sujetos atendieran a la posibilidad de la trampa o engaño, mientras que el contexto del otro experimento, la trampa o engaño, no era relevante. Entre las reglas del contrato social que fueron usadas en los experimentos se encuentra la siguiente norma, que es ampliamente conocida entre los excursionistas:

> Si alguien pasa la noche en una cabaña, entonces esa persona debe traer consigo un paquete de madera desde el valle.

La primera versión del contexto para el experimento, que los investigadores denominan "versión tramposa" es la siguiente:

> Hay una cabaña a cierta altitud en los Alpes Suizos que sirve a los excursionistas como un refugio para pasar la noche. Dado que hace frío y no hay leña disponible a esa altitud, la regla es que cada excursionista que pase una noche en dicha cabaña debe traer consigo su propia leña.
>
> Hay rumores de que la regla no es siempre respetada. Los sujetos en el experimento deben tener la perspectiva del guardia que debe averiguar cuál de los cuatro excursionistas ha violado la regla. Para los cuatro excursionistas hay una tarjeta que dice "permanece una noche en la cabaña", "no trae consigo leña", "trae consigo leña" y "no permanece una noche en la cabaña".

La segunda versión del contexto del experimento, que los investigadores denominan la "versión no tramposa", consiste en lo que sigue:

> Los sujetos en esta versión deben tener la perspectiva de ser miembros de la Asociación Alemana de Alpinismo que visita una cabaña suiza y tratan de entender cómo es que el Club Suizo de Alpinismo maneja dicho local. Un miembro de la Asociación Alemana observa que la gente trae leña a la cabaña, sugiriendo con ello que es posible que todos los que permanezcan en la cabaña deben traer su propia leña. Aunque es posible, en el experimento se les menciona esta alternativa, que los miembros del Club Suizo de Alpinismo que no pasen la noche en la cabaña podrían también traer consigo leña. La labor de los sujetos en el experimento es indagar cuál de las cuatro personas, cada una con su respectiva tarjeta, viola la regla sugerida por el miembro del la Asociación Alemana de Alpinismo.

La propuesta de Gigerenzer y Hug predice que a partir de la hipótesis de la detección del tramposo los sujetos tendrán mejores resultados en la "versión tramposa" del experimento que en la "versión no tramposa". Los resultados apoyan la propuesta de Gigerenzer y Hug (1992), ya que en la "versión tramposa" el 89 por ciento de los sujetos tuvo una respuesta correcta, mientras que en la "versión no tramposa" sólo el 53 por ciento de los sujetos obtuvo la respuesta correcta.

III.5. LA CONCLUSIÓN OPTIMISTA DE LA PSICOLOGÍA EVOLUCIONISTA

Es necesario hacer un recuento general de lo expuesto hasta aquí. Tanto la hipótesis frecuentista como la hipótesis de la detección del mentiroso propuestas por los psicólogos evolucionistas ofrecen una respuesta al problema que surge del estudio descriptivo de la razón, esto es, buscan especificar los mecanismos cognitivos que subyacen a la capacidad de razonar. La teoría de la modularidad masiva provee una idea general acerca de la arquitectura de la mente, así como un esquema general de los mecanismos cognitivos. La hipótesis frecuentista y la hipótesis de la detección del mentiroso, por su parte, hacen afirmaciones específicas acerca de dos mecanismos particulares de razonamiento modular.

Las hipótesis de los psicólogos evolucionistas aquí revisadas han sido ampliamente discutidas, por lo que es posible preguntarse: ¿qué tan verosímiles son estas tres hipótesis? Tanto la hipótesis frecuentista, como la hipótesis de la detección del mentiroso han incorporado un gran número de evidencia empírica a través de experimentos en el razonamiento, por lo que, a mi juicio, sería difícil rechazar dichas hipótesis únicamente de manera *a priori*. Por otra parte, gran parte del apoyo empírico que ha recibido la teoría de la modularidad masiva proviene no sólo de los trabajos señalados anteriormente, sino de campos de estudio tan diversos como la neuropsicología o los estudios psicológicos del desarrollo cognitivo sobre la "teoría de la mente". Samuels (2000) o Carruthers (2006) han argumentado, en algunos puntos de manera convincente, que de momento no existen buenas razones para rechazar la teoría de la modularidad masiva.

Algunos filósofos (Samuels, Stich y Faucher, 2004) consideran que aun cuando las hipótesis frecuentista y la detección del mentiroso, junto con la hipótesis de la modularidad masiva, sean candidatos verosímiles a explicar parte de la cognición humana, esto no significa que dichas hipótesis estén altamente confirmadas. No obstante, considero que las tres hipótesis aquí expuestas pertenecen a un rango de alternativas plausibles. Si bien, por ejemplo, existe evidencia psicológica que es compatible con la hipótesis de la detección del mentiroso o el tramposo, existen otras propuestas que han tratado de dar cuenta de esta evidencia. La propuesta de que los seres humanos poseen un módulo innato

y específico de dominio de razonamiento deóntico (Cummins, 1996) es una alternativa que podría explicar los datos psicológicos ofrecidos que apoyan a la hipótesis de la detección del tramposo. Otra propuesta que se ha sugerido para explicar las desviaciones en la ejecución del razonamiento, sin apelar a módulos darwinianos, es la propuesta de los esquemas de razonamiento pragmático [17] (Cheng y Holyoak, 1985).

En el campo filosófico es importante conocer cuáles son las implicaciones acerca de la racionalidad humana si la hipótesis de la modularidad masiva u otras hipótesis defendidas por los psicólogos evolucionistas fuesen correctas. Indagaré, primero, cuáles serían las implicaciones de aceptar las hipótesis defendidas por los psicólogos evolucionistas, especialmente aquellas relacionadas con el estudio de la racionalidad, para posteriormente, en las conclusiones de este capítulo, estudiar cómo tales hipótesis se vinculan con la noción de justificación. ¿Realmente la psicología evolucionista muestra que la interpretación pesimista de la racionalidad humana está equivocada? En la mayoría de escritos de los psicólogos evolucionistas se sugiere que sus descubrimientos y teorías indican que el razonamiento humano no está subsumido por heurísticas "rápidas y sucias" sino por "elegantes mecanismos" diseñados y refinados por la selección natural a lo largo de millones de años. De acuerdo con esta visión de la racionalidad humana, las preocupaciones en torno a la sistemática irracionalidad están infundadas. Comides y Tooby, junto con Gigerenzer, han explícitamente sostenido que no es necesario preocuparse por la racionalidad humana, ya que desde la psicología evolucionista se apoya una visión humana optimista o "básicamente racional" (Gigerenzer, 1991). Al igual que en el capítulo anterior, distinguiré aquí dos versiones de la "conclusión optimista". La primera de estas versiones sugiere que:

A) Muchos de los casos en los cuales los juicios de los sujetos cognoscentes concuerdan con las normas de racionalidad apropiadas son explicadas por el hecho de que los seres humanos activan módulos mentales que fueron diseñados por selección natural, y tales módulos hacen un buen trabajo cuando son provistos de los insumos adecuados.

Según esta primera versión, las tareas de razonamiento que ponen en marcha los mecanismos modulares darwinianos no se desvían de los principios normativos adecuados.

Una segunda versión de la conclusión defendida por los psicólogos evolucionistas es:

B) Todo (o casi todo) el razonamiento y toma de decisiones está subsumido en módulos normativamente no problemáticos, diseñados por la selección natural.

Si esta versión es correcta, entonces estaría infundada cualquier preocupación acerca de la sistemática irracionalidad de los seres humanos [18]. Las dos versiones de la conclusión optimista defendidas por la psicología evolucionista tienen en común hacer referencia a los mecanismos que subyacen al razonamiento humano, pero tienen un énfasis diferente en el número de mecanismos que satisfacen los principios normativos de razonamiento. Mientras que en la primera versión la satisfacción de los principios normativos por parte de los sujetos cognoscentes se explica porque los mecanismos de razonamiento hacen un buen trabajo sólo si son provistos los insumos adecuados; en la segunda versión se afirma que el razonamiento está determinado por mecanismos normativamente no problemáticos, por lo que está garantizada la satisfacción de los principios normativos de razonamiento por parte de los sujetos cognoscentes [19].

Con este telón de fondo de las distintas versiones de la conclusión optimista de la psicología evolucionista, es pertinente preguntase si realmente la psicología evolucionista, con cada una de sus hipótesis, contrarresta la tesis pesimista de la racionalidad humana expuesta en capítulo anterior.

III.6. LA PSICOLOGÍA EVOLUCIONISTA Y LA INTERPRETACIÓN PESIMISTA DE LA RACIONALIDAD HUMANA

En el capítulo anterior distinguí dos versiones de la interpretación pesimista. Estas dos versiones son las siguientes:

I. Muchas veces en que los juicios se desvían de las normas apropiadas de racionalidad son explicadas por el hecho de que los juicios y decisiones hechos por los sujetos se fundan en heurísticas que algunas veces conducen a errores sistemáticos y severos.

II. Los únicos mecanismos disponibles a los sujetos cognoscentes son heurísticas normativamente problemáticas.

Si se acepta la versión (A) de la conclusión optimista de la racionalidad humana, esto es, que muchos de los casos en los cuales los juicios de los sujetos cognoscentes concuerdan con las normas de racionalidad apropiadas porque se fundan en módulos mentales que fueron diseñados por selección natural, entonces se debe rechazar la versión pesimista (II). Asumiendo que la teoría de la modularidad masiva, la hipótesis frecuentista y la hipótesis de la detección del mentiroso fuesen correctas, las últimas dos hipótesis muestran que al menos habría dos mecanismos de razonamiento que responden a dos competencias que no son problemáticas. La respuesta no parece ser clara con respecto a la versión (I) de la interpretación pesimista debido a que los defensores de la interpretación pesimista no parecen distinguir claramente entre las versiones (I) y (II). Algunos defensores de la interpretación pesimista han hecho afirmaciones que parecerían sugerir que están comprometidos con (II). Sin embargo, según se ha expuesto, la psicología evolucionista no muestra que los seres humanos no cometan errores de razonamiento, es decir, que aceptar la conclusión (A) no supone el rechazo de la versión pesimista (I), es decir, (A) y (I) pueden ser compatibles. Aun cuando la psicología evolucionista esté en lo correcto al mostrar que existen algunos mecanismos que dan cuenta de competencias de razonamiento no problemáticas, no se sigue que los seres humanos no cometan errores de competencia o errores de ejecución fundados en competencias normativamente problemáticas.

> La afirmación de que poseemos competencias de razonamiento normativamente apropiadas, para el razonamiento frecuentista, la detección del tramposo y quizá otras tareas, es perfectamente compatible con la afirmación de que poseemos *otros* módulos de razonamiento que despliegan principios normativamente problemáticos que producen errores de competencia (Samuels, Stich y Faucher, 2004, p. 29 [20], la traducción es mía).

Si, como creen los psicólogos evolucionistas, la hipótesis de la modularidad masiva es correcta, entonces es posible que haya un número importante de mecanismos de razonamiento que la psi-

cología evolucionista aún no ha descubierto. Sin embargo, no hay razones para pensar que estos módulos den cuenta de competencias de razonamiento normativamente no problemáticas, ni para pensar que de ellos se tengan ejecuciones de razonamiento que satisfagan los principios normativos adecuados. La psicología evolucionista sostiene que la selección natural ha equipado al ser humano con un número importante de mecanismos de razonamiento que reflejan el empleo de principios normativos apropiados sobre la base del tipo de problemas que fueron importantes en el AAE. Aun si este tipo de consideraciones evolucionistas justifican la afirmación de que el ser humano posee principios normativos apropiados para lidiar con el tipo de problemas que fueron importantes en el AAE, es claro que hay muchos problemas de razonamiento que son importantes en el mundo moderno —por ejemplo, problemas que requieren de la probabilidad de sucesos singulares— con los cuales no pueden lidiar los mecanismos cognitivos del ser humano (Stanovich y West, 2003). En estos casos, incluso la psicología evolucionista sugiere que los módulos darwinianos no serían capaces de procesar este tipo de problemas. De donde, aceptar la versión optimista (A) conduce al rechazo de la versión pesimista (II), pero no de la versión pesimista (I).

Por último, la aceptación de la versión más optimista de la psicología evolucionista, esto es, la versión (B) que sostiene que casi todo el razonamiento humano está subsumido en una "elegante maquinaria" normativamente no problemática diseñada por la selección natural, supondría el rechazo de las versiones pesimistas (I) y (II) de la racionalidad humana. Aceptar que la cognición humana se funda en mecanismos normativamente no problemáticos supondría la falsedad de que todas o sólo algunas de sus competencias de razonamiento violan los principios normativos —es decir, la versión pesimista (II) y (I) respectivamente. De este modo, si se acepta (B), entonces se tendrían contrarrestadas de manera completa las interpretaciones pesimista (I) y (II) de la racionalidad humana.

Por lo visto anteriormente, con relación a la evaluación del razonamiento humano, es posible que la versión pesimista (I) junto con la versión optimista (A) sean todas verdaderas, pero no es posible que la versión (B) lo sea al mismo tiempo que las versiones pesimistas (I) y (II). Sin embargo, aun suponiendo que

la hipótesis de la modularidad masiva, así como la hipótesis de la detección de tramposo y la tesis frecuentista fuesen correctas, de ellas no se podría sostener la interpretación optimista (B). Las hipótesis frecuentista y la de detección del tramposo supondrían la existencia de sólo dos mecanismos en los cuales el razonamiento humano satisface los principios normativos adecuados, pero no que todos o casi todos los mecanismos de razonamiento no sean normativamente problemáticos. Con ello, si bien la versión optimista (B) supondría la falsedad de las dos distintas versiones de la interpretación pesimista de la racionalidad humana, aun suponiendo la verdad de las hipótesis defendidas por la psicología evolucionista, éstas no podrían justificar la verdad de la versión optimista (B). Si bien (B) y la interpretación pesimista no podrían ser verdaderas al mismo tiempo, (B) y la interpretación pesimista (II) sí podrían ser incorrectas al mismo tiempo.

> No hay nada en la tradición de heurística y sesgo que apoye la afirmación de que las heurísticas problemáticas son las *únicas* fuentes del razonamiento a las que los sujetos pueden recurrir, ni tampoco esta literatura provee alguna razón para pensar que no existen mecanismos normativamente no-problemáticos como los postulados por la psicología evolucionista. Por otro lado, los psicólogos evolucionistas no han ofrecido razones para pensar que *todo* nuestro razonamiento está subordinado a módulos normativamente no problemáticos (Samuels y Stich, 2004, p. 297, la traducción es mía).

Para recapitular, si la concepción del razonamiento humano que brinda la psicología evolucionista es correcta, ésta apoya a la versión optimista del razonamiento humano (A), pero no a la versión (B). Si se acepta (A), entonces es posible rechazar la versión pesimista (II) de la racionalidad humana que sostiene que todas las competencias de razonamiento que los seres humanos poseen son normativamente problemáticas. Pero si los defensores de la tradición heurística y sesgo defienden solamente la versión pesimista (I), entonces la psicología evolucionista realmente no ofrece razones para sostener que los seres humanos no cometen errores producto de heurísticas. Aun dando por supuestas las hipótesis más importantes de la psicología evolucionista, éstas resultan compatibles con la verdad de que los seres humanos se desvían de las normas apropiadas de la racionalidad [21]. Hay que dejar en

claro que lo que sostengo aquí es que la interpretación optimista (A) defendida por la psicología evolucionista puede ser compatible con la interpretación pesimista (I) defendida por la tradición de heurística y sesgo, y no que las teorías descriptivas defendidas por estas dos corrientes psicológicas sean en sí mismas compatibles[22].

III.7. CONCLUSIONES

En este capítulo he explorado una segunda teoría descriptiva del razonamiento, esto es, la psicología evolucionista. De esta teoría he destacado tres hipótesis centrales: *i.)* la hipótesis de la modularidad masiva, según la cual la mente humana está compuesta en su totalidad o en su mayoría por módulos darwinianos específicos de dominio; *ii.)* la hipótesis de la detección del tramposo o mentiroso, y *iii.)* la hipótesis frecuentista. Estas tres hipótesis, junto con un programa de investigación centrado en consideraciones adaptacionistas se presentan como una teoría rival de la tradición de heurística y sesgo. La psicología evolucionista defiende una interpretación optimista de la racionalidad humana, que puede ser entendida de distintas maneras. Una primera versión optimista (A) defendida por los psicólogos evolucionista afirma que muchos de los casos en los cuales los juicios de los sujetos cognoscentes concuerdan con las normas de racionalidad apropiadas son explicadas porque se activan módulos mentales que fueron diseñados por selección natural y tales módulos hacen un buen trabajo cuando son provistos de insumos adecuados. Una segunda versión optimista (B) defiende que todo (o casi todo) el razonamiento y toma de decisiones está subsumido en módulos normativamente no problemáticos diseñados por la selección natural. Si la versión (B) fuese correcta, estaría infundada cualquier preocupación acerca de la sistemática irracionalidad de los seres humanos. Según lo revisado en este capítulo, la evidencia proporcionada por los psicólogos evolucionistas no es suficientemente fuerte para sostener la versión optimista (B), pero sí para apoyar la versión (A). Sin embargo, la versión optimista (A) no es suficiente para eliminar por completo la interpretación pesimista de la racionalidad, puesto que habría al menos una versión pesimista que sería compatible con esta versión optimista proveniente de la psicología evolucionista, a saber, la interpretación pesimista

(I) que sostiene que muchas veces los juicios se desvían de las normas apropiadas por el hecho de que se fundan en heurísticas que conducen a errores sistemáticos de razonamiento. De este modo, la psicología evolucionista, de ser correcta, no demuestra que es errada la interpretación pesimista de la racionalidad humana, pero sí demuestra que, al menos, la versión más radical de la interpretación pesimista no puede sostenerse, i.e., que los únicos mecanismo disponibles a los seres humanos son heurísticas normativamente problemáticas.

En tanto que la psicología evolucionista defiende una versión optimista de la racionalidad humana no sólo ofrece una teoría descriptiva de los mecanismos que subyacen al razonamiento, sino una teoría evaluativa de la razón que, a su vez, requiere de una teoría normativa de la racionalidad. Al igual que la tradición de heurística y sesgo, la psicología evolucionista se compromete con la visión estándar de la racionalidad, según la cual los criterios normativos para evaluar el razonamiento humano provienen de los principios de la lógica y las matemáticas. Lo que la hipótesis de la detección del mentiroso y la hipótesis frecuentista muestran es que, dado los insumos adecuados, los mecanismos de razonamiento satisfacen los principios normativos provenientes de la probabilidad y la lógica. Así, la psicología evolucionista y la tradición de heurística y sesgo son dos teorías descriptivas del razonamiento distintas, que comparten la misma teoría normativa de la racionalidad, esto es, la visión estándar de la racionalidad. Un elemento que en principio no comparten la psicología evolucionista y la tradición de heurística y sesgo es su teoría evaluativa de la razón. Partiendo de descripciones del razonamiento distintas y de la misma visión de la racionalidad, los psicólogos evolucionistas defienden una conclusión optimista de la racionalidad humana, mientras que los defensores de la tradición de heurística y sesgo defienden una conclusión pesimista.

Estando en este punto, puede preguntarse, tal y como se hizo en el capítulo anterior, cuál es la relación entre la psicología evolucionista en tanto teoría del razonamiento con la noción de justificación epistémica. En general, ¿cuáles serían las consecuencias epistemológicas en teoría de la justificación, si se acepta la visión optimista de la racionalidad humana defendida por la psicología evolucionista? Para contestar esta pregunta regresaré

a las conclusiones más importantes del capítulo anterior, a saber, *i.)* la interpretación pesimista más radical (o la versión (II)) eliminaría cualquier posibilidad de construir una teoría de la justificación, y *ii.)* la tradición de heurística y sesgo no apoya claramente a ninguna noción de justificación. Veamos.

En primer lugar, en las conclusiones del capítulo anterior argumenté que si la interpretación pesimista (II) o más radical es correcta, esto es, que el grueso de las creencias de un individuo no son racionales, entonces no sólo se eliminaba cualquier noción de justificación, sino la noción tradicional de conocimiento, en tanto que supone que un elemento necesario para conocer es estar justificado. Sin embargo, si la psicología evolucionista está en lo correcto, entonces existe evidencia empírica que parece apoyar una conclusión optimista del razonamiento humano. Como lo he señalado, la versión optimista (A) bastaría para eliminar la verdad de la interpretación pesimista (II). De donde, si la psicología evolucionista es correcta, habría motivos para pensar que la interpretación pesimista (II), que conduce a un cierto tipo de escepticismo epistemológico, es inadecuada. Esta conclusión es importante porque implica que es posible desarrollar una teoría de la justificación, que resultaría imposible si la interpretación pesimista (II) fuese correcta.

En segundo lugar, sostuve que la tradición de heurística y sesgo no brinda apoyo a una noción internista ni a una noción externista de la justificación. Por tal motivo, cabe preguntarse: ¿existen conexiones teóricas importantes entre la psicología evolucionista y alguna noción de justificación? La noción internista de la justificación, como ya he expuesto, sostiene que los rasgos que justifican a una creencia pueden ser accesibles a los sujetos cognoscentes. Por su parte, la psicología evolucionista sostiene que los mecanismos de razonamiento son modulares y se desempeñan de manera apropiada cuando son provistos de los insumos adecuados. En principio, no parece haber alguna relación importante entre esta teoría epistemológica y la teoría psicológica que he expuesto en este capítulo. Las dos teorías podrían ser verdaderas al mismo tiempo y no parece haber alguna implicación lógica entre alguna de sus ideas centrales. La importancia de una postura como la defendida por los psicólogos evolucionistas serviría simplemente, desde la perspectiva internista, para posibilitar la

idea de que la construcción de una teoría de la justificación es posible ya que los sujetos son racionales.

Los psicólogos evolucionistas sostienen que módulos de razonamiento como el de la detección del tramposo o de razonamiento inductivo en formato de frecuencias son dos muestras de procesos adecuados de razonamiento, porque brindan resultados que satisfacen los estándares normativos provenientes de la visión estándar de la racionalidad. Al mismo tiempo, dichos módulos son procesos cognitivos eficaces en tanto que son producto de la evolución y ayudaron al ser humano a resolver problemas adaptativos en el AAE. Visto de esta manera, se podría pensar que los mecanismos modulares son procesos confiables, en el sentido de que suelen ser eficientes para resolver problemas particulares. Este rasgo podría ser un punto de contacto entre la noción externista de la justificación y la descripción modular del razonamiento ofrecida por los psicólogos evolucionistas. Recuérdese especialmente que para la noción externista de justificación, un proceso confiable de formación de creencias ofrece *outputs* justificados. En este sentido, hipótesis como la detección del tramposo o la hipótesis frecuentista podrían ser ejemplos de procesos que si reciben los insumos adecuados ofrecen creencias racionales y, en tanto que son procesos confiables, también ofrecen creencias justificadas según los estándares externistas.

Como he argumentado, es difícil sostener que a partir de la evidencia empírica proporcionada por los psicólogos evolucionistas se pueda sostener que todos los mecanismo de razonamiento que posee el ser humano satisfacen los estándares normativos adecuados, sino que sólo se podría sostener la versión optimista (A), esto es, la versión que es compatible con la versión débil de la interpretación pesimista de la racionalidad. Si lo anterior es el caso, una visión externista que se funde en la psicología cognitiva tendría que seguir de cerca el trabajo que los psicólogos evolucionistas realizan. En especial, aquel trabajo que distingue entre aquellos mecanismos cognitivos que satisfacen los estándares normativos y brindan creencias racionales, de aquellos que no lo satisfacen y producen creencias que no son racionales. Por ejemplo, según lo visto en este capítulo, los sujetos cuentan con mecanismos modulares que funcionan adecuadamente cuando la información se presenta en formato de frecuencia, pero de manera

inadecuada si la información se presenta en porcentajes, en el primer caso ofrecería creencias racionales, y por ello justificadas, mientras que en el segundo caso se ofrecen creencias irracionales, e injustificadas, según los estándares externistas.

Hay un rasgo que puede hacer pensar que hay una afinidad teórica entre la noción externista de justificación y la visión modular de la mente defendida por los psicólogos evolucionistas. Generalmente, una de las característica de los módulos es que éstos están encapsulados y son inaccesibles (Fodor, 1983). Hay filósofos que han sostenido que la noción módulo dentro de la modularidad masiva puede conservar estos rasgos de la noción fodoriana de módulo (Carruthers, 2006). Si lo anterior es el caso, los seres humanos no necesariamente son conscientes de cómo procesan información los mecanismos modulares, pero ello no evita que tales procesos sean eficientes o satisfagan los estándares normativos apropiados. Este rasgo sugiere una similitud con el externismo, pues esta noción de justificación acepta que un sujeto puede estar justificado aun cuando dicho sujeto no tenga acceso a los mecanismos que produjeron sus creencias, ni tenga control consciente sobre los mismos (en términos generales, para un externista un sujeto puede estar justificado, sin creer justificadamente que lo está).

Prima facie, si la noción de modularidad defendida por los psicólogos evolucionistas implica que los módulos están informacionalmente encapsulados y son insensibles a las metas cognitivas centrales, podría entrar en conflicto con la noción internista de la justificación. Si la conclusión optimista defendida por los psicólogos evolucionistas es correcta, entonces los módulos ofrecen creencias racionales, y si uno de los rasgos de los módulos es que son inaccesibles y están informacionalmente encapsulados, entonces los módulos producirían creencias que son racionales y que estaría fuera del acceso de los sujetos. De donde, partiendo de la idea de que los módulos implican encapsulamiento e inaccesibilidad, la hipótesis de que la mente es modular chocaría con las intuiciones internistas que se encuentran detrás de la noción de justificación, i.e., la idea de que alguien está justificado, si es capaz de "dar y ofrecer" razones y tiene control consciente sobre sus creencias [23].

Sostener que la hipótesis de la modularidad masiva y la conclusión optimista de la racionalidad humana defendidas por la psicología evolucionista, por un lado, choca con algunas intuiciones internistas que están detrás de la noción de justificación, y por otro lado, muestra afinidad teórica con el externismo, parte de la idea de que la noción de módulo en la psicología evolucionista implica encapsulamiento e inaccesibilidad. ¿Realmente la noción de módulo defendida por los psicólogos evolucionistas supone estos rasgos? La noción de módulo defendida por Cosmides y Tooby es la noción darwiniana de módulo que postula mecanismos computacionales productos de la selección natural y específicos de domino. Los dos ejemplos de módulos darwinianos revisados en este capítulo, esto es, el módulo frecuentista y el módulo especializado en detectar tramposos son muestra clara de que la noción de módulo que los psicólogos evolucionistas conciben no tiene que asumir que éstos están aislados informacionalmente ni que están encapsulados. Filósofos como Samuels, Stich y Faucher (2004, nota. 8) consideran que generalmente los psicólogos evolucionistas defienden una noción de módulo que niega que el encapsulamiento informacional sea un rasgo necesario dentro de dicha noción [24]. Si esto es correcto, la teoría de la modularidad masiva defendida por los psicólogos evolucionistas no contradice las intuiciones internistas de la justificación y tampoco apoya una noción externista. Por lo tanto, las tesis centrales de la psicología evolucionista no pueden ser usadas para dar salida al debate entre nociones internistas y externistas de la justificación.

La implicación más importante de la psicología evolucionista que quiero recuperar es que ofrece una visión optimista de la racionalidad humana. Si la versión optimista (A) o la versión optimista más débil es correcta, entonces hay razones para rechazar la interpretación pesimista más fuerte que implica un tipo de escepticismo en torno a la noción de justificación. Sin embargo, según lo visto en estas conclusiones, no existe un vínculo teórico importante entre la psicología evolucionista y las teorías de la justificación. A primera vista, la teoría de la modularidad masiva parece tener afinidad teórica con el externismo y, hasta cierto punto, parece incompatible con el internismo, si se piensa que la noción de módulo defendida por la psicología evolucionista su-

pone encapsulamiento informacional e inaccesibilidad. En tanto no es claro que la noción de módulo defendida por los psicólogos evolucionistas implique necesariamente los rasgos anteriores, es difícil afirmar que la noción darwiniana de módulo implica un rechazo al internismo. De este modo, tanto la noción externista de la justificación como la noción internista son compatibles con la teoría descriptiva ofrecida por la psicología evolucionista. Esta teoría del razonamiento no parece tener consecuencias relevantes con relación al debate entre internistas y externistas, su importancia radica en que, si sus tesis centrales son verdaderas, entonces se puede eliminar la versión más fuerte de la interpretación pesimista de la racionalidad humana.

La compatibilidad entre las teorías de la justificación estudiadas y las teorías del razonamiento motivan a explorar una teoría del razonamiento que rechace la interpretación pesimista de la racionalidad y que pueda decir algo sobre el debate entre las nociones internistas y externistas de la justificación. Por el momento, es necesario reconocer que los argumentos ofrecidos por la psicología evolucionista son insuficientes para rechazar una de las interpretaciones pesimistas de la racionalidad humana, por lo que en el siguiente capítulo analizaré otras estrategias para responder a esta interpretación. Del mismo modo, en el siguiente capítulo expondré un acercamiento distinto a la racionalidad, pues hasta el momento sólo me he centrado en la visión estándar de la racionalidad con la que se comprometen las teorías del razonamiento hasta ahora revisadas.

IV.
LA INTERPRETACIÓN PESIMISTA DE LA EVIDENCIA EMPÍRICA Y LAS VISIONES DE LA RACIONALIDAD

El propósito de este capítulo es revisar distintas objeciones a la interpretación pesimista de la racionalidad humana. Anteriormente argumenté que la psicología evolucionista, de ser correcta, no socava una de las versiones de la interpretación pesimista de la racionalidad humana, por lo cual en este capítulo indagaré y evaluaré otras estrategias que han tratado de desafiar a tal interpretación. Estas estrategias u objeciones conducen finalmente al cuestionamiento del criterio con el cual se ha evaluado al razonamiento humano, esto es, a la visión estándar de la racionalidad que asumen tanto los psicólogos comprometidos con la tradición de heurística y sesgo, como aquellos comprometidos con la psicología evolucionista. La revisión de esta teoría normativa y su contrastación con una teoría rival de la racionalidad brindará nuevas herramientas para entender la relación entre la racionalidad y la justificación.

La estructura del capítulo es la siguiente. Exploraré la objeción "pragmática" que consiste, a grandes rasgos, en sostener que diversos factores pragmáticos han sido olvidados en los datos de los experimentos que son usados para mostrar que los seres humanos cometen sistemáticos errores de razonamiento. Un segundo tipo de objeciones que revisaré en este capítulo surge de la interpretación y aplicación de la visión estándar de la racionalidad misma. Como se mencionó en el segundo capítulo de este trabajo de investigación, la visión estándar sostiene que ser racional es razonar de acuerdo con los principios normativos fundados en las reglas de la lógica y la teoría de la probabilidad. Este tipo de objeciones señala que no existe un consenso de cómo interpretar las reglas de las teorías formales ni tampoco hay consenso de

cómo derivar principios normativos de tales reglas. En una última parte de este capítulo exploraré el rechazo a la interpretación pesimista de la racionalidad a través del rechazo mismo de la visión estándar. Como argumentaré, el rechazo de la visión estándar se funda en *a)* el rechazo del acercamiento deontológico de la racionalidad, y en *b)* el énfasis puesto en las limitaciones de recursos desde los que actúa el sujeto cognoscente. En las conclusiones, tal y como se ha hecho en los capítulos anteriores, exploraré las implicaciones de lo anterior para la noción de justificación.

IV.1. UNA RESPUESTA PRAGMÁTICA A LA INTERPRETACIÓN PESIMISTA

Una posible solución a la interpretación pesimista es la respuesta que ofrece Adler (1984) al revisar la falacia de la conjunción cometida en el caso de "Linda, la cajera feminista". Adler argumenta, que en este experimento los sujetos hacen uso de los "principios griceanos" de implicación conversacionista. Dichos principios pueden explicar el error de los sujetos que sostienen que un suceso compuesto (como la opción "h) Linda es cajera en un banco y activista en un movimiento feminista"), como más probable que un suceso singular (como la opción "f) Linda es cajera en un banco") [1]. A grandes rasgos, los principios griceanos incorporan una "máxima de relevancia", entendida como el principio de que "lo que es dicho" debe ser asumido como relevante en el contexto lingüístico específico en el cual es expresado. Para algunos psicólogos cognitivos, lo más importante de la máxima griceana es que al entender al hablante, el oyente no sólo debe comprender el significado de lo que es dicho, sino también lo que esté implicado en el contexto, presuponiendo que el hablante es un hablante cooperativo con el oyente (Stanovich y West, 2000). En el contexto del experimento de "Linda, la cajera feminista" significa que si los sujetos se comportan como la teoría griceana predice, ellos deben interpretar la tarea de decidir si "h)" es más o menos probable con respecto a "f)" teniendo en cuenta que la descripción de Linda es relevante.

Si los sujetos interpretan la tarea —de la manera en que lo hacen la tradición de heurística y sesgo— como:

1. El término "probable" funciona de acuerdo con los principios de la teoría de la probabilidad,
2. la opción "h)" tiene la forma lógica de (A & B) y
3. la opción "f)" tiene la forma lógica de A,

entonces la descripción de Linda no es relevante para determinar si la opción "h)" es más o menos probable que la opción "f)". En esta interpretación, el juicio de que la opción "f)" es más probable que la opción "h)" es simplemente una instancia de una verdad matemática, dado que para cualquier enunciado A y para cualquier enunciado B, la probabilidad de A será mayor o igual que la probabilidad del enunciado (A & B), es decir, P (A) ≥ P (A & B). Dando por supuesto que el conjunto de cajeros de banco no es vacío, ninguna información contingente acerca de Linda —incluso la que es proporcionada en el experimento— es relevante para solucionar la tarea en cuestión. Por el contrario, si los sujetos del experimento se comportan como buenos griceanos, ellos deben rechazar esta interpretación a favor de una interpretación alternativa, en la cual la descripción de Linda proporcionada en el experimento es realmente relevante. En este caso, afirma Adler (1984) dada la descripción de Linda, es mucho más probable que Linda sea una cajera feminista a que sea solamente una cajera sin ser feminista. De acuerdo con lo anterior, se podría sostener que los sujetos no violan la falacia de la conjunción, sino que proveen una respuesta correcta a la pregunta hecha en el experimento, de acuerdo a como dichos sujetos la han interpretado. Además, esta interpretación de los sujetos es explicada por el hecho de que están haciendo justo lo que la teoría pragmatista griceana dice que se debe hacer. Dado este punto de vista, los resultados provenientes de "Linda, la cajera feminista" no apoyan la afirmación de que los seres humanos realizan sistemáticos errores de razonamiento. El caso de Linda apoyaría la afirmación verosímil de que los seres humanos actúan con respecto a la máxima de relevancia [2].

Considero que la propuesta de Adler para explicar cómo los sujetos realmente no cometen un error de razonamiento en el caso de "Linda, la cajera feminista" resulta interesante. Sin embargo, esta propuesta se enfrenta a dos problemas importantes. El primer problema es que Kahneman y Tversky han hecho los controles necesarios en los experimentos en caso de que los sujetos

consideren que la descripción de Linda es relevante para realizar la tarea que les pide el experimento, i.e., que interpretan la opción "f)" como "Linda es una cajera en el banco y no es feminista". Sin embargo, dichos matices no modifican la ejecución de los individuos en el experimento de "Linda, la cajera feminista" (Stein, 1996). El segundo problema que enfrenta la explicación que ofrece Adler de los efectos de la conjunción es que crea un enigma con respecto al porqué las ejecuciones de los sujetos mejoran considerablemente en versiones frecuentistas de "Linda, la cajera feminista" –como sucede en los experimentos de los psicólogos evolucionistas expuestos anteriormente. A juicio de Samuels, Stich y Faucher (2004), los principios griceanos parecen sugerir que el problema aparecía de la misma manera en las versiones de sucesos singulares como en las versiones frecuentistas ofrecidas por los psicólogos evolucionistas.

> De acuerdo con Adler, en el experimento original de eventos singulares, la descripción de Linda es irrelevante para ordenar las opciones (h) y (f). No obstante, en la versión frecuentista de dicha tarea la descripción de Linda también es irrelevante para decidir si más personas son cajeras feministas que solamente feministas. Así, la propuesta de Adler parece predecir que el efecto de la falacia de la conjunción también ocurrirá en la versión frecuentista del problema de la "cajera feminista". Pero, claramente, esto es precisamente lo que no ocurre (Samuels, Stich y Faucher, 2004, p. 33, la traducción es mía).

Es claro que la propuesta pragmatista de Adler debe ofrecer una explicación de por qué los sujetos cometen la falacia de la conjunción, como lo sugieren los principios griceanos, cuando se enfrentan con la versión de un evento particular, y no la cometen cuando se enfrentan a una versión frecuentista de la misma tarea. De lo anterior, es difícil entender cómo la objeción pragmatista puede ser una salida a la interpretación pesimista de la racionalidad humana.

IV.2. OBJECIONES RELACIONADAS CON LA INTERPRETACIÓN Y APLICACIÓN DE LA VISIÓN ESTÁNDAR DE LA RACIONALIDAD

> Quizá los únicos sujetos que sufren alguna illusión con relación a las ilusiones cognitivas son los psicólogos cognitivos (Ayton & Hardman).
> (Citado por Stanovich y West, 2000, la traducción es mía)

La visón estándar de la racionalidad, como reiteradamente he apuntado, sostiene que ser racional es razonar de acuerdo con los principios basados en las reglas de la lógica y la teoría de la probabilidad, en este sentido, los principios de razonamiento que están basadas en dichas reglas son los principios con los cuales los seres humanos deben razonar. Una objeción importante a la visión pesimista apunta que muchos de los defectos señalados por los psicólogos del razonamiento se fundan en la manera en que dichos psicólogos entienden y aplican esos principios normativos. En esta sección discutiré dos diferentes versiones de esta objeción. Una primera versión de esta objeción afirma que los defensores de la interpretación pesimista de la racionalidad usualmente adoptan "lecturas" específicas y altamente contenciosas de ciertos principios normativos, en especial, de los principios de la teoría de la probabilidad. Una segunda versión de esta objeción es lo que se ha denominado "problema de derivación", esto es, el problema de explicar cómo los principios normativos son derivados de los sistemas formales como la lógica o la teoría de la probabilidad. A continuación revisaré estas dos objeciones.

IV.2.1. EL RECHAZO DE INTERPRETACIONES NO FRECUENTISTAS DE LA TEORÍA DE LA PROBABILIDAD

Los resultados de los experimentos pueden considerarse, bajo una interpretación de los principios provenientes de las teorías formales, como violaciones a los principios normativos adecuados; mientras que, bajo una interpretación diferente, esos resultados no contarían como violaciones a dichos principios. Esta objeción ha sido ampliamente discutida en el campo de la probabilidad, donde existe un desacuerdo importante en torno a la interpretación del cálculo de probabilidad [3].

Kahneman y Tversky insisten en que la teoría de la probabilidad puede ser aplicada con completo sentido a sucesos

singulares, de ahí que los juicios sobre éstos pueden violar la descripción "subjetivista" o "bayesiana" de la probabilidad, que permite la asignación de probabilidades a sucesos singulares (Kahneman y Tversky, 1996). En contraste, Gigerenzer ha sostenido que la interpretación adecuada de la teoría de la probabilidad debe ser dada en términos frecuentistas, esto es, se debe ofrecer una interpretación frecuentista de acuerdo con la cual las probabilidades son construidas como frecuencias relativas de sucesos de una clase con respecto a eventos de otra clase. Gigerenzer sostiene que bajo la interpretación frecuentista

> no se puede hablar de probabilidad a menos que sea definida una clase de referencia... la frecuencia relativa de un suceso tal como la muerte es definida sólo con respecto a una clase de referencia tal como "todos los dueños de bares de cincuenta años viviendo en Bavaria" (Gigerenzer, 1993, p. 292, la traducción es mía).

Una consecuencia de dicha interpretación, que ciertamente Gigerenzer desea enfatizar, es que no tiene sentido asignar probabilidad a sucesos singulares. Gigerenzer subraya que las afirmaciones acerca de la probabilidad de eventos singulares son literalmente sin sentidos. A este respecto, este psicólogo afirma "el término 'probabilidad', cuando refiere a un evento singular, no tiene sentido para nosotros" (Gigerenzer, 1991, p. 88, la traducción es mía). Gigerenzer afirma, además, que si se sostiene esta interpretación, en sentido estricto, las leyes de la probabilidad dan cuenta únicamente de frecuencias y no de eventos singulares. De ahí que, agrega Gigerenzer (1993, p. 293), no habría juicios acerca de sucesos singulares que violen la teoría de la probabilidad.

Este desacuerdo acerca de la interpretación de la probabilidad surge de importantes y complejas cuestiones en los fundamentos de la estadística y la teoría de la decisión acerca de los alcances y límites del tratamiento formal de la probabilidad. La disputa entre frecuentistas y subjetivistas ha sido central en el debate sobre los fundamentos de la probabilidad en gran parte del siglo XX. No trataré aquí dicha disputa en la probabilidad, puesto que va más allá de los intereses propios de este trabajo. Sin embargo resulta importante comentar que resulta primordial el papel que juega la interpretación de la teoría de la probabilidad en el debate entre

los psicólogos evolucionistas y los defensores del programa de heurística y sesgo; en especial porque tanto los defensores de la tradición de heurística y sesgo, como los psicólogos evolucionistas comparten la visión estándar de la racionalidad, que ocuparán posteriormente para evaluar el razonamiento humano. Aunque se acepte que la visión estándar de la racionalidad es la teoría normativa adecuada, queda el problema de cómo deben ser interpretados los principios de las matemáticas (en este caso de la probabilidad). Gigerenzer interpreta los principios de la probabilidad de manera distinta a como lo hacen Kahneman y Tversky.

Algunos filósofos han argumentado que las consideraciones frecuentistas de las que hace uso Gigerenzer en esta disputa resultan ampliamente cuestionables (Piatelli-Palmarini, 2005). Como señalé anteriormente, Gigerenzer (1993) sostiene que si la interpretación frecuentista de la teoría de la probabilidad es correcta, entonces no tienen sentido las afirmaciones acerca de la probabilidad de los sucesos singulares; de ahí que los juicios acerca de la probabilidad de sucesos singulares no pueden violar la teoría de la probabilidad. Lo que claramente busca Gigerenzer con este argumento es socavar parte de la evidencia que es ocupada por los defensores de la tradición de heurística y sesgo para señalar que el razonamiento humano viola las normas probabilísticas apropiadas. De ese modo, Gigerenzer busca excluir los resultados experimentales que hagan uso de juicios acerca de eventos singulares como evidencia para la existencia de juicios y razonamientos probabilísticos normativamente problemáticos.

Piatelli-Palmarini (2005) considera que si bien esta estrategia de Gigerenzer parece persuasiva, se enfrenta a serias objeciones. En primer lugar, es necesario indicar que el frecuentismo mismo es una visión debatible en la teoría de la probabilidad, por lo que sería aventurado argumentar a favor de una postura que se apoye en él como lo hace Gigerenzer.

> En cualquier caso queda aún por demostrar que "no tiene sentido" hablar de la probabilidad de un solo acontecimiento. A todos nosotros, intuitivamente, nos parece que lo tiene, y algunos teóricos de la probabilidad nos lo confirman. Estar sometidos a las ilusiones cognitivas a la hora de *cómo* valorar esta probabilidad es una cosa; negar la validez misma de cualquier estimación de esta probabilidad, como

> lo están haciendo Gigerenzer y algunos otros, es otra cosa totalmente diferente. El valor de los descubrimientos de Tversky y Kahneman y de otros muchos, resiste bien ante estas y otras más antiguas objeciones (Piatteli-Palmarini, 2005, p. 76).

Pero aun suponiendo que el frecuentismo es una interpretación adecuada dentro de la teoría de la probabilidad, es posible sostener que la postura de Gigerenzer se enfrenta con dos problemas. El primero de ellos es que hay una tensión importante al sostener que *i.)* los sujetos cognoscentes *no cometen errores de razonamiento* en tareas donde intervienen sucesos singulares, en tanto que no tiene sentido calcular la probabilidad de sucesos singulares, según la postura frecuentista defendida por Gigerenzer, y *ii.)* la afirmación, frecuentemente hecha por los psicólogos evolucionistas y el mismo Gigerenzer, de que el razonamiento probabilístico de los seres humanos mejora considerablemente cuando los problemas a los que se enfrenta son presentados en formato de frecuencia. Sólo si hubiera algo malo en el razonamiento acerca de sucesos singulares, es como tiene sentido decir que los seres humanos tienen una mejor ejecución en problemas de razonamiento presentados en formato de frecuencia. En otras palabras, tiene sentido sostener "*ii.)*" sólo si los sujetos cognoscentes cometen errores de razonamiento (i.e., tiene sentido sostener "*ii.*)" si no es el caso que "*i.*)").

Relacionado con este problema, la propuesta de Gigerenzer se enfrenta a un segundo problema (Samuels, Stich y Faucher, 2004, Piatelli-Palmarini, 2005). Si los juicios acerca de la probabilidad de sucesos singulares carecen de sentido, entonces la respuesta a los problemas planteados por los psicólogos cognitivos acerca de la probabilidad de un evento singular no es un valor numérico o el puesto que ocupan en una escala, sino respuestas como: "¡eso no tiene sentido!" o "¿qué es lo que realmente quieres decir?". Samuels, Stich y Faucher (2004) ilustran lo anterior formulando preguntas como "¿es Linda más alta que?" o "¿qué tanto más alta es Linda que?". Las preguntas anteriores no tienen sentido porque están incompletas. Para poder contestarlas es necesario que tengan el correlato a "más alta que", a menos que esto sea hecho, contestar con un "sí", un "no" o proveer un valor numérico a la pregunta anterior sería normativamente inapropiado. Gigerenzer

parece señalar que la pregunta "¿qué tan probable es que Linda sea una cajera feminista?" no tiene sentido por las mismas razones que la pregunta "¿es Linda más alta que?" no tiene sentido. Así, parece que dar respuesta a sucesos singulares proveyendo un número o el puesto que ocupan dentro de una escala es problemático, pues la respuesta apropiada sería un "¿qué es lo que realmente quieres decir?". Irónicamente, Samuels, Stich y Faucher afirman:

> Los sujetos... saben, si bien sólo implícitamente, que las probabilidades de sucesos singulares son sinsentidos... Así, se podría pensar que el mensaje principal es: "no culpen a los sujetos por ofrecer respuestas absurdas. Culpen en primer lugar a los experimentadores por poner a los sujetos en situaciones absurdas" (Samuels, Stich y Faucher, 2004, p. 36, la traducción es mía).

De hecho, los sujetos comúnmente hacen uso de juicios acerca de la probabilidad de sucesos singulares en todo tipo de circunstancias tanto dentro como fuera del laboratorio de los psicólogos (Piatelli-Palmarini, 2005). De este modo, es difícil pensar que los sujetos ven a la probabilidad de un suceso singular como un sinsentido.

> Se pueden hacer contraobjeciones muy válidas a Gigerenzer y sus colaboradores. Básicamente, que a todos nosotros nos parece *extraordinariamente sensato* formular un juicio sobre la probabilidad de que un individuo ejerza una determinada profesión, tenga un cierto hobby, o desarrolle cierta actividad. Si, tal como afirma Gigerenzer, y como han sostenido también en alguna ocasión en el pasado cierto teóricos de la probabilidad, no tuviese sentido realmente estimar la probabilidad de *un* solo acontecimiento, en este caso esta sería probablemente la más *macroscópica* y la más *interesante* de todas nuestras ilusiones cognitivas. A nosotros, simples mortales, nos parece que precisamente sí *tiene* un sentido. No sólo esto, sino que estamos dispuestos a precisar *cuál* es, e incluso a apostar dinero constante y sonante por la probabilidad de que se produzca un único acontecimiento (Piatelli-Palmarini, 2005, p. 75).

Por último, se ha señalado que aun si los sujetos pensaran que la probabilidad de los sucesos singulares son sinsentidos, presumiblemente se debería esperar que las respuestas a las tareas hechas

por los psicólogos se ubicaran en un rango amplio de posibilidades y no el tipo de respuestas sistemáticas que se han observado en la literatura de la psicología cognitiva (Samuels, Stich y Bishop, 2002). Nuevamente, si se comparan las preguntas hechas por los psicólogos con "¿es Linda más alta que?" sería sorprendente que el resultado sistemáticamente fuera un "sí".

IV.2.2. EL PROBLEMA DE LA DERIVACIÓN

La idea de que los principios de razonamiento son derivados de los sistemas formales ha sido discutida ampliamente en la epistemología contemporánea (Goldman, 1986). Se ha argumentado que el problema de la "derivación" de los principios normativos de los sistemas formales presenta tantos problemas como para rechazar la visión estándar de la racionalidad. En este sentido, el problema de la derivación de los principios normativos es un intento por rechazar la interpretación pesimista de la racionalidad humana al mostrar que el criterio que asume dicha interpretación podría ser el apropiado, pero que de él no se pueden obtener los principios normativos que establecen cuándo se razona adecuadamente.

Uno de los primeros problemas con la visión estándar de la racionalidad consiste en que dentro de las teorías formales existen diversas maneras de entender lo que son la probabilidad y la lógica. La visión estándar de la racionalidad es imprecisa con respecto al modo en que se debe entender la matemática, la lógica y el cálculo de probabilidad, pues sólo se limita a afirmar que los estándares de razonamiento adecuados son aquellos que se fundan en los principios de una teoría formal. Pero dado que hay distintas maneras de entender a las diversas teorías formales y de ellas surgen principios, que algunas veces resultan incompatibles, surge el problema de indagar sobre cuáles principios formales se deben fundar los estándares de razonamiento. Un ejemplo claro se encuentra en la lógica, por un lado están el cálculo de predicados de primer orden, la lógica relevante, la lógica epistémica, la lógica paraconsistente, la lógica intuicionista y muchas más; si lo que la visión estándar de la racionalidad afirma es que los estándares normativos se derivan de los principios lógicos, surge el problema de cuál de las diversas maneras de hacer lógica deben tomarse los estándares normativos de razonamiento. En la sec-

ción anterior presenté otro ejemplo de esta dificultad, viz., Gigerenzer argumenta que la teoría de la probabilidad debe ser interpretada de una manera particular, distinta a la interpretación de Kahneman y Tversky. Tanto Gigerenzer como la tradición de heurística y sesgo comparten la visión estándar de la racionalidad; sin embargo, creen que los principios normativos adecuados se deben fundar en distintas interpretaciones de la misma.

Otro problema con el origen de los principios normativos es precisamente entender en qué sentido dichos principios se "derivan" de tales teorías. Suponiendo que se ha seleccionado un sistema de teorías canónicas como la fuente de las normas y reglas de razonamiento, entonces no queda claro cómo es que las normas de razonamiento se derivan de las teorías formales, pues ellas mismas no son lógicamente implicadas por las teorías (Goldman, 1986). De los axiomas y teoremas del cálculo de probabilidad no se sigue lógicamente que los sujetos deban razonar de acuerdo con lo que ellos expresan. Goldman (1986) ofrece dos razones para mostrar cómo de la lógica no se pueden derivar principios de razonamiento:

1. La lógica no estudia estados psicológicos u otras cuestiones normativas. Lo que es propio de las reglas de razonamiento es el tránsito entre ciertos estados psicológicos, por ejemplo, entre creencias. Sin embargo, la lógica no nos dice nada acerca de creencias (1986, p. 82).
2. La lógica es una disciplina descriptiva en donde se formulan ciertos hechos —presumiblemente necesarios— para estudiar sus propiedades y relaciones sintácticas y semánticas. En contraste, las reglas de razonamiento son proposiciones normativas, que permiten la transición entre —para usar la terminología de Goldman— estados doxásticos (1986, p. 82).

Con lo anterior es prudente preguntarse cuál es la legitimidad de los principios de razonamiento, si ellos no son inferidos lógicamente o probabilísticamente de las teorías formales [4].

Se puede agregar aún un tercer problema con la derivación de los principios normativos a través de las teorías formales. Supóngase que sin dificultad se ha seleccionado canónicamente la teoría formal que da origen a los principios normativos adecuados y que

éstos en algún sentido son derivados de la teoría formal elegida, aun si esto fuera posible no parece claro, ha argumentado Goldman (1986), que los principios así derivados sean los correctos. Se ha señalado que los principios normativos correctos no pueden ser derivados de la lógica formal, debido a que algunas veces las creencias actuales de los sujetos implican de hecho otras proposiciones, pero ello no significa que los sujetos deban creer tales proposiciones implicadas. Goldman presenta esta postura de esta manera:

> Supóngase que *p* es implicado por *q*, y *S* cree que *q*. ¿Se sigue que *S* debe creer que *p*, o incluso que S podría creer *p*? De ningún modo... ¡quizá lo que *S* debería hacer, después de notar que *q* implica *p*, es abandonar su creencia en *q*! Después de todo, algunas veces aprendemos cosas que nos hacen recomendable abandonar creencias previas (Goldman, 1986, p. 83, la traducción es mía) [5].

Para Goldman, no sólo hay dificultades para entender en qué sentido los principios normativos son derivados de las teorías formales, sino que aun si éstos fueron derivados en algún sentido, las reglas así derivadas podrían ser incorrectas (Goldman, 1986, p 81).

Los defensores de la visión estándar de la racionalidad podrían responder al tipo de problemas que apunta Goldman sugiriendo que los principios normativos son derivados de las reglas formales a través de "reglas de conversión" (Samuels, Stich y Faucher, 2004). Las reglas de conversión tendrían como propósito transformar o redefinir a los axiomas y teoremas de los sistemas formales en principios de razonamiento. Por ejemplo, se podría adoptar el siguiente esquema de conversión:

> Antecede a cada oración del lenguaje formal con la expresión "S cree que" y convierte a todas las instancias de "no es posible" por "S no tiene permitido".

Dada esta regla de conversión, por ejemplo, se podría rescribir la regla de la conjunción, es decir, de la regla:

> "no es posible que P(A) sea menor que P(A&B)",

se podría pasar al principio normativo:

> "S no tiene permitido creer que P(A) es menor a P(A&B)".

Una propuesta como la anterior señalaría el sentido en el cual los principios normativos son derivados de las teorías formales, a saber, un principio normativo de razonamiento es aquel que surge de aplicar un conjunto de reglas de conversión a una proposición dentro de un sistema formal. Si se sostiene que los principios normativos son el producto de la conjunción entre un sistema formal determinado y un conjunto de reglas de conversión, los argumentos de Goldman serían insuficientes para garantizar el rechazo de la idea de que los principios normativos son derivados de las teorías formales, debido a que habría algunas reglas de conversión que no conducen a las consecuencias que Goldman considera problemáticas (Samuels, Stich y Faucher, 2004). Supóngase, por ejemplo, que se adopta una regla de conversión que permita rescribir el *Modus ponens* como el siguiente principio de inferencia:

> Si *S* cree *p* y *S* cree que (Si *p*, entonces *q*), entonces *S* no debe creer que no *q*.

Tal regla de conversión no comprometería al sujeto a creer en la consecuencia lógica de las creencias "*p*" y "si *p*, entonces *q*", sino que sólo le exigiría evitar creer en la negación de lo que ellas implican. Así, se podría evadir la objeción de Goldman.

Si bien las reglas o esquemas de conversión permiten resolver las objeciones descritas anteriormente, los defensores de la visión estándar de la racionalidad se enfrentan con una serie de problemas. Uno de estos problemas radica en ofrecer una descripción correcta de las reglas de conversión que ofrezcan la interpretación de los principios formales a los principios normativos. Se ha argumentado que hasta que se ofrezca tal descripción, la teoría normativa del razonamiento que se desea defender está profundamente subespecificada (Samuels, Stich y Faucher, 2004, p 39). Del mismo modo, es posible plantear una serie indefinida de reglas que se pueden proponer para rescribir las proposiciones del lenguaje formal a reglas normativas. De modo que los defensores de la visión estándar de la racionalidad tendrían que escoger entre *a*) proponer una manera principal de seleccionar los esquemas de conversión, o *b*) enfrentar el problema de encontrarse con un número variado de visiones de racionalidad, cada una con su conjunto consistente de teorías formales unidas a un conjunto de

reglas específicas de conversión. Mientras que la alternativa *a*) lleva al problema de ofrecer una justificación de un esquema de conversión sobre los otros, lo que a su vez plantearía el problema de elaborar un criterio para que dicha justificación sea posible; la alternativa *b*) parece simplemente como opuesto a la visión estándar de la racionalidad, en tanto habría una multiplicidad de estándares normativos, todos provenientes de alguna teoría formal junto con diversas reglas de conversión.

A mi juicio, los problemas a los que conducen la diversidad de teorías formales y su modo de interpretarlas, la dificultad de derivar los principios de razonamiento de los principios lógicos y la propuesta de reglas de conversión para los principios de las teorías formales, sugieren que la visión estándar de la racionalidad tiene que ser rechazada o por lo menos que debe ser explorada otra teoría normativa. En esta sección, he intentado mostrar cómo la aceptación de la visión estándar es problemática, en tanto que dicha visión de la racionalidad no específica el modo en que las reglas o principios de la lógica y la matemática dan lugar a principios normativos que determinen cómo razonar. Por tal motivo, en el siguiente apartado exploraré una teoría normativa distinta a la visión estándar de la racionalidad.

IV.3. EL RECHAZO DE LA VISIÓN ESTÁNDAR DE LA RACIONALIDAD

En esta sección reviso dos argumentos a favor del rechazo de la visión estándar de la racionalidad. Con estos argumentos se intenta eliminar la visión estándar de la racionalidad y, con ello, la interpretación pesimista, dado que dicha interpretación supone la aceptación de la visión estándar de la racionalidad. El primer argumento desafía la concepción deontológica de la racionalidad implícita en la visión estándar de racionalidad. El segundo argumento que revisaré subraya el hecho de que la visión estándar de la racionalidad no toma en consideración los limitados recursos con los que cuenta el ser humano.

IV.3.1. EL ACERCAMIENTO DEONTOLÓGICO DE LA RACIONALIDAD

De acuerdo con la visión estándar de la racionalidad, lo que significa ser racional es razonar conforme a los principios norma-

tivos derivados de las teorías formales; de este modo, si un sujeto falla en razonar de esta manera, entonces dicho sujeto es irracional. Piatelli-Palmarini apunta lo anterior en el siguiente pasaje:

> Los principios universales de la lógica, aritmética y el cálculo de probabilidad... nos dicen lo que *debemos*... pensar, no lo que de hecho pensamos... si nuestra intuición nos conduce de hecho a resultados incompatibles con la lógica, concluimos que nuestra intuición está equivocada (Piatelli-Palmarini, 1994, p. 158, la traducción es mía).

Si bien ya he expuesto algunas dificultades de la visión estándar de la racionalidad, en el pasaje citado se encuentra implícito lo que se ha denominado el "acercamiento deontológico de la racionalidad". De acuerdo con este acercamiento, lo que significa razonar correctamente —lo que es constitutivo del buen razonamiento— es *solamente* razonar en concordancia con algún conjunto de reglas o principios. Dentro del acercamiento deontológico un razonamiento es el adecuado por el simple hecho de *seguir* el conjunto de reglas que determina lo que significa razonar correctamente [6].

El acercamiento deontológico, sin embargo, no es la única concepción de racionalidad que se ha defendido. Un acercamiento a la racionalidad importante es el denominado "consecuencialismo", el cual sostiene que razonar correctamente es hacerlo de manera tal que probablemente se obtengan o alcancen ciertas metas o resultados. En el consecuencialismo lo aceptable se define a partir de las consecuencias. La postura consecuencialista no está en contra de las reglas, esto es, no se afirma que las reglas no juegan un papel importante en las teorías normativas del razonamiento, sino que mantiene que razonar de acuerdo con algún conjunto de reglas no es en sí mismo constitutivo de un buen razonamiento (Foley, 1993). En el consecuencialismo la racionalidad de un proceso se evalúa por sus consecuencias, independientemente de que se haga uso o no de reglas. En otras palabras, aunque la aplicación de las reglas de razonamiento podría ser un medio para alcanzar ciertos fines, lo que es constitutivo de un proceso de razonamiento racional es —desde el acercamiento consecuencialista— ser un medio efectivo para alcanzar una meta o un conjunto de metas. En este sentido, la visión estándar de la racionalidad

no tiene cabida dentro del acercamiento consecuencialista de la racionalidad, pues afirma que el buen razonamiento consiste en satisfacer los principios que se derivan de la lógica y la matemática. Un proceso *sin más* es un proceso racional, desde la visión estándar, si sigue los principios normativos fundados en las teorías formales y no por las consecuencias que se obtengan de seguir tales principios.

Las divergencias entre los acercamientos normativos consecuencialista y deontológico quizá puedan verse con mayor claridad en la ética. Para la postura deontológica, lo que hace una elección moralmente correcta es su conformidad con una norma moral. Tales normas deben ser simplemente obedecidas por un agente moral (Larry y Moore, 2008). En contraste con las posturas consecuencialistas, las teorías deontológicas juzgan que la moralidad de los actos por criterios diferentes a los estados de cosas que esos actos morales pueden traer como consecuencia (Larry y Moore, 2008). Por su parte, el consecuencialismo en la ética sostiene que la corrección moral de un acto depende únicamente de las consecuencias que tenga dicho acto (Sinnott-Armstrong, 2006). El paradigma del consecuencialismo en la ética es el utilitarismo, cuyos proponentes clásicos son Bentham y Stuart Mill.

Dejando los aspectos éticos de lado, con respecto a la racionalidad distinguiré tres tipos de consecuencialismo. Un primer modo de entender al consecuencialismo es aquel que denominaré *veritista.* Este tipo de consecuencialismo sostiene que la meta o la consecuencia con la cual se debe evaluar el razonamiento es qué tanto un proceso cognitivo acerca al sujeto a la verdad y/o lo aleja de la falsedad. Un ejemplo de este tipo de consecuencialismo es el confiabilismo, expuesto en el capítulo I, donde un proceso de razonamiento adecuado es aquel que tiende a producir mayor número de creencias verdaderas que falsas (Goldman, 1986). Otra forma de consecuencialismo es la *pragmatista*, que sostiene que un proceso de razonamiento es bueno si es adecuado para alcanzar los objetivos pragmáticos de satisfacer las metas y deseos personales. Un ejemplo de consecuencialismo pragmatista es la propuesta de Stich.

> Al evaluar sistemas de procesos cognitivos, el sistema que debe ser preferido es aquel en el que sería más probable alcanzar aquellas

> cosas que son intrínsecamente valoradas por la persona cuyos intereses son relevantes para el propósito de la evaluación. En la mayoría de los casos, la persona relevante será aquella que está o que podría estar usando el sistema. Así, por ejemplo, si el asunto en cuestión es la evaluación del sistema de procesos cognitivos de Smith en comparación con alguna alternativa factual o hipotética, el sistema que, bajo la explicación pragmatista, tiene una evaluación cognitiva alta es aquel que más probablemente conduzca a las cosas que Smith encuentra intrínsecamente valiosas (Stich, 1990, pp. 131-32, la traducción es mía).

Una tercera manera de entender al consecuencialismo es de un modo *biológico*. Este tipo de consecuencialismo evalúa el razonamiento a partir de qué tanto nos conduce a metas como la supervivencia, la reproducción o la propagación de los genes. Como expondré en el próximo capítulo, Stanovich y West (2000 y 2002) sostienen un criterio consecuencialista de corte biológico, que ellos denominan "racionalidad evolucionista", para evaluar parte del razonamiento humano.

Una vez hecha la distinción entre el acercamiento consecuencialista y el deontológico es posible regresar a la objeción en contra de la visión estándar de la racionalidad. La objeción en contra de esta visión, tomada de Samuels, Stich y Faucher (2004), consiste en mostrar que dado que dicha visión se encuentra enmarcada en un acercamiento deontológico a la racionalidad, entonces si es rechazado el acercamiento deontológico a favor de uno consecuencialista, se tendría que rechazar también a la visión estándar de la racionalidad. La crítica al acercamiento deontológico se ha llevado a cabo por diferentes frentes; uno de los más socorridos es el relacionado con el *valor* de un buen razonamiento (Samuels, Stich y Faucher, 2004). Según algunos filósofos, el problema acerca de la racionalidad y de la calidad del razonamiento humano es importante, sólo si razonar correcta o incorrectamente realmente importa. Esto ha llevado a algunos filósofos a sostener que cualquier teoría normativa debe ofrecer una "condición de valor" que explique por qué es importante razonar de un modo normativamente correcto, es decir, que explique por qué el buen razonamiento es deseable (Samuels, Stich y Faucher, 2004). Si se acepta la condición de valor, los acercamientos consecuencialistas están por encima de los deontológicos en tanto que con más sencillez

pueden cumplir el *desideratum* expuesto por tal condición. Al construir una teoría consecuencialista del razonamiento se procede a identificar las metas o fines del buen razonamiento —los "bienes cognitivos" (Kitcher, 1992). Se debe, según Kitcher,

> reflexionar sobre la empresa cognitiva (incluyendo la empresa científica), sobre su historia y sobre las capacidades de aquellos que participan en ella, para alcanzar formulaciones *corregibles* de las metas de la empresa y explicaciones *corregibles* de las estrategias prometedoras para alcanzar tales metas (1992, p. 52, la traducción es mía).

Así, si perseguir el logro de metas personales o buscar la adquisición de creencias verdaderas es valioso, entonces el acercamiento consecuencialista (en este caso el pragmatismo o el confiabilismo, respectivamente) puede identificarlas como metas que se desean obtener. Para completar el proyecto consecuencialista, además de especificar los fines apropiados, es necesario especificar los métodos o procesos que permitan eficientemente obtener tales fines. Así, el acercamiento consecuencialista provee una clara explicación de por qué es importante el buen razonamiento, a saber, el buen razonamiento es aquel que conduce a la posesión de aquello que se valora.

En contraste con el acercamiento consecuencialista, no parece claro cómo el acercamiento deontológico puede dar cuenta de la condición de valor. Podría preguntársele a un defensor del acercamiento deontológico por qué se debe razonar en concordancia con un conjunto preespecífico de principios normativos. Sin embargo, cualquier intento desde el acercamiento deontológico de subrayar las consecuencias del razonamiento en un modo racional mostraría que el consecuencialismo es una mejor explicación de la racionalidad. Debido a que desde el acercamiento deontológico se afirma que razonar en concordancia con el conjunto de reglas *R* (por ejemplo, aquellas provenientes de la lógica y la matemática) es constitutivo del buen razonamiento, dicho acercamiento se compromete con la afirmación de que un sujeto *S* que razona de acuerdo con *R* está razonando correctamente aún si hay maneras más eficientes de lograr sus fines deseables. De hecho, desde el acercamiento deontológico el propósito de *R* no es la obtención de metas.

Una posición deontológica, según Samuels, Stich y Faucher (2004), trae consigo una serie de problemas. Según estos filósofos, el primero de ellos es que *prima facie* resulta más deseable lograr los fines y metas que le interesan a un sujeto que estar en mera concordancia con un conjunto de estándares normativos, por ejemplo, aquellos que provienen de la lógica. Sería difícil entender cómo, bajo esta perspectiva, se puede explicar, desde el acercamiento deontológico, que ser racional resulta mucho más valioso que no ser racional. El segundo problema es que sería contraintuitivo afirmar que la racionalidad puede obligar a los sujetos a evitar los medios eficientes para lograr los fines deseados. En este sentido, a diferencia del acercamiento deontológico, el acercamiento consecuencialista de la racionalidad parece capturar la intuición correcta de que no se debe estar obligado a estar en consonancia con principios de razonamiento en contextos donde tales principios no son los medios eficientes para alcanzar las metas y fines deseados.

IV.3.2. LA LIMITACIÓN DE RECURSOS DE LOS SUJETOS COGNOSCENTES

En los estudios sobre racionalidad de los últimos años ha habido una mayor preocupación por comprender los criterios de razonamiento como dependientes de recursos. Los sujetos cognoscentes están inmersos en una gran variedad de limitaciones de recursos como limitaciones de tiempo, energía, poder computacional, memoria, atención e información (Goldman, 1993). Ha sido a partir del trabajo de Herbert Simon (1957) desarrollado en los años cincuenta, que cada vez es más común que los teóricos dedicados al estudio del razonamiento fijen su atención en las limitaciones de recursos cuando deciden cuáles son los estándares de razonamiento que se deben aceptar. Lo que se busca es mostrar que las teorías normativas dependen de sistemas cognitivos particulares con limitaciones de recursos específicos; en otras palabras, se busca adoptar concepciones de racionalidad acotadas o relativas a recursos y no independientes de éstos (Goldman, 1993; Gigerenzer y Selten, 2002).

Si bien se han señalado distintas críticas a la concepción de racionalidad independientes de recursos, quizá las más comúnmente invocadas se fundan en la defensa del principio en el cual los sujetos cognoscentes deben *hacer* lo que *pueden* o lo que *está a*

su alcance hacer [7]. A grandes rasgos, la idea es que así como en cuestiones éticas las obligaciones están restringidas por lo que se pude hacer, así también en cuestiones epistémicas no se está obligado a satisfacer los estándares normativos que estén más allá de las capacidades humanas (Kitcher, 1992). La adopción del principio "si no se puede hacer A, entonces no es el caso que se deba hacer A" requiere el rechazo de una concepción de estándares normativos independientes de recursos a favor de una concepción dependiente de éstos.

El principio "el deber implica el poder" se apoya en el hecho de que no todos los sujetos cognoscentes son capaces de ejecutar las mismas tareas de razonamiento. Los seres humanos no tienen las capacidades que una computadora y, al mismo tiempo, otros seres —por ejemplo, los primates no humanos— no pueden realizar algunas de las tareas de razonamiento que el ser humano puede realizar. Si es el caso que "el deber" está implicado por "el poder", entonces, dado un conjunto de estándares normativos, habría ciertos sujetos cognoscentes que estarían obligados a seguirlos, mientras que otros no lo estarían; la obligación de los primeros descansa en que estos sujetos podrían satisfacer tales estándares, mientras que los segundos no tendrían la capacidad para satisfacerlos y, por lo tanto, no estarían obligados a seguirlos. Así, la adopción de un principio como "el deber implica el poder" sugiere el rechazo de estándares independientes de recursos a favor de estándares dependientes de éstos [8]. Stich apunta de manera humorística la relación entre la dependencia de recursos y los estándares normativos: "es simplemente perverso juzgar que los sujetos están haciendo un mal trabajo en el ámbito del razonamiento, porque no están usando una estrategia que requeriría un cerebro del tamaño de un dirigible" (1990, p. 27, la traducción es mía). Desde una concepción de estándares normativos dependientes de recursos, la preocupación está centrada en cómo deben razonar los sujetos cognoscentes dados los recursos que poseen, y una infinidad de tiempo no es un recurso que poseen esos sujetos (Gigerenzer y Selten, 2002).

Una crítica a la visión estándar de la racionalidad desde la dependencia de recursos es la idea de que los sujetos deben conservar la consistencia de cada una de sus creencias para ser considerados racionales. Cherniak (1986) ha sostenido que aun

concibiendo de modo conservador un número estimado de creencias para un sujeto *S*, sería una tarea computacionalmente inalcanzable determinar si sus creencias son consistentes [9]. De ese modo, si no se pueden satisfacer las normas de la visión estándar de la racionalidad, dado que "el deber implica el poder", se sigue que dicha visión no es un criterio adecuado de racionalidad. Cherniak (1986) considera que las nociones de racionalidad que se han aceptado explícita o tácitamente en la filosofía son demasiado idealizadas como para ser aplicadas a los seres humanos. Por tal motivo, se han explorado nociones de racionalidad que toman en cuenta los recursos cognitivos tales como el tiempo y la memoria.

Con el propósito de hacer más fuerte al argumento de la limitación de recursos en contra de la visión estándar, se puede defender una racionalidad dependiente de recursos comprometida con el acercamiento consecuencialista de la normatividad. Esta postura tendría importantes implicaciones en la manera en que comúnmente se conciben los estándares normativos. Una de estas implicaciones es que se tiene que negar que los principios normativos de razonamiento son universales en dos sentidos. El primero de estos sentidos es que el vínculo entre dependencia de recursos y el acercamiento consecuencialistas niega que el mismo conjunto de reglas de razonamiento deba ser empleado por *todos los sujetos cognoscentes*. Por el contrario, las reglas de razonamiento serán normativamente correctas con respecto a un tipo de sujeto cognoscente. De acuerdo con el consecuencialismo, un buen razonamiento consiste en desplegar los procesos cognitivos eficientes para alcanzar ciertas metas —por ejemplo, creencias verdaderas, desde una postura confiabilista, o la satisfacción de metas y deseos personales, desde el pragmatismo individualista. La adopción de un consecuencialismo dependiente de recursos no supone en sí mismo que las metas del buen razonamiento sean relativas a diferentes clases de sujetos cognoscentes, ya que un consecuencialista veritista podría sostener, por ejemplo, que la adquisición de creencias verdaderas y el rechazo de creencias falsas es siempre una meta del buen razonamiento. Lo que sí supone el consecuencialismo dependiente de recursos es que el conjunto de reglas o procesos para adquirir ciertas metas podrían ser apropiadas para un tipo de individuos pero no para otros. Esta implica-

ción contrasta con la visión estándar de la racionalidad, la cual mantiene que la misma clase de reglas es normativamente correcta, independientemente de los recursos cognitivos disponibles a los sujetos. De este modo, la dependencia de recursos socava una idea importante de la visión estándar de la racionalidad, esto es, la idea de que los criterios normativos son principios universales.

La adopción de un consecuencialismo dependiente de recursos también supone la relativización de las evaluaciones a rangos de *medios ambientes* específicos. Por ejemplo, si se adopta una forma de confiabilismo dependiente de recursos, entonces se tendría que especificar el tipo de ambiente en el cual la evaluación es hecha para determinar si un proceso de razonamiento es un proceso normativamente adecuado. Lo anterior se justifica por varias razones, entre ellas la idea de que diferentes ambientes pueden afectar la eficiencia de un proceso de razonamiento. En primer lugar, los diferentes ambientes proporcionan diferentes tipos de información a los sujetos. Por ejemplo, algunos ambientes sólo contenían información probabilística que es presentada en forma de frecuencias (como consideran los psicólogos evolucionistas que fue el AAE), mientras que en otros ambientes la información probabilística podía encontrarse en formatos no frecuentistas. Es interesante tener en mente que parece ser una posibilidad empírica —según los estudios de los psicólogos evolucionistas que fueron revisados en el capítulo anterior— que la diferencia de la información que se encuentra en el ambiente puede afectar la eficacia de un proceso de razonamiento. Del mismo modo, diferentes ambientes pueden imponer diferentes tipos de restricciones de tiempo. En algunos ambientes puede haber demasiado tiempo para que el sujeto cognoscente ejecute una tarea de razonamiento dada, mientras en otros ambientes el tiempo puede ser insuficiente. Es más que verosímil sostener que tales restricciones de tiempo pueden afectar la eficacia de los distintos procesos de razonamiento para alcanzar metas cognitivas como la adquisición de creencias verdaderas. La adopción de una forma de consecuencialismo dependiente de recursos conduce al rechazo de que los estándares de buen razonamiento son aplicados en todos los ambientes, esto es, que los estándares son independientes de cualquier contexto.

En suma, en este apartado he mostrado que la visión estándar de la racionalidad ignora el papel del medio ambiente y los recursos cognitivos de los seres humanos. El principio "el deber implica el poder" sugiere que las reglas de razonamiento adecuadas no pueden exigir a los sujetos razonar de maneras que vayan más allá de sus capacidades y recursos. Una racionalidad consecuencialista y dependiente de recursos, además, supone que las normas de razonamiento pueden ser relativas a los sujetos particulares y a ambientes específicos, esto es, niega la posibilidad de que tales normas sean universales.

IV.4. LA DEPENDENCIA DE RECURSOS Y LA INTERPRETACIÓN PESIMISTA DE LA RACIONALIDAD HUMANA

¿Cuáles serían las implicaciones de aceptar un acercamiento consecuencialista con relación a la interpretación pesimista de la racionalidad humana y, en general, con el estudio que tiene como propósito evaluar el razonamiento dado un determinado estándar de la racionalidad? ¿Supone la concepción dependiente de recursos que la interpretación pesimista de la racionalidad humana es inadecuada? Diversos filósofos (Stein, 1996, Samuels, Stich y Faucher, 2004) consideran que es perfectamente compatible la interpretación pesimista de la racionalidad humana con la idea de que los estándares normativos son dependientes de recursos. Estos filósofos sostienen que de la adopción de una noción de racionalidad dependiente de recursos no se sigue que la evidencia empírica proveniente de la tradición de heurística y sesgo no apoye la interpretación pesimista de la racionalidad humana. Según Stein (1996), la adopción de una racionalidad dependiente de recursos no implica que los seres humanos nunca deban satisfacer algún principio estipulado por la visión estándar de la racionalidad. Por poner un ejemplo, no parece claro que evitar la falacia de la conjunción en el caso de "Linda, la cajera feminista" sea algo que los sujetos cognoscentes no puedan ejecutar. Una idea similar a la de estos filósofos es defendida por Stanovich y West (2000) quienes afirman que no todos los errores cometidos en las tareas de razonamiento pueden ser explicados por las limitaciones computacionales de los sujetos.

> Hay algunas tareas en la literatura de heurística y sesgo que carecen de cualquier tipo de asociación con la habilidad cognitiva. Colectivamente, estos resultados indican que las limitaciones computacionales parecen lejos de ser absolutas (Stanovich y West, 2000, p. 649, la traducción es mía).

Lo que estos psicólogos tienen en mente es que para que una racionalidad dependiente de recursos socave la idea de que los experimentos apoyan alguna variante de la interpretación pesimista de la racionalidad humana, se debe mostrar que dichos experimentos presentan tareas que están muy por encima de las capacidades cognitivas de los seres humanos. Sin embargo, para Stanovich y West (2000), lo anterior nunca ha sido mostrado.

A mi juicio, si bien un acercamiento consecuencialista de la racionalidad, que además sea dependiente de recursos, puede ser lógicamente compatible con alguna versión de la interpretación pesimista de la racionalidad humana (como lo sugieren Stanovich y West, 2000 y Stein, 1996), de ahí no se sigue que la evaluación del razonamiento humano no se modifique al aceptar otros criterios que determinen qué significa razonar correctamente. Lo anterior tiene sentido, en tanto que el estudio evaluativo de la razón depende de una descripción del razonamiento humano y de un criterio que establezca cómo se debe razonar. De hecho, el principio "el deber implica el poder" sugiere que los estándares con los que se evalúa el razonamiento están determinados, en parte, por la manera en que los seres humanos razonan [10]. Regresando nuevamente a la distinción que hice en el capítulo II, el principio "el deber implica el poder" sugiere que la tarea normativa de la racionalidad está influenciada fuertemente por la tarea descriptiva del razonamiento, esto es, que los estándares normativos que nos dicen cómo se debe razonar dependen, en parte, de cómo de hecho los seres humanos razonan.

Así, considero que lo que sí se sigue de la adopción de un acercamiento consecuencialista dependiente de recursos es que se vuelve una tarea extremadamente compleja determinar en qué medida los procesos de razonamiento usados por los sujetos violan los estándares normativos. Si se ha adoptado una postura normativa dependiente de recursos, resulta inadecuado invocar a los principios formales de la lógica y la estadística como los

únicos parámetros del buen razonamiento. Por el contrario, al adoptar una postura dependiente de recursos es necesario establecer un amplio entramado conceptual y empírico para determinar cuáles son los estándares bajo los cuáles se puede calificar el razonamiento de los sujetos cognoscentes [11]. Eso supone que se deben especificar varios parámetros relativos a conjuntos de sujetos cognoscentes y a rangos ambientales, antes de que un estándar pueda ser aplicado. Dado que es una cuestión empírica indicar cuáles son los métodos que resultan ser los medios eficientes para alcanzar los fines normativos de los sujetos cognoscentes en un determinado ambiente, no es del todo claro indicar cuáles son las reglas o procesos que los sujetos deben emplear en su razonamiento. Gigerenzer y su grupo (Gigerenzer y Selten, 2002) han investigado la eficacia de ciertas estrategias de razonamiento conocidas como *heurísticas rápidas y frugales*. Como su nombre lo indica, estos procesos de razonamiento son computacionalmente económicos y veloces, de ahí que a diferencia de los métodos tradicionales asociados con la visión estándar son fácilmente utilizados por los seres humanos. Lo importante aquí es que a pesar de su frugalidad, estos algoritmos son extremadamente eficaces al ejecutar tareas de razonamientos dentro de un cierto rango de ambientes. Si se adopta una forma de consecuencialismo dependiente de recursos, es posible que se ofrezca una posibilidad empírica genuina de que las heurísticas rápidas y frugales sirvan de modelos para conformar principios normativos adecuados, esto es, aquellos con los cuales se pueda juzgar la ejecución de razonamiento.

Otra propuesta en psicología cognitiva para dar cabida a los experimentos provenientes de la tradición de heurística y sesgo, así como los distintos acercamientos a la racionalidad aquí expuestos es la teoría dual de sistemas. Los defensores de esta propuesta muchas veces han sostenido que esta teoría tiene la virtud de unir las teorías descriptivas del razonamiento con las teorías normativas de la racionalidad (Stanovich y West, 2000). La teoría dual de sistemas, como examinaré en el próximo capítulo, recoge algunos principios de la tradición de heurística y sesgo, así como de la psicología evolucionista. Asimismo, esta teoría asume una distinción entre dos distintas maneras de entender los criterios normativos para evaluar el razonamiento. En otras palabras,

la teoría dual, en su nivel normativo, sostiene que no existe un único criterio normativo para evaluar al razonamiento humano. Como sea, antes de poder conocer cuáles son las tesis centrales de la teoría dual, es preciso recapitular lo visto hasta aquí, así como valorar cuáles son las relaciones que tiene un acercamiento consecuencialista con las nociones de justificación.

IV.5. CONCLUSIONES

A lo largo de este capítulo he expuesto tres grandes estrategias en contra de la interpretación pesimista de la racionalidad. La primera consiste en señalar que los experimentos provenientes de la tradición de heurística y sesgo no toman en cuenta los elementos pragmáticos. Esta estrategia, de ser correcta, explica cómo los sujetos pueden interpretar las pruebas de razonamiento presentadas por los psicólogos. De ser este el caso, los sujetos no ejecutan incorrectamente las tareas de razonamiento, sino que simplemente interpretan dichas tareas de distintas maneras. Un problema con esta estrategia radica en que es difícil pensar que cada uno de los experimentos de razonamiento pueda ser explicado a través de principios pragmáticos. El problema central es que la estrategia pragmatista parece chocar con la evidencia proporcionada por los psicólogos evolucionistas, en donde una formulación distinta de las tareas de razonamiento provee resultados sustantivamente mejores que en las formulaciones de los psicólogos defensores de la tradición de heurística y sesgo.

La segunda estrategia en contra de la interpretación pesimista revisada en este capítulo discute la existencia de diversas interpretaciones de las normas provenientes de la visión estándar de la racionalidad, así como los problemas mismos que surgen para derivar las normas de razonamiento de los principios provenientes de la lógica y la matemática. Estas dos posturas (o subestrategias) no ponen en entredicho la visión estándar de la racionalidad, sino que aceptándola señalan los problemas para evaluar el razonamiento humano por parte de dicha visión. En el fondo, estas posturas muestran que la aceptación de la visión estándar como teoría normativa de la racionalidad por sí misma no es suficiente para poder explicar las normas correctas de razonamiento, y por lo tanto no pueden ser usadas para evaluar o determinar si el ser

humano satisface o no los principios normativos adecuados. En esa sección del capítulo señalé algunos de los problemas con los que se enfrenta esta segunda estrategia, los cuales son el punto de partida para explorar una nueva teoría de la racionalidad.

La última estrategia que reviso en este capítulo es el rechazo mismo de la visión estándar de la racionalidad. En esta sección muestro cómo es que la visión estándar de la racionalidad se encuentra enmarcada en un acercamiento deontológico. Este acercamiento puede ser contrastado con el acercamiento consecuencialista, dentro del que he distinguido una versión veritista, pragmatista y biológica. Si se acepta que las teorías normativas deben explicar por qué es importante razonar de un modo normativamente correcto, entonces parece que los acercamientos consecuencialistas de la racionalidad están por encima de los deontológicos. A este acercamiento se le agregó el reconocimiento que se debe tener acerca de los recursos tanto cognitivos como medioambientales, por lo que bajo esta perspectiva las teorías normativas deben estar relativizadas a sistemas cognitivos particulares con limitaciones de recursos específicos.

Al evaluar esta última estrategia es posible percatarse que al aceptar un acercamiento consecuencialista dependiente de recursos, la cuestión de la evaluación del razonamiento se hace más compleja, en tanto que no parece haber un único estándar que pueda ser usado por todos los sujetos en cualquiera de los ambientes en donde se encuentren. De este modo, la aceptación de un acercamiento consecuencialista y dependiente de recursos requiere de una teoría que recoja la evidencia empírica necesaria, así como un entramado conceptual en donde se integren diferentes criterios para evaluar el razonamiento humano. Como señalé anteriormente, dos posibles teorías que realizan esta tarea son la propuesta de la racionalidad ecológica de Gigerenzer y su grupo de investigación (Gigerenzer y Selten, 2002), por un lado, y los defensores de la teoría dual de sistemas (Evans y Over, 1996, Stanovich y West, 2000, 2003), por el otro. En este trabajo revisaré sólo la propuesta de la teoría dual, en parte, porque los defensores de esta teoría pretenden combinar algunos de los principios de la tradición de heurística y sesgo, y la psicología evolucionista, así como crear un "puente" entre las teorías del razonamiento y racionalidad (Stanovich y West, 2000). Este "puente" teórico entre

las teorías del razonamiento y la racionalidad brindará una conexión importante entre las nociones de justificación y de racionalidad que guían este libro.

Si bien en este capítulo no se ha expuesto una nueva teoría de razonamiento, se ha introducido un nuevo acercamiento a la racionalidad, i. e., el acercamiento consecuencialista. Además, se ha mostrado cuál es el acercamiento normativo que supone la visión estándar de la racionalidad, esto es, el acercamiento deontológico. En el capítulo anterior argumenté que tanto la tradición de heurística y sesgo como la psicología evolucionista asumen la visión estándar de la racionalidad, y si lo visto en este capítulo es correcto, entonces las dos teorías del razonamiento hasta aquí expuestas presuponen de manera implícita un acercamiento deontológico de la racionalidad. Así, tenemos que los defensores de la tradición de heurística y sesgo y los psicólogos evolucionistas comparten las mismas teorías normativas de la racionalidad, a saber, la visión estándar de la racionalidad y, en un sentido más profundo, el acercamiento deontológico a la racionalidad.

En las conclusiones de los capítulos II y III he analizado cuáles serían las relaciones entre las nociones internista y externista, y las dos teorías psicológicas estudiadas (la tradición de heurística y sesgo y la psicología evolucionista). En este capítulo que he presentado un acercamiento distinto a la racionalidad deseo hacer algunas observaciones acerca de cómo cada una de las nociones de justificación se relaciona con los dos acercamientos a la racionalidad. En términos generales, cuando se dice que una creencia está justificada, por ejemplo, en la noción tradicional de conocimiento, se considera que el sujeto cree racionalmente dicha creencia, pero, ¿con qué acercamiento de la racionalidad se está comprometiendo exactamente la noción de justificación? Además, dado que he distinguido varias nociones de justificación, podría preguntarse si dichas nociones están asumiendo un acercamiento particular de la racionalidad. Para poder responder a esta pregunta analizaré las relaciones entre las nociones internistas y externistas de la justificación y los dos acercamientos a la racionalidad expuestos en este capítulo.

En reiteradas ocasiones he insistido que hay un estrecha relación entre las nociones de racionalidad y de justificación. Si se acepta que la noción de racionalidad debe entenderse únicamen-

te como la visión estándar de la racionalidad, se tendría que aceptar que los criterios que indican que una creencia está justificada o que es racional serían aquellos que satisfagan los principios provenientes de la lógica y la matemática. ¿Esta idea de racionalidad y justificación apoya a una noción particular de justificación? El internismo, por una parte, sostiene que una creencia está justificada, si el sujeto tiene acceso a los rasgos o procesos que hacen que una creencia esté justificada. Si los criterios provenientes de la lógica, la matemática y la probabilidad son algo a lo que el sujeto puede tener acceso, en principio parecería compatible la visión estándar de la racionalidad con el internismo. Por ejemplo, el papel de la coherencia en una teoría internista de la justificación como la de Bonjour (1985) se encuentra íntimamente ligada con la lógica. La relación coherente entre las creencias de un sistema es algo a lo que, según este filósofo, puede tener acceso el sujeto.

La visión estándar de la racionalidad y la noción internista de la justificación parecen compartir la importancia en la argumentación y la relación lógica que tienen las creencias entre sí como un elemento importante para determinar si una creencia es racional o está justificada. Sin embargo, la visión estándar no se compromete con la idea de que un sujeto es racional en creer algo si el sujeto es consciente de que la creencia que sostiene satisface criterios provenientes de la lógica y la matemática. La ausencia del rasgo anterior parecería minar la conexión entre la visión estándar de la racionalidad y el internismo. No obstante, esta aparente disociación no es lo suficientemente fuerte para socavar la conexión entre estas dos posturas, ya que el internismo no sostiene que cada vez que un sujeto esté justificado, éste haya sido consciente de que las creencias cumplen cierto criterio para considerarlas justificadas, sino que en principio pueda tener acceso a dichos principios y que el sujeto esté justificado en seguir tales principios (Pappas, 2005). Es precisamente en este sentido como se puede entender la conexión entre el internismo y la visión estándar de la racionalidad, a saber, que dicha visión de la racionalidad puede servir como guía para decidir qué creer y permite aceptar sólo las creencias que el sujeto cognoscente puede creer. Nótese, además, que de esta manera se recoge la intuición preanalítica detrás de la noción internista de justificación, esto es, la responsabilidad epistémica y el "dar y ofrecer" razones.

El externismo, por su parte, sostiene que una creencia está justificada no por el acceso interno que el sujeto tiene a los rasgos que justifican una creencia, sino por el hecho de que la creencia esté producida por los procesos de formación de creencias adecuados. Esta noción de justificación supone que una creencia es racional si es producida por un mecanismo adecuado de producción de creencias. Sin embargo, la visión estándar de la racionalidad no hace referencia al modo en que las creencias fueron causadas o producidas sino a que las creencias cumplan con criterios fundados en una teoría formal. Se podría sostener que hay una conexión fuerte entre la visión estándar de la racionalidad y el externismo, si se afirma que los procesos de formación de creencias adecuados son procesos que satisfacen los criterios normativos derivados de la lógica y la matemática. Sin embargo, Goldman (1986) uno de los defensores más importantes del confiabilismo ha defendido que las reglas normativas que justifican a las creencias no pueden ser aquellas que se originen de la lógica o la probabilidad.

El argumento que ofrece Goldman (1986) para sostener que los criterios de justificación no pueden ser criterios que se deriven de la lógica y la matemática explica cuál es la relación entre la noción externista de la justificación y la visión estándar de la racionalidad. En el capítulo quinto de *Epistemology and Cognition*, Goldman presenta un contraejemplo para mostrar por qué la lógica no puede estar detrás de los criterios de justificación:

> Imaginemos a Martha una mujer que cree justificadamente que *p* e imaginemos también que Ø, un conjunto de proposiciones, puede ser lógicamente derivado de *p*. Imaginemos también que incluso para un experto en lógica le sería difícil entender cómo es que Ø se sigue de *p* y Martha no sabe de lógica, ni conoce a alguna autoridad que le advierte que de Ø se sigue *p*. No obstante, Martha cree que Ø [12].

Goldman se pregunta "¿está Martha justificada en creer que Ø? Su respuesta es un "no". El caso de Martha, según Goldman, muestra que la lógica no puede brindar un criterio que determine cuándo una creencia está justificada, pues la transición de *p* a Ø en el caso de Martha es lógicamente válida, sin embargo, no puede decirse que ella está justificada en creer que Ø. Goldman (1986,

pp. 90-91) presenta un ejemplo similar para la probabilidad, con el fin de indicar que tampoco la probabilidad puede fundar a los criterios adecuados de justificación. Goldman (1986) afirma que el problema con la idea de que los criterios de justificación pueden reducirse a aquellos que provienen de la lógica y la matemática es que permite cualquier tipo de transición proveniente de las teorías formales, sin poner restricciones en *cómo* es hecha dicha transición ni en los *procesos* para llegar a la creencia en cuestión.

Lo interesante de los contraejemplos de Goldman es que éstos chocan con la idea de que la visión estándar de la racionalidad explica el origen de los estándares de racionalidad y justificación. De modo que el confiabilismo de Goldman choca con la visión estándar de la racionalidad. Bajo esta perspectiva, una creencia está justificada o es racional independientemente de que satisfaga criterios fundados en la lógica y la matemática. Lo que sí hace la diferencia entre criterios adecuados y aquellos que no lo son, bajo una perspectiva confiabilista, es su conexión con la verdad. Un proceso de formación de creencias adecuado es aquel que conecta a las creencias con su posible verdad. Por ello, según Goldman (1986), el confiabilismo es un tipo de consecuencialismo, pues evalúa a los procesos de formación de creencias en tanto que conducen o no a la verdad. Así, el confiabilismo se compromete con un tipo particular de consecuencialismo, a saber, el consecuencialismo veritista.

El caso de Martha para mostrar cómo la noción de justificación no se reduce a principios que deriven su fuerza normativa de las teorías formales, muy bien pudo haber sido usada por los internistas para sostener que cuando un sujeto está justificado, éste debe ofrecer razones para decir que está justificado. Pero al sostener esta tesis se estarían comprometiendo con la idea, que también es defendida por Goldman, de que la visión estándar de la racionalidad no podría por sí misma ofrecer una explicación adecuada de las normas de justificación. Lo interesante de la relación entre el consecuencialismo y los criterios que determinan cuándo una creencia está justificada es que incluso, habría sentidos en los que los internistas defienden posturas consecuencialistas.

Bonjour (1985) sostiene que el metacriterio para determinar cuándo una teoría de la justificación es la adecuada depende de que las creencias justificadas por los criterios de la teoría a evaluar

sean probablemente verdaderas. Este filósofo muchas veces sostiene que el valor de la noción de justificación es instrumental. Para Bonjour, el valor de la noción de justificación se debe a la noción verdad. Un conjunto dado de estándares epistémicos es correcto o razonable sólo al saber que los estándares en cuestión son conductores genuinos a la meta cognitiva de la verdad (Bonjour, 1985, p. 9) [13]. El énfasis en la noción de coherencia de la postura internista de Bonjour hace creer que los criterios que determinan cuándo una creencia está justificada dependen de principios lógicos. Sin embargo, Bonjour insiste en que la coherencia es importante porque conduce a la verdad (1976, pp. 303 ss). Esto se relaciona con las críticas hechas al coherentismo de Bonjour, revisadas en el primer capítulo de este trabajo, según las cuales para que las creencias de los sujetos estén justificadas éstas tienen que ser coherentes con un sistema de creencias y además deben estar justificados *a priori* de que la coherencia conduce a la verdad (Goldman, 1989).

En suma, la noción internista de la justificación es en algunos aspectos compatible con la visión estándar de la racionalidad. Aún así, no puede sostenerse que la visión estándar de la racionalidad aceptada implícitamente por los defensores de la tradición de heurística y sesgo o los psicólogos evolucionistas apoye a una noción internista de la justificación. Goldman ofrece un argumento convincente para mostrar cómo la visión estándar de la racionalidad no puede ofrecer un criterio adecuado que determine qué rasgos debe tener una creencia justificada. La pauta para definir los criterios adecuados de racionalidad y justificación es, según este filósofo, la noción de verdad. Un criterio que determine cuándo una creencia está justificada será el correcto, si es un buen medio para alcanzar la verdad. Por lo anterior, la postura externista de Goldman se compromete explícitamente con el consecuencialismo. Sin embargo, el acercamiento consecuencialista de la racionalidad no tiene únicamente afinidad teórica con la noción externista de la justificación, puesto que una postura como la de Bonjour considera que la noción de la justificación tiene un valor instrumental. A pesar de que en este capítulo he revisado importantes críticas a la visión estándar de la racionalidad, y he visto cómo el externismo y algunas nociones internistas se comprometen con una visión consecuencialista de la racionalidad, no he

explorado alguna teoría del razonamiento que se comprometa explícitamente con la postura consecuencialista o, al menos, sostenga que la visión estándar de la racionalidad no es la única teoría normativa que explique la racionalidad. Por tal motivo, y algunos otros que he mencionado en estas conclusiones, en el próximo capítulo exploraré con cierto detalle la teoría dual de sistemas.

V.
LA TEORÍA DUAL DE SISTEMAS

Como he señalado anteriormente, una versión de la interpretación pesimista de la racionalidad humana defendida por la tradición de heurística y sesgo, y una versión de la interpretación optimista defendida por los psicólogos evolucionistas pueden ser compatibles[1]. Además, he argumentado que tanto la tradición de heurística y sesgo como la psicología evolucionista comparten la visión estándar de la racionalidad con la cual se evalúa el razonamiento humano. Algunos psicólogos y filósofos han sostenido que los rasgos que tienen en común estas teorías pueden ser recogidas en una familia de teorías de procesamiento dual acerca de los mecanismo mentales que subyacen al razonamiento y toma de decisiones (Evans, 2003). Aunque esas teorías difieren ligeramente entre cada una de ellas, en general sostienen que todos los procesos de razonamiento están subsumidos en dos diferentes tipos de sistemas: uno de estos sistemas es rápido, automático, en gran medida inconsciente y requiere de poca capacidad cognitiva; el otro sistema, es relativamente lento, consciente y requiere de mayor capacidad cognitiva[2]. En este capítulo exploro y evalúo las tesis centrales de la teoría dual de sistemas, y sugiero que los distintos problemas de razonamiento expuestos a lo largo de este libro pueden explicarse y entenderse de manera más adecuada a través de esta teoría psicológica. A mi juicio, la teoría dual de sistemas puede lidiar tanto con el pesimismo proveniente de la traición de heurística y sesgo, como con el optimismo proveniente de la psicología evolucionista, porque asume dos visiones distintas para evaluar el razonamiento humano.

La estructura del capítulo es la siguiente. En una primera sección expongo las tesis centrales de la teoría dual, para luego evaluar cómo esta teoría explica la interpretación que se ha dado a los experimentos provenientes de la tradición de heurística y

sesgo. Una vez señalado cómo la teoría dual de sistemas se enfrenta con la evidencia que, se cree, apoya la interpretación pesimista de la racionalidad humana, presento algunas de las dificultades a las que esta teoría se enfrenta. La teoría dual asume explícitamente distintas visiones de la racionalidad, por lo que al final de este capítulo expondré el tipo de compromisos normativos presupuestos por la teoría dual.

V.1. LA TEORÍA DUAL DE SISTEMAS

En esta sección mostraré algunas de las tesis centrales de la teoría descriptiva del razonamiento defendida por la teoría dual de sistemas, y para ello me centraré en la teoría defendida por Evans y Over (1996), Stanovich y West (2000, 2003) y Sloman (1996). Las coincidencias entre estos trabajos psicológicos sugiere que existe un relativo consenso acerca de lo que significa una teoría dual del razonamiento [3]. Pero antes de explorar cuáles son las tesis centrales de la familia de teorías que han defendido la teoría dual de sistemas es preciso indicar cuál es el propósito general de estas teorías.

El trabajo de Evans y Over (1996), como el trabajo de otros teóricos, está motivado por la existencia de una "aparente paradoja" en torno a la idea que se tiene de la racionalidad humana y una interpretación pesimista de los resultados provenientes de los estudios en psicología cognitiva. Esta paradoja surge al notar que, por una parte, los seres humanos cuentan con herramientas lingüísticas y de representación superiores a cualquier especie sobre la tierra; la información compartida y su capacidad de creación ha dado lugar a teorías científicas y una organización política compleja. En este sentido, los seres humanos son una especie altamente exitosa, lo que hace pensar que son racionales. Por otra parte, los seres humanos dentro de laboratorios psicológicos donde se evalúan sus habilidades de razonamiento, muestran que cometen errores sistemáticos de razonamiento que se desvían de los principios normativos adecuados.

> Parece haber una paradoja. Sobre la base de su comportamiento exitoso, los seres humanos son evidentemente muy inteligentes. Por otro lado, el estudio psicológico de la deducción parece sugerir que

> son ilógicos. Aunque algunos autores dedicados al estudio de los sesgos han sido cuidadosos al calificar sus afirmaciones acerca del comportamiento humano, otros han hecho afirmaciones bastante fuertes con relación a que su trabajo muestra que la gente es irracional (Evans y Over, 1996, p. 4, la traducción es mía).

La tensión entre una interpretación de los resultados llevados a cabo en los laboratorios psicológicos y la idea intuitiva que considera al ser humano como un ser racional, está en cierta forma detrás de las preocupaciones de diversos teóricos cognitivos, como los psicólogos evolucionistas. Para Evans y Over la solución de esta paradoja es uno de los propósitos centrales de la teoría dual de sistemas [4].

Evans y Over afirman que la cognición humana se funda en dos sistemas de razonamiento, uno de ellos implícito o tácito y otro explícito. El sistema tácito o implícito [5], de acuerdo con estos teóricos, es computacionalmente poderoso y presenta sólo sus productos finales a la conciencia. Si bien podría tener bases innatas es moldeado por interacciones con el ambiente, esto es, los procesos tácitos o implícitos pueden, para Evans y Over (1996, p. 174), ser innatos o aprendidos, pero aún en aquellos procesos que fueron modelados por el aprendizaje están sujetos a restricciones biológicas. En contraste, el sistema explícito es empleado en el razonamiento verbal de los sujetos, en el que conscientemente toman parte y en el que pueden ofrecer un reporte verbal. Los seres humanos son conscientes y, hasta cierto punto, pueden controlar los procesos explícitos de razonamiento, mientras que no pueden hacer lo mismo con los procesos del sistema tácito o implícito.

Si bien muchas veces se cree que el sistema explícito da cuenta de gran parte de la cognición humana, Evans y Over sostienen que en el razonamiento y la toma de decisiones de los sujetos el sistema tácito juega un papel central. El sistema tácito o implícito, según estos psicólogos, es responsable de la representación de información y del foco de atención del sujeto cognoscente —que Evans y Over denominan "procesos de pertinencia [6]". Otro rasgo que parece diferenciar a los sistemas del pensamiento es que resulta relativamente difícil la adquisición de ciertas habilidades explícitas de razonamiento, mientras que los procesos tácitos o

intuitivos son automáticos y adquiridos sin ningún esfuerzo (Evans y Over, 1996). Por último, es necesario indicar que, según estos defensores de la teoría dual, los procesos subyacentes al sistema tácito son compartidos con los animales, mientras que el sistema explícito es exclusivo de los seres humanos (Evans y Over, 1996, p 154 [7]).

La teoría dual de procesamiento defendida por Evans y Over responde a un modelo interactivo entre los dos sistemas de razonamiento [8]. Los diversos defensores de la teoría dual de sistemas consideran que la relación entre los dos sistemas de razonamiento es interactivo, sin embargo, no explican de manera precisa en qué consiste tal interacción. Es posible que parte del trabajo que desarrollarán los defensores de la teoría dual en el futuro se enfoque entre la relación de los dos sistemas de razonamiento. Los procesos responsables para la representación de problemas y de concentración son tácitos o implícitos (Evans y Over, 1996). Del mismo modo, creen Evans y Over, la mayoría de las inferencias y acciones suelen ser el resultado de procesos implícitos sin la intervención del sistema explícito. En el caso de la toma de decisiones, éstas pueden ser tomadas de modo intuitivo y rápido, o bien de una manera consciente y elaborada.

> En una tarea de razonamiento los sujetos podrían hacer un esfuerzo de razonamiento explícito de acuerdo con las instrucciones y podrían tener éxito al derivar la conclusión de esta manera. De manera alterna, podrían rendirse y escoger la conclusión que "sientan" correcta. La naturaleza interactiva de los procesos descansa en el hecho de que el pensamiento consciente está siempre moldeado, dirigido y limitado por procesos... tácitos (Evans y Over, 1996, p. 146, la traducción es mía).

Una de las ventajas de la teoría dual de sistemas es que el pensamiento consciente provee flexibilidad y previsión del futuro, que el sistema tácito no puede proveer.

> En la mayoría de ocasiones nuestras decisiones son automáticas y habituales, determinadas por el aprendizaje previo, pero no tiene que ser necesariamente de esta manera. Podemos tomar decisiones conscientes fundadas en el análisis de un nuevo problema, la proyección de modelos mentales de mundos futuros posibles y los

> cálculos de riesgos, costos y beneficios... Sin embargo, el punto crucial está en que la conciencia nos da la posibilidad, tanto de lidiar con lo novedoso como de anticipar el futuro (Evans y Over, 1996, p. 154, la traducción es mía).

Además de Evans y Over, otro dueto de psicólogos ha defendido de manera relativamente independiente una teoría dual de razonamiento. Como Evans y Over, Stanovich y West (2000, 2003) sostienen que todos los procesos de razonamiento y toma de decisiones están subsumidos en dos diferentes tipos de sistemas: uno de estos sistemas (Sistema 1 o S1) es rápido, holista, automático, en gran medida inconsciente y requiere de poca capacidad cognitiva, este sistema podría ser modular y parcialmente encapsulado [9]; el otro sistema (Sistema 2 o S2), por su parte, es relativamente lento, basado en reglas y requiere de mayor capacidad cognitiva [10]. Stanovich y West afirman que una clara distinción entre S1 y S2 se hace evidente en muchos de los experimentos usados por la tradición de heurística y sesgo en donde los dos sistemas señalan diferentes respuestas. Estos psicólogos ven un caso claro de lo anterior en el problema de "Linda, la cajera feminista" en donde los dos sistemas apuntan a respuestas diferentes.

Stanovich especula que S1 es en su mayoría innato, el cual surgió relativamente rápido en la evolución humana y —como de algún modo los psicólogos evolucionistas sugieren— ha sido diseñado por medio de la selección natural (Stanovich, 1999). S2, en cambio, evolucionó recientemente, es mayormente influenciado por la cultura y la educación formal, y es frecuentemente utilizado para lidiar con problemas presentados por la sociedad contemporánea tecnológicamente avanzada [11]. Debido a que el nuevo sistema requiere de mayor capacidad cognitiva, además de estar más influenciado por la cultura y la educación, Stanovich y West argumentan que podría haber diferencias particulares entre la habilidad o inclinación de los sujetos cognoscentes para usar dicho sistema. Específicamente, Stanovich sostiene que los sujetos con más alta capacidad cognitiva, como es medida en instrumentos como la prueba *Scholastic Aptitud Test* [12] (SAT, por sus siglas en inglés), realizarán tareas de mejor manera —tareas que el viejo sistema no está diseñado para resolver. Con lo anterior, se sugiere que los sujetos con mayor puntuación en SAT tienen mayor habi-

lidad para desarrollar tareas en donde interviene S2. En diversos estudios de estas tareas, Stanovich ha mostrado que mientras que el promedio de la ejecución de estas tareas es muy pobre, los sujetos que responden a la respuesta correcta tienen significativamente una puntuación alta en SAT.

Un punto que es importante señalar es que, también para Stanovich y West, la teoría dual de sistemas debe entenderse como una continua interacción entre los dos sistemas. La interacción surge del hecho de que S2 es moldeado por S1. S2 requiere de S1 para ahorrar tiempo e identificar posibles problemas o errores. El punto está en que S1 es primario, en el sentido de que permea todo nuestro pensamiento (Stanovich y West, 2000). Sin embargo, el hecho de que S1 sea central en el razonamiento humano no significa que los resultados producto de S1 no puedan ser revisados o corregidos por S2. "Una de las funciones del Sistema 2 es servir como un sistema de control para algunos de los resultados automáticos y mandatorios provistos por el Sistema 1" (Stanovich y West, 2000, p. 662, la traducción es mía).

Otro defensor de la teoría dual del razonamiento ha sido Steven Sloman (1996). Este psicólogo cognitivo ha distinguido entre el razonamiento asociativo —el cual asume que está basado en sistemas conexionistas— y un razonamiento basado en reglas [13]. Las características centrales del sistema asociativo son la contigüidad y la similaridad (Sloman, 1996, p. 15). Sloman sostiene que en el razonamiento asociativo los seres humanos generalmente no son conscientes de los procesos que aquí interviene sino solamente del producto de éstos [14]. Con ello, de alguna manera se relaciona al sistema asociativo con procesos automáticos, mientras que el sistema basado en reglas se relaciona con procesos controlados (Sloman, 1996, p. 6). Parte de la importancia del trabajo de Sloman consiste en mostrar la evidencia empírica y la plausibilidad de la teoría dual de sistemas, por ejemplo, con relación a los trabajos de categorización en psicología cognitiva.

Un apoyo sustancial que proporciona fundamentos suficientes para defender la idea de que existen dos sistemas de razonamiento se encuentra, según Sloman, en aquellas tareas de razonamiento que satisfacen lo que este psicólogo denomina "criterio S" (1996). El criterio S es satisfecho por aquellos problemas de razonamiento que hacen que un mismo sujeto crea simultáneamente

en dos respuestas contradictorias. Sloman expone una serie de problemas de razonamiento —muchos de ellos usados por la tradición de heurística y sesgo— para ilustrar la satisfacción del criterio S [15]. En el caso de "Linda, la cajera feminista" la mayoría de los sujetos suelen considerar que la oración de "Linda es cajera en un banco y feminista" es más probable que la oración de "Linda es cajera en una banco". Sloman sostiene que el sistema asociativo, en tanto que trabaja con relaciones de similaridad, señala que "Linda es cajera y feminista" es más probablemente verdadera, ya que es mucho más cercana a la descripción de Linda presentada en la prueba. Sin embargo, la aplicación de las reglas de la probabilidad indican que la respuesta correcta es que es más probable la oración de "Linda es cajera en un banco". Una vez explicada la prueba a los sujetos, éstos aceptan, según Sloman, que cometen la falacia de la conjunción, pero ello no evita que exista una disposición a creer que "ser cajera y feminista" se acerca más a la descripción de Linda.

> Existen dos mecanismos que conducen a conclusiones divergentes. Por un lado, una heurística intuitiva conduce a la conclusión de que T & F es más probable. Por el otro lado, un argumento probabilístico conduce a la conclusión de que T es más probable. Los dos mecanismos tienen fuerza psicológica (Sloman, 1996, p. 12, la traducción es mía).

Un rasgo que comparte la postura de Sloman con las dos versiones de la teoría dual revisadas anteriormente es que defiende un tipo de interacción entre los dos sistemas de razonamiento. Si bien los dos sistemas pueden ofrecer respuestas contradictorias, muchas veces los dos sistemas interactúan para ofrecer una misma respuesta. De hecho, Sloman cree que muchas veces es difícil distinguir cuál de los dos sistemas ha producido una respuesta particular debido a que generalmente los dos sistemas interactúan (1996, p. 5 [16]). A pesar de que los dos sistemas de razonamiento están interconectados, Sloman argumenta que el sistema basado en reglas puede llegar a suprimir al sistema asociativo, pero nunca lo puede inhibir completamente. Como se vio anteriormente, este rasgo es compartido, tanto por las teorías de Evans y Over como de Stanovich y West.

Las teorías duales del razonamiento ofrecidas por Sloman, Stanovich y West, y Evans y Over comparten rasgos comunes, lo que sugiere que existe un grupo de afirmaciones en el que los defensores de la teoría dual convergen. Los rasgos centrales de la teoría dual de razonamiento son que todos los procesos de razonamiento están subsumidos en dos diferentes tipos de sistemas: uno de estos sistemas es rápido, automático, en gran medida inconsciente y requiere de poca capacidad cognitiva; el otro sistema, por su parte, es relativamente lento, reflexivo y requiere de mayor capacidad cognitiva. Los psicólogos defensores de la teoría dual comparten también la idea de que existe una relación interactiva entre los dos sistemas de razonamiento, de hecho, en general sostienen que si bien el sistema tácito (S1 o sistema asociativo) muchas veces subyace a gran parte del razonamiento, el sistema explícito (S2 o sistema basado en reglas) puede corregir o inhibir hasta cierto punto al sistema tácito. En la siguiente página presento un cuadro sinóptico en donde se aprecian algunas de las características centrales de los dos sistemas de razonamiento.

Si bien es claro que existe una gran similitud entre las diferentes teorías duales, es preciso indicar algunas de las diferencias entre los distintos acercamientos. Los defensores de la teoría dual enfatizan diversos aspectos teóricos para explicar cuestiones particulares. Por ejemplo, Stanovich y West (2000) enfatizan el papel que juegan las diferencias individuales en la solución de problemas de razonamiento. De hecho, estos psicólogos presentan a la teoría dual como una interpretación que se puede dar al patrón consistente de tareas expuestas por la tradición de heurística y sesgo, donde puede defenderse una construcción alternativa por parte de lo sujetos. Stanovich y West afirman:

> para el propósito de la presente discusión, la diferencia más importante entre los dos sistemas es que tienden a conducir a diferentes tipos de construcción de tareas. Las construcciones puestas en marcha por el Sistema 1 son altamente contextualizadas, personalizadas y socializadas... En contraste, los procesos más controlados del Sistema 2 sirven para descontextualizar y despersonalizar los problemas (Stanovich & West, 2000, p 658-59, la traducción es mía).

Evans & Over (1996)	Procesos de pensamiento tácito *Propiedades:* Procesos implícitos, computacionalmente poderosos, operan de modo rápido (p. 143), pueden ser innatos o aprendidos (p. 147), son más antiguos evolutivamente hablando, son compartidos con algunos mamíferos superiores, son robustos e invariantes (p. 150), son no-verbales (p. 157).	Procesos de pensamiento explícito *Propiedades*: Procesos cognitivos explícitos, que están sujetos a la influencia de instrucciones verbales (p. 143), son procesos lentos, con capacidades limitadas (p. 153), procesos conscientes (p. 161), son procesos propiamente humanos.
Stanovich & West (2000, 2003)	Sistema 1 *Propiedades*: Procesos automáticos, en gran medida inconscientes que demandan relativamente poca capacidad computacional (2000, p. 658). Son procesos universales, compartidos con todos los seres humanos (2000, p. 660). Personalizan y contextualizan la construcción de tareas. Mecanismos modulares de dominio específico (2003, p. 2), parcialmente encapsulados (2003, p. 8)	Sistema 2 *Propiedades*: Procesos cercanos a la inteligencia analítica, controlados y que demandan de capacidad cognitiva (2000, p. 659). Despersonalizan y descontextualizan la construcción de tareas. Es un sistema que puede controlar algunos de los resultados provenientes de S1
Sloman (1996)	Sistema asociativo *Propiedades*: Procesos asociativos, automáticos (p. 15). El principio computacional que lo rige es de tipo asociativo, esto es, sus cómputos reflejan similaridad en estructura y relaciones de contigüidad temporal (p. 4). Algunas veces concuerdan con las reglas sin seguirlas explícitamente (p. 5).	Sistema dependiente de reglas *Propiedades*: Procesos dependiente de reglas, controlados (p. 15), en el que algunas veces se tiene conciencia tanto de los resultados de los procesos como de los procesos mismos (p. 6).

Una diferencia de mayor grado entre las distintas versiones de la teoría dual expuestas aquí es la relación de estas teorías con visiones particulares de la racionalidad. En el trabajo de Evans y Over, así como en el de Stanovich y West, se hacen vínculos explícitos entre la teoría dual de razonamiento con diversas maneras de entender la racionalidad. Evans y Over conectan su postura con una teoría normativa de la racionalidad, que se diferencia del vínculo que Stanovich y West ven entre la teoría dual del razonamiento y la racionalidad. Sloman, por su parte, no intenta hacer alguna conexión explícita de su teoría con alguna teoría de la racionalidad. Veamos.

a) Evans y Over (1996) consideran que la paradoja entre la inteligencia humana y los resultados de los estudios sobre el razonamiento y su interpretación puede ser explicada a partir de una distinción entre dos formas de racionalidad, viz., una manera personal y otra impersonal de la racionalidad. La "racionalidad personal" es usada básicamente para evaluar los sistemas tácitos, mientras que la "racionalidad impersonal" es usada para evaluar el sistema explícito. De hecho, Evans y Over (1996, p. 141) sostienen que su teoría dual de la racionalidad lleva a la teoría dual del razonamiento, y viceversa [17]. Desde la visión impersonal de la racionalidad o racionalidad 2, un sujeto es racional si sistemáticamente su razonamiento satisface los principios normativos provenientes de la lógica, la matemática y la teoría de la probabilidad (o bien de otro estándar normativo). Sin embargo, este acercamiento se complementa desde la visión personal de la racionalidad o racionalidad 1, en donde un sujeto es racional si su razonamiento sistemáticamente es una ayuda confiable para alcanzar sus metas personales. Para Evans y Over, ni la racionalidad personal ni la impersonal, tomadas individualmente, dan cuenta de manera cabal de la noción de racionalidad.

Bajo esta perspectiva, la respuesta de un sujeto a un cierto problema cognitivo podría ser vista como irracional desde la visión de la racionalidad 2, pero podría ser racional desde la racionalidad 1. Se puede sostener que para Evans y Over, el producto del sistema tácito será un resultado generalmente racional desde el punto de vista de la racionalidad 1, pero no necesariamente será racional desde el punto de vista de la racionalidad 2.

Del mismo modo, el buen procesamiento del sistema explícito dará resultados que generalmente son racionales según la racionalidad 2, pero no necesariamente racionales según la racionalidad 1 [18].

b) Stanovich y West (2000) distinguen entre *1)* adaptación evolucionista y *2)* la racionalidad instrumental (maximización de la utilidad dadas ciertas metas y creencias). La explicación de esta distinción que hacen estos psicólogos es:

1) La optimización adaptativa se encuentra a nivel de los genes. En términos de Dawkins, según Stanovich y West (2000), la adaptación evolucionista concierne a la optimización de los procesos relevantes también llamados replicadores [19] (o genes).

2) La maximización es a nivel de los individuos. En términos de Dawkins, según Stanovich y West (2000), la racionalidad instrumental se interesa por la maximización de la utilidad de los vehículos que dan cabida a los genes, esto es, los individuos [20].

Mientras a *2)* se le puede denominar, según estos psicólogos, "racionalidad normativa", a *1)* se le denomina "racionalidad evolucionista". A este respecto estos psicólogos afirman:

> Planteamos como hipótesis que los rasgos del Sistema 1 están diseñados muy cercanamente para buscar los incrementos en la probabilidad de reproducción de los genes. El Sistema 2, aun cuando es un producto evolutivo, es también primariamente un sistema de control concentrado en los intereses de la persona en su conjunto. Es el maximizador primordial de la utilidad personal del individuo... Maximizar esto último ocasionalmente resultará en el sacrificio de la adecuación genética... (Stanovich y West, 2000, pp. 660-61, la traducción es mía).

Si bien esta distinción puede verse como dos tipos distintos de racionalidad, Stanovich y West consideran que también se puede entender como dos procedimientos de optimización operando en un nivel personal y subpersonal. Mientras que la racionalidad normativa se podría ver como el primero de estos procedimientos de optimización, la racionalidad evolucionista se podría ver como el segundo de estos procedimientos (Stanovich y West, 2000, p. 661 [21]).

En principio, haciendo esta distinción entre dos tipos de racionalidad supone que lo que puede ser racional desde el punto de

vista de la racionalidad evolucionista, podría no serlo desde el punto de vista de la racionalidad normativa, y viceversa. Pero lo anterior no supone que todas las respuestas dadas por S1 sólo satisfagan la racionalidad evolucionista, sino que en la mayoría de situaciones de la vida cotidiana los procesos de este sistema también satisfacen la racionalidad normativa. Así, Stanovich y West, ven como una última diferencia entre los dos sistemas de razonamiento sus conexiones con la racionalidad que cada una de estos sistemas tiene:

> La última diferencia entre el Sistema 1 y el 2... es que el Sistema 1 instancia metas genéticas de corto alcance, mientras que el Sistema 2 instancia una jerarquía de metas que están orientadas a la maximización de la satisfacción de metas a nivel del organismo en su conjunto (2003, p. 10, la traducción es mía).

El siguiente cuadro representa los diferentes vínculos entre los sistemas de razonamiento y las visiones de la racionalidad.

Evans y Over (1996)	Procesos de pensamiento tácito	Racionalidad 1 o racionalidad personal El grado con el cual los sujetos confiablemente alcanzan sus metas individuales (p. 8).
	Procesos de pensamiento explícito	Racionalidad 2 o racionalidad impersonal El grado con el cual los sujetos ejecutan inferencias o actúan por razones sancionadas por algún estándar normativo y abstracto (p. 8).
Stanovich y West (2000)	Sistema 1	Racionalidad evolucionista Optimización a nivel del replicador mismo (los genes) (p. 661).
	Sistema 2	Racionalidad normativa Optimización a nivel del vehículo (el individuo) (p. 611).

En suma, si bien existe una similitud entre los rasgos con los cuales se distinguen los dos sistemas de razonamiento, la diferencia más importante radica en los compromisos que las distintas versiones de la teoría dual —principalmente por parte de Evans y Over, y de Stanovich y West— tienen con la racionalidad. En tanto que el propósito de este libro es conocer los vínculos entre la racionalidad y la noción de justificación a través de distintas teorías del razonamiento, los diferentes estándares de racionalidad que proponen los defensores de la teoría dual son de capital importancia. Como argumentaré en la sección final de este capítulo, las diferentes maneras de relacionar la teoría dual del razonamiento con la racionalidad implican compromisos con dos distintos acercamientos a la racionalidad, viz., el acercamiento deontológico y consecuencialista a la racionalidad.

V.2. LA TEORÍA DUAL DE SISTEMAS Y LA INTERPRETACIÓN PESIMISTA DE LA RACIONALIDAD HUMANA

En esta sección expondré cuál es la salida que ofrece la teoría dual a la interpretación pesimista de la racionalidad humana [22]. En primer lugar, quiero enfatizar que los defensores de esta teoría psicológica argumentan a favor de la tesis de que el ser humano razona correctamente, a pesar de la evidencia proporcionada por la tradición de heurística y sesgo. Por ese motivo, la teoría dual intenta ofrecer una explicación del razonamiento humano donde se tome en cuenta tanto la evidencia proporcionada por la tradición de heurística y sesgo que se piensa apoya la idea de que los seres humanos son irracionales, como la evidencia proporcionada por la psicología evolucionista, que señala que los seres humanos muchas veces razonan de manera correcta. Para Evans y Over (1996), como se verá en esta sección, esta explicación depende en gran medida de dar cabida a dos distintos criterios para evaluar el razonamiento humano.

¿Cómo explica la teoría dual los sesgos o las desviaciones normativas de los sujetos en los experimentos que supuestamente apoyan la visión pesimista de la racionalidad humana? En las primeras formulaciones de la teoría dual de sistemas, Evans (1991) sostenía que el razonamiento se divide en dos etapas. En una primera etapa, el problema de razonamiento al que se enfrenta un sujeto es procesado por heurísticas preconscientes que tienen

como resultado codificar los rasgos psicológicamente relevantes de dicho problema. En la segunda etapa, intervienen procesos de razonamiento analítico que toman en cuenta la competencia lógica del sujeto.

> El punto es que la competencia manifiesta en el razonamiento está críticamente determinada por la representación de relevancia producida por los pre-procesamientos heurísticos. De ahí que los sesgos surgirán, si los rasgos lógicamente críticos no son vistos como relevantes, o si los sujetos atienden a aquello que es lógicamente irrelevante (Evans, 1991, p. 92, la traducción es mía).

Según esta primera formulación, el origen de los sesgos se debe no a que la competencia de razonamiento sea inadecuada, sino a que el sistema de razonamiento tácito (lo que Evans (1991) denominaba "razonamiento heurístico") proporciona datos no relevantes al sistema de razonamiento explícito.

La teoría dual de sistemas defendida posteriormente por Evans y Over (1996) conserva la manera de entender el razonamiento humano defendida por Evans (1991), en donde las tareas de razonamiento se pueden dividir en dos etapas. Para estos psicólogos un conjunto de procesos pertenecientes al sistema tácito o S1 —denominados procesos de relevancia— producen una representación de la información relevante que puede posteriormente ser usada por procesos explícitos y conscientes.

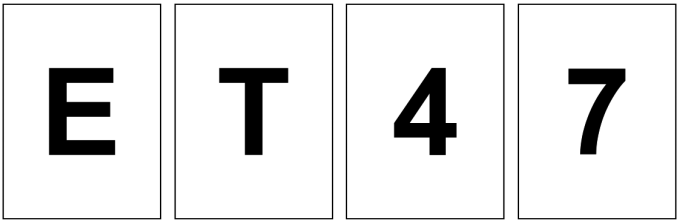
> Nuestra idea básica es que el pensamiento explícito o consciente está centrado en representaciones altamente seleccionadas que parecen "relevantes", pero esta relevancia está determinada por procesos preconscientes y tácitos (Evans y Over, 1996, p. 48, la traducción es mía).

A diferencia de la primera formulación defendida por Evans (1991) estos psicólogos añaden la distinción entre dos tipos de racionalidad. La racionalidad personal o racionalidad 1 y la racionalidad impersonal o racionalidad 2 juegan el papel de dos teorías normativas que posteriormente son usadas para evaluar si los sujetos razonan de manera adecuada. Precisamente esta distinción es central para entender la relación entre la teoría dual de sistemas y la interpretación pesimista de la racionalidad humana.

Los sesgos que cometen los seres humanos en los laboratorios, los cuales han sido usados como evidencia para apoyar la inter-

pretación pesimista de la racionalidad humana son, para Evans y Over, errores sistemáticos relativos a un estándar normativo proveniente de una teoría formal. Por lo tanto, según estos psicólogos, los sesgos son propios de la racionalidad impersonal o racionalidad 2. Lo que estos psicólogos sostienen es que muchos sesgos son productos de procesos que son adaptativos en el mundo real (fuera de los laboratorios), por lo que son eficaces para satisfacer las metas de los sujetos (Evans y Over, 1996, p. 49). En otras palabras, lo que estos psicólogos defienden es que los llamados errores de razonamiento o sesgos pueden ser originados por procesos tácitos o implícitos o bien por la combinación de estos con los procesos explícitos, que si bien son considerados como "errores" o "sesgos" según la racionalidad 2, según la racionalidad 1 tienen un carácter racional. Veamos con cierto detalle un ejemplo de lo anterior.

En la tarea de selección de tarjetas, como expuse el segundo capítulo, se le presentan a los sujetos diferentes tarjetas y se les pregunta cuáles de éstas tienen que voltear para comprobar la verdad de una oración condicional (de la forma "si..., entonces"). En el ejemplo revisado anteriormente se les presenta a los sujetos cuatro tarjetas como aparecen a continuación y posteriormente se explica lo que deben realizar.



Todas las tarjetas anteriores tienen de un lado un número y del otro una letra del abecedario. Tu trabajo es determinar cuáles de estas tarjetas tienes que voltear para probar la verdad de la oración: si hay una E en un lado de la carta, entonces hay un número 4 en su otro lado.

Wason y otros psicólogos [23] descubrieron que comúnmente los sujetos no responden correctamente a este tipo de tareas, pues aun cuando la respuesta correcta a este problema son las tarjetas E y 7, la mayoría de sujetos sólo consideran que se debe voltear la tarjeta E, o las tarjetas E y 4. Evans y otros psicólogos señalaron que ejemplos como este daban lugar al "sesgo de emparejamiento" (*matching bias*). Los sujetos simplemente seleccionan las tarjetas E y 4 porque ellas están mencionadas en la oración condicional de la que deben probar su verdad. Como la oración a probar se encuentra en un condicional, la partícula "si" llama la atención de lo sujetos a los estados de cosas que hacen verdaderas la oración, en este caso la tarjeta E. La hipótesis que desean probar los defensores de la teoría dual es que los procesos de relevancia (procesos que pertenecen al sistema tácito o S1) son los responsables de que los sujetos dirijan su atención a las tarjetas E y 4 que son mencionados en el condicional.

Los argumentos que ofrecen Evans y Over para señalar que los procesos de relevancia son los que determinan la elección de las tarjetas se funda en dos grupos de experimentos. El primero de ellos consiste en mostrar que la respuesta de los sujetos mejora notablemente en las versiones de la selección de tarjetas en donde aparecen negaciones explícitas. En la versión ordinaria de la tarea de selección, como la expuesta anteriormente, se tiene que probar la verdad de las oración "si hay una E en un lado de la carta, entonces hay un número 4 en su otro lado", en un caso de negación explícita se ofrecen las mismas cartas reformulando la oración a probar como "si hay una E en un lado de la carta, entonces no hay un número 7 en su otro lado". Esta reformulación de la tarea de selección mejora significativamente los errores de razonamiento de los seres humanos, eliminando el sesgo de emparejamiento (Evans y Over, 1996, p. 58).

El segundo conjunto de experimentos usados para apoyar la idea de que los sujetos hacen uso de procesos tácitos en la tarea de selección de tarjetas es mostrando que los sujetos ponen atención sólo a las tarjetas que de hecho seleccionan. Después de diversos experimentos [24] Evans y Over (1996) descubrieron que en la tarea de selección las tarjetas que no son seleccionadas simplemente no son tomadas en cuenta, los sujetos simplemente rechazan estas tarjetas ya que no analizan qué ocurriría si dichas

tarjetas se voltearan. De este modo, estos experimentos señalan que las tarjetas seleccionadas son de hecho las cartas que son inspeccionadas por más tiempo, mientras que las cartas que no se seleccionan son escasamente inspeccionadas (Evans y Over, 1996, p. 59). El razonamiento explícito o analítico es utilizado escasamente para resolver problemas como la tarea de selección, pues éste no parece determinar la elección de las tarjetas.

> Los sujetos realmente ocupan el razonamiento analítico en la tarea de selección pero *eso de hecho no afecta la elección que realizan*. Primeramente, las tarjetas que no son seleccionadas no son atendidas. De ahí que el razonamiento analítico no es usado para rechazar las tarjetas. En segundo lugar, cuando los sujetos realmente piensan sobre una carta, ellos se convencen a sí mismo de seleccionarla (Evans y Over, 1996, p. 60, la traducción es mía).

Hasta aquí estos psicólogos han explicado cómo en la tarea de selección de tarjetas intervienen procesos de relevancia propios del sistema de razonamiento implícito (o de S1), sugiriendo además que los procesos explícitos no parecen ser determinantes para la elección de las tarjetas. Sin embargo, Evans y Over deben explicar en qué sentido el uso de tales procesos hace que la ejecución de razonamiento en la selección de tarjetas sea racional.

El argumento usado por estos psicólogos para mostrar que el sesgo de emparejamiento —y en general otros sesgos estudiados por los psicólogos— es racional, es que dicho sesgo es producto de mecanismos tácitos o implícitos útiles para alcanzar las metas personales de los seres humanos.

> La noción de racionalidad 1 está incrustada en la idea de que hemos evolucionado de una manera adaptativa que nos permite alcanzar metas en el mundo real, no solucionar problemas de laboratorio. De hecho, el sesgo de emparejamiento es una consecuencia de mecanismos útiles que nos llevan a dirigir nuestra atención de modo apropiado cuando entendemos el lenguaje natural (Evans y Over, 1996, p. 61, la traducción es mía).

En tanto que la racionalidad personal evalúa el grado con el cual los seres humanos satisfacen sus metas personales y los mecanismos de relevancia que producen el sesgo del emparejamiento

sirven para satisfacer las metas de los seres humanos, los resultados proporcionados por la tarea de selección de tarjetas sólo sugieren que los seres humanos no satisfacen las normas provenientes de un estándar formal (proveniente en este caso de la lógica). Por el contrario, habría un sentido en el cual el sesgo del emparejamiento es racional [25], puesto que los procesos cognitivos que lo causan son medios eficaces para alcanzar la metas particulares de los sujetos.

Al igual que con la respuesta que dan Evans y Over al sesgo producto de la tarea de selección, estos psicólogos revisan otros sesgos provenientes de diversos experimentos que han sido usados para apoyar la interpretación pesimista de la racionalidad humana [26]. El argumento de Evans y Over es básicamente el mismo que el revisado para el sesgo de emparejamiento. Intentan primero señalar cómo es que S1 o el sistema tácito interviene en estos experimentos, para después mostrar en qué sentido la intervención de procesos pertenecientes a S1 son eficientes para alcanzar las metas y objetivos de los sujetos. Es aquí en donde se puede entender cómo en gran parte, según los defensores de la teoría dual de sistemas, la clave para entender que lo seres humanos son racionales, aun cuando existe evidencia que parece socavar esta idea, se encuentra en la aceptación de dos tipos de racionalidad.

> Un análisis desde la racionalidad 1 es el punto de inicio y además el que provee una perspectiva desde la cual podemos entender mucho del aparentemente extraño comportamiento en la literatura sobre razonamiento y toma de decisiones (Evans y Over, 1996, p. 21, la traducción es mía).

Pero, ¿cómo se relaciona la respuesta ofrecida por la teoría dual de sistemas con la interpretación pesimista de la racionalidad humana? Como argumenté anteriormente, existen varias versiones de la interpretación pesimista, pero sólo la versión (I) podría ser correcta dada la evidencia proporcionada por la psicología evolucionista. La versión pesimista (I) sostiene que muchas veces en que los juicios se desvían de las normas apropiadas de racionalidad se debe a que los juicios de los sujetos se fundan en heurísticas que algunas veces conducen a errores sistemáticos y severos. Por su parte, la psicología evolucionista, que ofrece un

panorama optimista de la racionalidad humana, puede defender adecuadamente una interpretación optimista, a saber, la versión optimista (A): muchos de los casos en los cuales los juicios de los sujetos cognoscentes concuerdan con las normas de racionalidad apropiadas son explicadas por el hecho de que al hacer tales juicios, los seres humanos activan módulos mentales que hacen un buen trabajo cuando son provistos de los insumos adecuados. A diferencia de la tradición de heurística y sesgo y la psicología evolucionista que se comprometen con la visión estándar de la racionalidad, la teoría dual de sistemas asume dos teorías normativas de la racionalidad [27]. La distinción que hacen los psicólogos defensores de la teoría dual entre la racionalidad personal y la racionalidad impersonal da pie a una respuesta que se encuentra en medio de la versión pesimista (I) y la versión optimista (A) del razonamiento humano.

A mi juicio, la teoría dual de sistemas puede integrar en una sola explicación a la interpretación optimista y a la interpretación pesimista de la racionalidad. Esta reconciliación es factible porque al final del tercer capítulo he argumentado que la versión (I) de la interpretación pesimista de la racionalidad humana, así como la versión (A) de la interpretación optimista de la racionalidad humana pueden ser compatibles. Si se observan los sesgos de razonamiento, por ejemplo el sesgo del emparejamiento, se tienen elementos para pensar que los seres humanos son irracionales 2, pero no irracionales 1. En este sentido, Evans y Over, así como los otros defensores de la teoría dual de sistemas, no están negando la existencia de sesgos de razonamiento, es decir, no están negando que muchas veces los seres humanos se aparten de los estándares normativos adecuados. En otras palabras, los defensores de la teoría dual de sistemas aceptarían la versión pesimista (I). De hecho, algunas veces el razonamiento humano es irracional, pero sólo desde la perspectiva de la racionalidad 2 o racionalidad impersonal. Sin embargo, hay un sentido, el evaluado por la racionalidad personal, en el que los seres humanos son racionales y este sentido tiene que ver con la eficacia en que alcanzan las metas que desean. Es decir, los defensores de la teoría dual de sistemas aceptarían la versión optimista (A). No obstante, lo anterior no implica que los seres humanos siempre sean racionales [28] como suponen las versiones más optimistas.

La idea que deseo subrayar aquí es que la interpretación que se hace del razonamiento humano —por ejemplo, aquel observado en los experimentos revisados en el segundo capítulo de este libro— depende en gran medida del estándar bajo el cual se evalúa el razonamiento. La postulación de dos sistemas de razonamiento y de dos visiones distintas de la racionalidad no responde a la interpretación pesimista de la racionalidad humana mostrando que los seres humanos *finalmente* son racionales, sino que presenta una descripción empírica y normativa en donde se puede entender en su justa dimensión el sentido en que el razonamiento humano es algunas veces racional y algunas veces irracional. De este modo, la teoría dual de sistemas podría integrar el optimismo proveniente de la psicología evolucionista y el pesimismo proveniente de la tradición de heurística y sesgo, haciendo uso de dos sistemas de razonamiento y postulando dos teorías normativas de la racionalidad.

La importancia de combinar distintas visiones de la racionalidad en la teoría dual responde a dos necesidades. Primeramente, es un modo de lidiar con la paradoja de la racionalidad, esto es, con la idea de que los seres humanos son muy exitosos para alcanzar ciertas metas en el mundo real, junto con la interpretación que usualmente se hace de la evidencia psicológica que muestra que son irracionales. En segundo lugar, combinar dos visiones de la racionalidad supone por parte de la teoría dual una manera de dejar de lado la visión estándar de la racionalidad como única teoría normativa. En el capítulo anterior presenté algunos argumentos que muestran que la visión estándar de la racionalidad no puede brindar una explicación adecuada de las normas de razonamiento. Sin embargo, hasta ese momento no había explorado alguna teoría psicológica que rechazara o combinara la visión estándar de la racionalidad con otra teoría normativa. Así, la teoría dual de sistemas combina varias de las estrategias revisadas en capítulos anteriores para aceptar la versión optimista (A) defendida por los psicólogos evolucionistas y la versión pesimista (I) defendida por la tradición de heurística y sesgo. Esta es, bajo mi perspectiva, para tomar prestada una frase de Samuels, Stich y Bishop (2002), la manera de "dar término a las guerras de la racionalidad". Por tal motivo, considero que la teoría

dual es la teoría de razonamiento que de momento resulta más apropiada.

V.3. ALGUNAS DIFICULTADES PRESENTES EN LA TEORÍA DUAL DE SISTEMAS

En las páginas anteriores he expuesto las tesis centrales de la teoría dual y he explicado cómo esta teoría da cuenta de la evidencia que ha sido usada para apoyar la interpretación pesimista de la racionalidad humana. Si bien en la actualidad parece formarse un consenso acerca de que la teoría dual de sistemas —o una teoría similar— es la teoría del razonamiento que mejor describe el modo en que los seres humanos razonan, es necesario señalar algunas de las dificultades que esta teoría tiene. Veamos.

La distinción entre S1 y S2 radica en que principalmente se desea contrastar procesos que son rápidos, automáticos y que requieren de poco esfuerzo cognitivo, de procesos que resultan lentos, secuenciales y controlados. Debido a la automaticidad y rapidez de los procesos que pertenecen a S1 se considera que dicho sistema es evolutivamente antiguo, es compartido con otros animales y que en términos generales es independiente de las diferencias individuales de inteligencia. Por su parte, S2 es evolutivamente reciente, distintivamente humano y tiene diferencias individuales sustanciales. Algunos psicólogos consideran que algunas de las diferencias entre estos dos sistemas no son tan claras como comúnmente se ha sugerido (Evans, 2006). A continuación presento algunas críticas hechas a la teoría dual.

1. Como apunté anteriormente, se ha sostenido que S1 es un sistema antiguo en términos evolutivos, mientras que S2 es un sistema moderno. En estudios neuropsicológicos —Evans (2006, p. 2003) cita el trabajo de Goel y Dolan— se señala que en los experimentos donde operan procesos que típicamente se considera que pertenecen a S1, las creencias surgen de la parte frontal del cerebro que en términos evolutivos es moderno, lo cual hace suponer que algunos procesos subyacentes a S1 no pueden ser "antiguos" en su origen. Otra crítica que podría señalar que S1 no es más antiguo que S2, evolutivamente hablando, es a partir de la idea de algunos psicólogos cognitivos —relacionados con filosofía de la mente— al considerar que en S1 se encuentran mecanismos

modulares, esto es, específicos de dominio y encapsulados (lo que ha llevado a muchos psicólogos a decir que S1 es específico de dominio, mientras que S2 es de dominio general (Evans, 2006, p. 204)). La crítica radica en que existe evidencia —para ello Evans (2006, p. 203) cita el trabajo de Mithen en arqueología cognitiva- de que un mecanismo encapsulado tal y como supone la noción de modularidad debió haber evolucionado más tardíamente. Si bien el trabajo de Mithen es altamente especulativo, su referencia es suficiente para señalar que si S1 está compuesto de mecanismos modulares, entonces es posible que dichos mecanismo sean evolutivamente más recientes que un procesador de dominio general.

2. El rasgo revisado en (1), es decir, que S1 podría ser un sistema que evolutivamente hablando es más antiguo que S2, está generalmente relacionado con la idea de que S1 es un sistema que los seres humanos comparten con cierto tipo de animales, mientras que S2 es un sistema que únicamente pertenece a los humanos [29]. Sin embargo, los defensores de la teoría dual han puesto poca atención a la literatura sobre cognición en animales. Citando el trabajo de Toates, Evans (2006, p. 202) sostiene que hay evidencia a favor de dos sistemas de razonamiento en los animales, a saber, una cognición fijada a los estímulos y otra cognición de orden superior. El primer tipo de cognición, según Toates, se encuentra en el condicionamiento y aplicación del comportamiento instintivamente programado y fijado por la naturaleza; mientras que la cognición de orden superior tiene como rasgo principal el control, esta cognición eventualmente se desarrolló o transformó en la conciencia en el caso de los seres humanos. Evans considera que si Toates está en lo correcto, no sería posible sostener la idea de que la distinción de dos tipos de razonamiento es únicamente humana, a lo mucho tendría que defenderse la idea de que S2 está mayormente desarrollado en los seres humanos.

3. Como se ha expuesto anteriormente, se considera que los procesos subyacentes a S1 son inconscientes, mientras que aquellos que subyacen a S2 son conscientes y controlados. Evans reconoce que estos rasgos para distinguir a S1 de S2 han sido influenciados por los estudios en atención que sugieren que lo que no es implícito y automático, es consciente y se puede controlar. El problema radica en que no parece claro qué significa estar bajo

control (Evans, 2006, p. 203). Por un lado, parece que los procesos subyacentes en S1 controlan el comportamiento humano por *default*, a menos que un esfuerzo consciente y explícito intente controlarlo. Lo interesante parece ser que no importaría el esfuerzo puesto por S2 para controlar los procesos inconscientes, pues el resultado podría no ser exitoso. Por otro lado, el razonamiento analítico de S2 es muchas veces usado para racionalizar el comportamiento inconscientemente controlado. "Así, nuestra experiencia consciente y nuestras creencias expresadas... podrían algunas veces proveer una explicación a nivel intencional de nuestro comportamiento controlado por el Sistema 2, pero podría proveer una mera racionalización de comportamientos controlados por el Sistema 1. A mi parecer, no existe un dispositivo metodológico que confiablemente nos separe uno del otro" (Evans, 2006, p. 204, la traducción es mía).

4. Al parecer S2 es un sistema singular, secuencial que requiere de atención consciente, que es relativamente lento, etc., sin embargo, Evans (2006) considera que se han puesto en S1 todos los procesos que no se encuentran en S2. Tal parece que habría distintos mecanismos subyacentes a S1 entre ellos *i.*) procesos de aprendizaje experimental como los procesos asociativos; *ii.)* algunos mecanismos modulares —*à la* Fodor— como los sistemas perceptuales y el lenguaje, y *iii.)* procesos automatizados, es decir, aquellos procesos que fueron conscientes pero que posteriormente se hicieron automáticos (Sloman, 1996, Evans, 2006).

> Parece que hay procesos implícitos que fueron una vez conscientes, otros que nunca fueron conscientes pero que entregan un producto consciente y otros que influyen nuestro comportamiento sin ser consciente en ninguno de los sentidos. ¿Tiene sentido clasificarlos a todos dentro de un "Sistema 1"? Seguramente no. Todo lo que realmente liga a las teorías de procesamiento dual es la naturaleza del Sistema 2 y la manera en que los procesos automáticos e implícitos (de cualquier tipo) parecen competir con él por el control de nuestro comportamiento (Evans, 2006, p. 205, la traducción es mía).

Si lo anterior es correcto, no queda claro en qué sentido podemos denominar como "sistema" a S1, ni tampoco es claro por qué no se debería de aceptar una teoría múltiple de razonamiento en

donde cada uno de los procesos que en la actual teoría dual de sistemas pertenecen a S1 puedan ser entendidos como sistemas independientes.

De los señalamientos anteriores, Evans sostiene que de momento no se tiene una teoría coherente fundada en dos sistemas de razonamiento (Evans, 2006, p. 206). En un interesante artículo en donde se comenta solamente la teoría dual defendida por Sloman, Gigerenzer y Regier (1996) afirman que la distinción conceptual entre dos sistemas de razonamiento es, de momento, ambigua e imprecisa, rasgos que también Evans sostendría que posee la teoría dual de sistemas. Uno de los puntos interesantes del artículo de Gigerenzer y Regier es, además, que pone en duda el apoyo empírico ofrecido a favor de la teoría dual de sistemas [30]. Como expuse anteriormente, Sloman (1996) afirma que la satisfacción del criterio S, según el cual un sujeto sostiene al mismo tiempo dos creencias contradictorias, muestra la existencia de dos distintos sistemas de razonamiento. Para ilustrar la satisfacción del criterio S, Sloman recurre a los experimentos usados por la tradición de heurística y sesgo, entre ellos el problema de "Linda, la cajera feminista" (Stanovich y West, 2000, Sloman, 1996). A juicio de Gigerenzer y Regier (1996), la respuesta de que "Linda es cajera y feminista" podría ser explicada sin comprometerse con la existencia de un sistema de razonamiento complementario al sistema basado en reglas. Una explicación podría ser que las instrucciones ofrecidas a los sujetos por los psicólogos cognitivos son lo suficientemente ambiguas y ello da lugar a la elección de la opción "cajera y feminista [31]". La crítica de Gigerenzer y Regier parece correcta, al indicar que la satisfacción de lo que Sloman denomina el criterio S no apoya unívocamente la existencia de dos sistemas de razonamiento.

De lo anterior, es posible sostener que en la actualidad diversos acercamientos psicológicos que abordan la cognición están siendo estudiados y valorados no sólo a la luz de nueva evidencia empírica, sino de su coherencia conceptual. La teoría dual se encuentra dentro de estos acercamientos teóricos y dada la evidencia empírica usada para apoyar esta teoría sería difícil rechazarla únicamente por consideraciones *a priori*. El tipo de dificultades que esta teoría presenta marcarán los desarrollos que dicha teoría tenga en

el futuro (García Campos, 2008); de momento es necesario analizar cuáles serían las consecuencias filosóficas relacionadas con la racionalidad de aceptar una teoría dual del razonamiento.

V.4. LA TEORÍA DUAL DE SISTEMAS Y LA RACIONALIDAD

En este apartado deseo mostrar cómo la teoría dual de sistemas se compromete con diversas teorías de la racionalidad. La teoría dual de sistemas se presenta como una teoría descriptiva del razonamiento vinculada fuertemente a diferentes posturas acerca de la racionalidad, pero como expuse anteriormente, los defensores de la teoría dual no comparten los estándares con los que debe ser evaluado el razonamiento humano. En esta sección argumentaré que los postulados normativos propuestos por los defensores de la teoría dual los conducen a dos acercamientos distintos de la racionalidad, esto es, con el acercamiento deontológico y consecuencialista.

V.4.1. RACIONALIDAD PERSONAL Y RACIONALIDAD IMPERSONAL

Evans y Over (1996) sostienen que una teoría dual conduce a una visión dual de la racionalidad y viceversa. En esta sección muestro que lo que Evans y Over denominan racionalidad 1 o racionalidad personal es un criterio normativo de tipo consecuencialista, mientras que lo que denominan racionalidad 2 o racionalidad impersonal es un criterio normativo de tipo deontológico. Veamos.

La racionalidad 1 o la racionalidad personal es definida como el grado en el que los sujetos pueden alcanzar confiablemente sus metas personales (Evans y Over, 1996). En repetidas ocasiones, Evans y Over sostienen que el razonamiento humano no puede ser evaluado únicamente por la lógica formal. La visión estándar de la racionalidad, como reiteradamente he señalado, sostiene que ser racional es razonar conforme a los principios de la lógica y la matemática. Sin embargo, la racionalidad en general, según la teoría dual de Evans y Over, no puede entenderse como una teoría normativa que considere que los principios se fundan únicamente en las teorías formales. Lo anterior no implica que no haya reglas normativas que provengan de la lógica o de la teoría de la probabilidad que puedan usarse para evaluar el tipo de

razonamiento que llevan a cabo los sujetos cognoscentes, sino que la lógica no puede agotar el conjunto de reglas normativas.

> Como una analogía, considérese la distinción entre las reglas del ajedrez y la estrategia de jugar ajedrez. Alguien que meramente sabe las reglas, no sabe cómo jugar ajedrez, excepto en el sentido de producir un juego legal... el uso de la lógica misma en el razonamiento ordinario requiere de un buen juicio práctico para seleccionar una conclusión que sea relevante a la meta que deseamos obtener. La lógica no provee de mejor manera una estrategia de razonamiento de lo que lo hacen las reglas del ajedrez que te dicen cómo ser un buen jugador de ajedrez (Evans y Over, 1996, p. 19, la traducción es mía).

En esta cita se pude apreciar cómo para Evans y Over la lógica no puede abarcar en su conjunto a la racionalidad humana, pues requiere de buenos juicios prácticos que sean pertinentes para alcanzar las metas o deseos que el sujeto desea obtener. Es precisamente en este punto que es relevante la racionalidad 1, pues este tipo de racionalidad se funda en los medios confiables bajo los cuales un sujeto satisface sus metas. Pero si la lógica parece no ser suficiente para alcanzar las metas deseadas y la racionalidad 1 tiene como propósito señalar cuáles son los medios confiables para alcanzar una meta, entonces la dimensión normativa de la racionalidad 1 no puede estar esbozada por la visión estándar de la racionalidad.

El que la racionalidad 1 no pueda ser explicada por la lógica podría sugerir que la probabilidad o alguna otra teoría formal podrían dar cuenta de este tipo de racionalidad. De hecho, según Evans y Over, muchas veces se cree que la racionalidad 1 puede ser capturada por los principios de una teoría formal de la decisión. Sin embargo, estos psicólogos sostienen que una teoría de la decisión "tiene serias deficiencias para explicar el comportamiento que conduce a la consecución de metas personales" (Evans y Over, 1996, p 25, la traducción es mía). Evans y Over tratan de argumentar que la teoría de la decisión normativa —al igual que la lógica— no puede capturar cabalmente la noción de racionalidad 1 (Evans y Over, 1996, p. 148). Los seres humanos pueden ofrecer un reporte verbal de sus metas en la vida cotidiana, pueden tener conocimiento explícito de sus metas, por ejemplo, que están buscando comida. No obstante, según Evans y Over,

los seres humanos tienen un conocimiento restringido de los procesos que les ayudan a adquirir ciertas metas o de los procesos confiables de percepción que les ayudan a aprender del mundo. Al igual que otros animales, los seres humanos confían en gran medida en sus procesos implícitos como la percepción y el aprendizaje.

> Para ayudarnos a aprender cómo satisfacer nuestros deseos básicos hasta un punto razonable, no tenemos que tener representaciones mentales de probabilidades y utilidades, ni hacer cálculos en algún punto dado de todas las posibilidades con que contamos, ni confiar solamente en el aprendizaje explícito. El procesamiento tácito o implícito es muy característico de este tipo de racionalidad personal, y puede algunas veces estar en consonancia con los principios de la decisión teórica sin seguirla... (Evans y Over, 1996, p. 29, la traducción es mía).

La teoría de la decisión normativa supone, según Evans y Over, que los seres humanos saben todo acerca de las acciones que están bajo su poder, pero esto parece imposible. Algunas de las habilidades de los seres humanos les son desconocidas y requerirían de una investigación científica extensa para tener una idea de lo que son estas habilidades. Además, sostienen Evans y Over, existe para casi cualquier decisión un número ilimitado de consecuencias que son imposibles de anticipar, aun para acciones muy simples.

> ...no podemos hacer elecciones eficientes en tiempo real (aun tan simples como el de girar el coche a la derecha o a la izquierda) aplicando una teoría normativa de la decisión. Podemos aprender a estar en consonancia con esta teoría, sin siquiera seguir sus reglas para manipular proposiciones y hacer cálculos, en casos simples pero importantes. Ese es el motivo por el cual la mayoría de nosotros puede girar a la izquierda o a la derecha de manera segura (Evans y Over, 1996, p. 30, la traducción es mía).

Si la lógica y la teoría normativa de la decisión no dan cuenta de la racionalidad 1, entonces la dimensión normativa de la racionalidad 1 no puede estar esbozada por la visión estándar de la racionalidad ni tampoco puede estar comprometida con un acercamiento deontológico de la racionalidad. Si lo anterior es correc-

to, ¿es la racionalidad 1 un criterio normativo de tipo consecuencialista? La visión consecuencialista de la racionalidad, como expuse en el capítulo anterior, sostiene que lo que es constitutivo de un proceso cognitivo racional es que sea un medio eficiente para alcanzar una meta o fin. La racionalidad 1 se enmarca en el acercamiento consecuencialista, en tanto considera que los estándares normativos adecuados surgen de la eficacia en que los sujetos cognoscentes razonan o actúan para alcanzar sus metas (Evans y Over, 1996, p. 7). De manera implícita, Evans y Over suponen un compromiso de la racionalidad 1 con un acercamiento consecuencialista de la racionalidad cuando afirman:

> Se podría ser racional en términos de alcanzar metas personales (racionalidad 1) sin ser racional en el sentido de estar en conformidad con un sistema normativo tal como la lógica (Evans y Over, 1996, p. X, la traducción es mía).

En el capítulo anterior distinguí tres tipos de consecuencialismo, a saber, veritista, pragmatista y biológico. Dada esta distinción, el consecuencialismo que la racionalidad 1 o racionalidad personal presupone es de tipo pragmatista, puesto que considera que un proceso de razonamiento es adecuado si conduce a la satisfacción de deseos y metas personales.

El tipo de críticas a los principios normativos de decisión hechos por Evans y Over no sólo suponen un rechazo a la visión estándar de la racionalidad, sino que al mismo tiempo apoyan una noción de racionalidad dependiente de recursos, la cual, como se explicó en el capítulo anterior, se encuentra relacionada íntimamente con el acercamiento consecuencialista de la racionalidad. Evans y Over mismos reconocen que su rechazo a que la lógica y los estándares propuestos por una teoría formal de la decisión den cuenta de la racionalidad 1 fue influenciada por Herbert Simon. Muchas veces los seres humanos tienen a lo mucho vagas preferencias basadas en un conocimiento parcial de lo que su lenguaje hace referencia, por lo que hacen juicios de probabilidad vagos y limitados. Por tal motivo, sería equivocado, creen Evans y Over, que la teoría normativa de la decisión sea el estándar general de la racionalidad 1, pues aceptarlo sería violar el principio de que el "deber" implica el "poder" (Evans y Over, 1996, p. 31), tal y como sugieren los defensores de una visión de la racionalidad depen-

diente de recursos. Así, la racionalidad 1 es una teoría normativa que es usada para evaluar el razonamiento. Dada la estructura de la mente, los resultados de los procesos implícitos o tácitos de razonamiento propios de S1 generalmente satisfacen este estándar normativo. Según he argumentado, dicho estándar se compromete con un acercamiento consecuencialista pragmatista y dependiente de recursos.

La racionalidad 2 es aquella que, según Evans y Over, evalúa el grado con el cual los sujetos llevan a cabo inferencias o actúan por razones que son evaluadas por un estándar normativo abstracto como lo son la lógica formal y la teoría de la probabilidad. Es común observar en *Rationality and Reasoning* afirmaciones como "la teoría de la decisión bayesiana provee el marco para entender la racionalidad 2" (Evans y Over, 1996, p. 43, la traducción es mía). La racionalidad 2, a diferencia de la racionalidad 1, es una teoría normativa que claramente se compromete con la visión estándar de la racionalidad. Las referencias que apoyan esta tesis enfatizan el contraste entre el origen de la normatividad propia de la racionalidad 1 de aquel proveniente de la racionalidad 2. Mientras los criterios de la lógica y las teorías formales sirven de base para evaluar la racionalidad impersonal o racionalidad 2, estos mismos criterios no pueden ser usados para explicar la racionalidad personal o la racionalidad 1. Haciendo la distinción entre la racionalidad personal e impersonal Evans y Over afirman:

> El acercamiento impersonal se pregunta si estamos siguiendo los principios lógicos y de otras teorías normativas en nuestro razonamiento y toma de decisiones. Flanagan ofrece una buena declaratoria de una visión impersonal de la racionalidad: "Muchas veces la racionalidad es tomada como un equivalente de logicidad. Esto es, eres racional sólo en caso de que sistemáticamente instancies las reglas y principios de la lógica inductiva, la estadística y la teoría de la probabilidad, por una parte, y la lógica deductiva y todas las ciencias matemáticas, por otra parte" (Evans y Over, 1996, p. 7, la traducción es mía).

Nótese cómo es que Evans y Over encuentran en la descripción de Flanagan una buena manera de entender la racionalidad impersonal o la racionalidad 2. Esta descripción es precisamente el

punto que es defendido desde la visión estándar de la racionalidad, pues en ella ser racional consiste en satisfacer las normas provenientes de la lógica y la matemática, especialmente de la teoría de la probabilidad. Con ello parece haber una clara similitud entre la visión estándar de la racionalidad y lo que Evans y Over entienden por racionalidad 2. En el capítulo anterior argumenté que la visión estándar de la racionalidad se encuentra enmarcada en un acercamiento normativo deontológico, puesto que, según este acercamiento a la racionalidad, lo que es constitutivo del buen razonamiento es únicamente la satisfacción de los principios normativos adecuados. Por lo que, si lo dicho hasta aquí es correcto, la aceptación de la racionalidad 2 supone un compromiso con el acercamiento deontológico de la racionalidad, que puede ser contrastado con la racionalidad 1 y su compromiso consecuencialista.

Si lo que he argumentado es correcto, para la teoría dual defendida por Evans y Over (1996), la postulación de los sistemas S1 y S2 representa su teoría descriptiva del razonamiento. Del mismo modo, la distinción entre racionalidad 1 y racionalidad 2 representa para Evans y Over su teoría normativa; dicha teoría normativa supone la aceptación de un acercamiento consecuencialista y deontológico de la racionalidad. Lo interesante dentro de la teoría dual ofrecida por Evans y Over es que en la evaluación del razonamiento se pueden combinar distintas teorías normativas de la racionalidad, esto es, ni el razonamiento humano se puede reducir a un único sistema de pensamiento ni la racionalidad a una única teoría normativa. La correcta manera de entender el razonamiento, según estos psicólogos, supone una dualidad en los procesos de pensamiento. Del mismo modo, la manera correcta de entender a la racionalidad, según Evans y Over, implica una dualidad de criterios normativos.

V.4.2. RACIONALIDAD EVOLUCIONISTA Y RACIONALIDAD NORMATIVA

Stanovich y West, como mencioné anteriormente, postulan dos teorías de la racionalidad, a saber, la racionalidad evolucionista y la racionalidad normativa. La racionalidad evolucionista evalúa los procedimientos de optimización a nivel subpersonal o de los genes. La racionalidad normativa, por su parte, evalúa los procedimientos de optimización a nivel personal o de los individuos.

Los dos tipos de racionalidad hacen referencia, según Stanovich y West, a dos niveles de optimización distintos, en otras palabras, hacen referencia a la eficiencia con que los distintos procesos mentales alcanzan diferentes tipos de metas. Si bien estos psicólogos no han desarrollado con cuidado las diferencias entre estos dos tipos de racionalidad, la manera en cómo lo presentan sugiere fuertemente que el criterio para determinar si un proceso o mecanismo —ya sea perteneciente a S1 o S2— es adecuado, es evaluando si dicho proceso o mecanismo conduce de manera eficiente a una meta o conjunto de metas particulares. El sistema S1, en el cual se encuentran procesos rápidos, modulares, inconscientes, etc., generalmente satisface una racionalidad en donde el nivel de optimización se encuentra a nivel de los genes. En cambio, los procesos que pertenecen al sistema S2 son eficientes para alcanzar las metas personales, esto es, generalmente satisfacen los estándares de la racionalidad normativa.

Lo interesante de la perspectiva de Stanovich y West es que la explicación que ofrecen para hacer la distinción entre estos dos tipos de racionalidad se enmarca fácilmente dentro del acercamiento consecuencialista revisado en el capítulo anterior. El consecuencialismo es la teoría normativa de la racionalidad que sostiene que el buen razonamiento consiste en razonar de manera tal que probablemente se alcancen ciertas metas o resultados. Ahora bien, estas metas o resultados pueden ser de distintos tipos o a distintos niveles. Lo que Stanovich y West hacen al distinguir entre la racionalidad evolucionista y la racionalidad normativa es simplemente hacer una distinción entre la optimización de metas a diferentes niveles, i. e., un nivel subpersonal y un nivel personal —o para hacer uso de la terminología de Dawkins que muchas veces estos psicólogos ocupan, una optimización a nivel de los genes o a nivel de los vehículos de genes o los individuos.

Si mi apreciación es correcta, los dos tipos de racionalidad que distinguen Stanovich y West (2000, 2003) suscriben un acercamiento consecuencialista de la normatividad. En el capítulo anterior distinguí tres tipos de consecuencialismo, a saber, veritista, biológico y pragmatista. Las dos nociones de consecuencialismo que Stanovich y West suscriben son el consecuencialismo biológico y pragmatista. El consecuencialismo pragmatista, según lo visto en el capítulo anterior, sostiene que un proceso de razona-

miento es adecuado si con él se pueden alcanzar los objetivos pragmáticos de satisfacer las metas y deseos personales. El consecuencialismo biológico, por su parte, evalúa el razonamiento a partir de qué tanto conduce a metas como la supervivencia, la reproducción o la propagación de los genes. De este modo, lo que Stanovich y West denominan "racionalidad evolucionista" pertenece al consecuencialismo biológico, mientras que lo que denominan "racionalidad normativa" pertenece al consecuencialismo pragmatista.

Así, bajo una perspectiva meramente descriptiva, el razonamiento humano se puede entender, según Stanovich y West, a partir de mecanismos que subyacen en los sistemas S1 y S2. El primero de estos sistemas generalmente satisface los estándares provenientes de la racionalidad evolucionista, mientras que el segundo sistema generalmente satisface los estándares provenientes de la racionalidad, normativa [32]. Sin embargo, estos dos tipos de racionalidad se comprometen con un acercamiento consecuencialista de la racionalidad en donde el origen de la normatividad proviene de la eficacia en que cada uno de estos sistemas alcance las metas o resultados particulares. Lo que hace a un mecanismo un *buen* mecanismo de razonamiento no es satisfacer (o seguir) las normas provenientes la lógica, la probabilidad y las matemáticas, sino el grado de eficacia con el cual dicho mecanismo cumple con ciertas metas. Así como algunas veces ciertos mecanismos pueden ser eficaces, por ejemplo, para alcanzar cierta meta a nivel de los genes, pero dichos mecanismo podrían ser poco satisfactorios según la racionalidad normativa; algunas veces ciertos mecanismos podrían ser eficaces para alcanzar metas a nivel de los individuos, que sin embargo sean poco racionales según la racionalidad evolucionista. De este modo, la teoría dual de sistemas defendida por Stanovich y West asume una teoría dual de la racionalidad, no obstante, esta teoría dual se compromete únicamente con el acercamiento consecuencialista de la racionalidad en sus versiones pragmatista y biológico.

V.5. CONCLUSIONES

A lo largo de este capítulo se ha podido revisar con cuidado la teoría dual de sistemas. En general, se puede sostener que dentro

de las distintas versiones de la teoría dual exploradas aquí existe un núcleo de principios que las diferentes versiones aceptan. En términos generales se postula un sistema (S1) de razonamiento tácito, inconsciente, rápido, antiguo en términos evolutivos, que el ser humano comparte con otros animales, que puede claramente distinguirse de un sistema (S2) de razonamiento explícito, consciente, el cual puede ser controlado, lento, de capacidad limitada y distintivamente humano. *Grosso modo,* el razonamiento se entiende como la interacción entre estos dos sistemas, que si bien pueden ofrecer respuestas divergentes, suelen trabajar en conjunto. Estos son los principios centrales que comparten las distintas versiones de la teoría dual revisadas hasta ahora.

La principal línea de distanciamiento entre las teorías duales es el compromiso que estas teorías tienen con la racionalidad, principalmente la teoría de Evans y Over, y la teoría de Stanovich y West. Estas dos teorías duales sostienen que la distinción entre dos sistemas de razonamiento conduce a una distinción entre dos teorías de la racionalidad. Para hacer uso de sus propios términos, el sistema tácito y el sistema explícito postulados por Evans y Over conducen, según estos psicólogos, respectivamente, a la racionalidad personal (o racionalidad 1) y a la racionalidad impersonal (o racionalidad 2). Por su parte, Stanovich y West afirman que los sistemas S1 y S2 conducen a la racionalidad evolucionista y la racionalidad normativa. A mi juicio, los defensores de la teoría dual no ofrecen un argumento claro que muestre que de la existencia de dos sistemas de razonamiento se siga lógicamente que se deban aceptar dos criterios normativos distintos. En parte, la postulación de tales teorías normativas está motivada por dar una respuesta a la denominada "paradoja de la racionalidad", pero este no es un argumento fuerte para aceptar la postulación particular que los psicólogos defensores de la teoría dual hacen. En el próximo capítulo ofreceré dos argumentos para aceptar al menos dos teorías normativas para evaluar el razonamiento. De momento deseo enfatizar la idea de que no existe consenso con relación a las teorías normativas que se deben aceptar, si concebimos que la cognición humana consta de dos sistemas de razonamiento.

En el capítulo he argumentado, además, que la teoría dual de Evans y Over presupone una teoría dual de la racionalidad y la

teoría de Stanovich y West se compromete con dos teorías normativas de tipo consecuencialista. Por un lado, la teoría dual de Evans y Over asume dos acercamientos distintos de la racionalidad, a saber, la racionalidad personal como un tipo de racionalidad enmarcada en el acercamiento consecuencialista de corte pragmatista, mientras que la racionalidad impersonal se encuentra enmarcada en lo que se denomina la visión estándar de la racionalidad (y con ello con un acercamiento deontológico). Por otro lado, la teoría dual de Stanovich y West sólo se compromete con un acercamiento a la racionalidad, esto es, tanto la racionalidad evolucionista como la racionalidad normativa son tipos de racionalidad comprometidos con el acercamiento consecuencialista. Sin embargo, mientras que la racionalidad evolucionista se puede entender como un consecuencialismo de tipo biológico, la racionalidad normativa se pude entender como un consecuencialismo de tipo pragmatista. La idea central que está detrás de las propuestas normativas de Evans y Over y de Stanovich y West es que el razonamiento humano debe ser evaluado por al menos dos estándares normativos distintos, i. e., el razonamiento humano no puede ser evaluado con un único estándar normativo. Estas conclusiones son de suma importancia para este trabajo ya que serán usadas en el próximo capítulo para indagar las relaciones teóricas que hay entre las nociones de racionalidad y justificación desde la teoría dual de sistemas.

Se enumeraron aquí algunos de los problemas con los que la actual teoría dual de sistemas se enfrenta. Según lo revisado en dicha sección, de momento no se tiene una postura completamente coherente que pueda distinguir claramente dos sistemas de razonamiento con el conjunto de rasgos que generalmente se han descritos de cada uno de los sistemas. Uno de los problemas centrales radica en la idea misma de un sistema tácito o S1, pues en este sistema se encuentran distintos mecanismo que no pueden ser parte del sistema explícito o S2, pero que son lo suficientemente distintos para agruparlos en un "sistema". Las dificultades revisadas, si bien importantes, indican precisamente parte de las posibles líneas de investigación dentro de la teoría dual de sistemas.

Un último punto que deseo señalar en estas conclusiones es la respuesta que los defensores de la teoría dual de sistemas —en

especial en la versión defendida por Evans y Over— dan a la interpretación pesimista de los resultados provenientes de la tradición de heurística y sesgo. Si bien en el segundo capítulo expuse dos versiones de esta interpretación, la psicología evolucionista puede mostrar que en realidad la versión más fuerte (o la versión II) no es adecuada. Asimismo, anteriormente he argumentado que una versión optimista de la racionalidad humana ofrecida por la psicología evolucionista (la versión optimista A) y una versión de la interpretación pesimista defendida por la tradición de heurística y sesgo (la versión I) pueden ser compatibles. Es precisamente frente a esta intersección donde se puede ubicar a la teoría dual de sistemas. La respuesta de la teoría dual a estas interpretaciones depende, a grandes rasgos, de la importancia de los procesos de pertinencia. En diversos experimentos llevados a cabo por los defensores de la teoría dual se intenta mostrar cómo los procesos que pertenecen al sistema tácito o S1 intervienen en la resolución de problemas de razonamiento. Se señala después el sentido en que estos procesos de razonamiento son racionales. Los defensores de la teoría dual de sistemas no niegan la existencia de sesgos de razonamiento, pero de ello no se sigue, según estos psicólogos, que los seres humanos sean invariablemente irracionales, pues los sesgos son errores sólo con respecto a uno de los estándares normativos con los que se puede evaluar el razonamiento humano.

Al revisar la teoría dual de sistemas y mostrar cómo esta teoría puede explicar tanto la evidencia optimista como pesimista proveniente de dos corrientes psicológicas —en principio— antagónicas, y ver cómo la teoría dual de sistemas postula un intrincado conjunto de procesos cognitivos divididos en dos grandes sistemas, considero que la teoría dual es de momento la alternativa más adecuada para explicar el razonamiento humano. Esta convicción es reforzada por los argumentos en contra de la visión estándar de la racionalidad presentados en capítulo anterior. Dadas las críticas expuestas en dicho capítulo, la visión estándar de la racionalidad no da cuenta de manera cabal de las normas con las cuales debe razonar el ser humano. Por tal motivo, es importante que una teoría del razonamiento tenga en cuenta estándares normativos distintos de los postulados por la visión estándar. Del mismo modo, una teoría del razonamiento debe

tener en consideración la idea de una racionalidad dependiente de recursos (cognitivos y medio ambientales). Según lo revisado en este capítulo, la teoría dual de sistemas es una teoría que toma en cuenta estos elementos. A partir de esto, en el próximo capítulo indagaré si esta teoría del razonamiento apoya a una noción particular de justificación.

VI.
LA TEORÍA DUAL DE SISTEMAS Y LAS NOCIONES DE JUSTIFICACIÓN Y DE RACIONALIDAD

En el primer capítulo de este trabajo sostuve que detrás del debate actual entre el internismo y el externismo están distintas intuiciones epistémicas de lo que significa estar justificado. Estas intuiciones tienen en común que la noción de justificación se relaciona con la noción de racionalidad y el razonamiento, por ello he estudiado distintas teorías del razonamiento a lo largo de este trabajo. Si lo revisado en el capítulo anterior es correcto y la teoría dual de sistemas es la teoría psicológica que, de momento, brinda la mejor explicación sobre el razonamiento humano, entonces es necesario investigar si esta teoría apoya a una noción de justificación particular. Mostraré que la teoría dual de sistemas, a diferencia de otras teorías del razonamiento, sí sugiere una dirección en el debate entre las teorías internistas y externistas [1]. En particular, argumentaré que los productos de los dos sistemas de razonamiento postulados por los defensores de la teoría dual están justificados según estándares externistas. Además, sostendré que si una noción de justificación se funda sobre la teoría dual de sistemas, en su nivel descriptivo y normativo, entonces esta noción tendrá que ser externista no veritista, y tendrá que aceptar un tipo de dualidad a nivel normativo.

La estructura de este capítulo es la siguiente. En primer lugar, indagaré cuál es la relación entre la teoría dual de sistemas, a un nivel descriptivo, y las nociones internistas y externistas de la justificación. Distinguiré tres maneras de vincular, dicha teoría psicológica con las nociones de justificación, pero argumentaré que una de ellas (la idea de que la teoría dual apoya una postura externista de la justificación) es la más adecuada. En segundo

lugar, analizaré la relación conceptual entre la noción de justificación y la noción de racionalidad desde la perspectiva de la teoría dual. Como he expuesto, los defensores de la teoría dual comparten la idea de que debe haber al menos dos teorías normativas para evaluar el razonamiento humano, de modo que analizaré cómo esta idea puede ser compatible con una noción de la justificación. Al final de este capítulo esbozaré un ejemplo de teoría externista formulada a partir de las intuiciones normativas provenientes de la teoría dual.

VI. 1. LA NOCIÓN DE JUSTIFICACIÓN Y LA TEORÍA DUAL DE SISTEMAS

En el primer capítulo de este trabajo sostengo que existen dos ejes en torno a las distintas teorías de la justificación. El primer eje fundacionismo-coherentismo se entiende como un cuestionamiento acerca de la naturaleza de las creencias justificadas y la dirección en que las creencias "trasmiten" la justificación. El segundo eje de discusión se centra no tanto en la naturaleza de las creencias justificadas, sino en la relación entre el sujeto cognoscente y las condiciones que justifican a sus creencias. Para el internismo la justificación de las creencias de *S* depende de estados internos (reflexión, razonamiento o memoria) a los cuales *S* tiene acceso inmediato. Mientras que para el externismo la justificación de las creencias de *S* depende del estado externo de cosas que hacen que su creencia sea causada por un proceso cognitivo adecuado [2], en este sentido, *S* no necesariamente tienen acceso a dichos estados. Para un internista, siempre que *S* está justificado en creer que *p*, *S* cree justificadamente que lo está; por el contrario, para un externista *S* puede estar justificado en creer que *p*, aun cuando no crea que está justificado en creer que *p* (King, 2000, Steup, 2005). Ya que para el externismo un sujeto puede estar justificado sin creer justificadamente que lo está, la principal crítica hecha a esta noción de justificación es que no recoge lo que los seres humanos afirman tener cuando hablan de creencias justificadas, esto es, razones para decir que se está justificado (Bonjour, 1980). Por su parte, ya que el internista considera que siempre que un sujeto está justificado éste cree justificadamente que lo está, uno de los argumentos en contra del internismo es que su noción de

justificación es demasiado estrecha, pues calificaría como injustificadas muchas de las creencias que los sujetos cognoscentes consideran justificadas (Goldman, 1980).

Los defensores del internismo consideran que la noción de justificación debe recoger la intuición de la responsabilidad epistémica y la *racionalidad*. Por su parte, los defensores del externismo consideran que la genuina intuición detrás de la noción de justificación es la referencia a los procesos cognitivos y métodos adecuados de formación de creencias, como los distintos procesos de razonamiento inductivo y deductivo. Así, el internismo y el externismo se presentan como dos modos distintos de entender la noción de justificación, que a su vez parecen responder a intuiciones epistémicas distintas. Si bien las dos intuiciones que consideran los defensores del internismo y externismo son distintas, las intuiciones que defienden estas dos posturas comparten una referencia, de manera directa o indirecta, a la noción de razonamiento y racionalidad. Es por ello que en el primer capítulo sugiero que el estudio de la razón puede echar luz sobre el debate contemporáneo en torno a la noción de justificación epistémica.

En esta sección del capítulo deseo mostrar que hay tres posibles maneras de acercarse al debate en teoría de la justificación desde la teoría dual de sistemas. Muy a grandes rasgos, una manera de entender la conexión entre la teoría dual de sistemas y la noción de la justificación es argumentando que la noción internista de la justificación responde a un sistema de razonamiento, mientras que la noción externista de la justificación responde a otro sistema de razonamiento (Eraña, en proceso b). En segundo lugar, exploro la propuesta de que la teoría dual de sistemas sólo ofrece creencias que están justificadas según el criterio internista, por lo que se debe hacer una distinción entre creencias justificadas y creencias racionales (Eraña, en proceso a). En tercer lugar analizaré la idea de que la noción externista de la justificación puede dar cuenta de los dos sistemas de razonamiento. Esta última manera de entender la relación entre la teoría dual y un tipo de externismo es la relación que considero más adecuada. Veamos.

VI. 1.1. LA TEORÍA DUAL DE SISTEMAS Y UNA NOCIÓN NO EXCLUYENTE DE JUSTIFICACIÓN

Una primera manera de relacionar la teoría dual de sistemas con las nociones de justificación es sostener que cada uno de los sistemas de razonamiento ofrece creencias justificadas según distintas nociones de justificación. Una propuesta como esta es defendida por Eraña (en proceso b) en el artículo "La noción de justificación, ¿un concepto dual?". Según Eraña, dadas las características de S1 es posible conectar a este sistema con el criterio de justificación defendido por el externismo, ya que

> los procesos que subyacen a la operación de S1: (i) no son directamente accesibles a la conciencia; (ii) son automáticos, y (iii) son innatos, pero están moldeados por nuestra interacción con el medio ambiente (Eraña, en proceso b, p. 19 [3]).

La primer característica de los procesos pertenecientes a S1 tiene estrecha relación con la idea externista de que los sujetos no tienen acceso a las condiciones que justifican a sus creencias. Eraña (en proceso b) cree que estos productos o *outputs* no pueden estar justificados, según la noción internista, en tanto que (i) pone de manifiesto que los sujetos no pueden tener acceso a tales procesos. Esta postura choca con la idea internista de que un sujeto que está justificado, cree que lo está. La segunda característica mencionada hace patente que los procesos de S1 ofrecen respuestas automáticas, esto es, independientemente del sujeto, los procesos que conforman su sistema tácito producen creencias racionales que no están bajo su control consciente. Eraña (en proceso b) argumenta además que el rasgo (iii) de los procesos que subyacen a S1, que muestra que hay una relación entre los resultados que S1 entrega y los criterios externistas de la justificación, apunta al hecho de que los seres humanos son muy eficientes para resolver los problemas en la vida cotidiana [4]. Eraña afirma

> los procesos que subyacen a S1 toman como insumo hechos del mundo que tiene una conexión causal con su operación y es esta conexión lo que hace que ellos sean altamente confiables, i.e., tienden a proveer respuestas eficientes a problemas regularmente encontrados (Eraña, en proceso b, p. 20).

Si entiendo adecuadamente a Eraña, ella considera que el hecho de que los procesos de S1 sean innatos (pero moldeados por el medio ambiente) permite que tales procesos generen una operación causal confiable. De ahí que, según Eraña (en proceso b), es posible sostener que los productos ofrecidos por S1 tienen afinidad con la idea externista (de corte confiabilista) de que una creencia está justificada si es producto de un proceso cognitivo confiable.

De (i), (ii) y (iii) Eraña sostiene que se puede "mostrar que hay una relación entre los resultados que S1 entrega y los criterios externistas de la justificación epistémica" (en proceso b, p. 19 [5]). Se podría preguntar ahora si S2 tiene el mismo tipo de relación con el internismo que al parecer tiene S1 con el criterio de justificación externista. Eraña responde afirmativamente al cuestionamiento anterior, y las razones que ofrece se fundan en las características de S2 como un sistema

> (a) guiado por ciertos principios que están explícitamente representados en él; (b) entrega resultados que son sistemáticamente correctos de acuerdo con los principios normativos de razonamiento; y (c) recibe como insumos hechos del mundo (igual que S1) y creencias producidas por los procesos subyacentes a S1 [6] (Eraña, en proceso b, p. 25).

Para Eraña (en proceso b), la descripción de S2 presupone una noción internista de justificación, porque los procesos subyacentes a S2 pueden servir como guías para decidir qué creer y porque permiten aceptar sólo las creencias que el sujeto cognoscente puede creer [7]. Además de lo anterior, las razones que los sujetos cognoscentes tienen para aceptar las creencias producidas por S2 son accesibles de manera especial, en el sentido en el que el contenido de S1 no es accesible (Eraña, en proceso b). El internismo sostiene que una creencia está justificada si el sujeto puede mostrar que su creencia *p* está relacionada de modo adecuado con otras creencias o estados mentales [8]. De lo anterior, es verosímil sostener que las creencias producidas por S2 están justificadas en virtud de su relación con otras creencias o estados mentales producto de S1 o S2 (Eraña, en proceso b). Bajo esta perspectiva, se puede sostener que los *outputs* provenientes de S2 estarían justificados bajo criterios internistas (Eraña, en proceso b).

Una manera de entender estas ideas es que las creencias producto de S1 son creencias justificadas por el hecho de que los procesos que operan en S1 son procesos racionales y eficaces para alcanzar el tipo de metas que los individuos tienen. Por su parte, las creencias que son producto de S2 están justificadas debido a que son resultado de procesos conscientes que permiten a los sujetos cognoscentes revisar previamente las creencias sostenidas y asegurarse que las creencias son consistentes entre ellas. Esta noción dual de la justificación parece preferible, según Eraña (en proceso a) que defender de manera exclusiva, o el internismo, o el externismo, para cualquier tipo de creencia, ya que en principio la noción no exclusiva de la justificación parecería lógicamente compatible con una teoría descriptiva del razonamiento.

En suma, una posible conexión entre la noción de justificación y la teoría dual de razonamiento traza, por un lado, una línea que une a la noción internista de la justificación con S2 y, por otro lado, una línea que une a la noción externista de la justificación con S1 (Eraña, en proceso b). Si la teoría que mejor da cuenta del razonamiento humano es aquella que postula la existencia de los sistemas S1 y S2, entonces, según esta primera propuesta, quizá se debe concebir una noción de justificación no excluyente que incluya al internismo y al externismo. Lo anterior puede ser resumido esquemáticamente en el siguiente cuadro.

Teorías de la justificación	Internismo	Externismo
⇕	⇕	⇕
Teoría dual del razonamiento	S2	S1

Si lo argumentado en esta sección fuese correcto, entonces la noción de justificación sería un concepto dual como defiende Eraña (en proceso b). Pero, ¿en realidad es dual la noción de justificación? Una de las razones para sostener la propuesta de que S1 ofrece creencias justificadas, según criterios externistas, mientras que S2 ofrece creencias justificadas, según criterios internistas, es poniendo énfasis en el acceso que los sujetos tienen a los procesos que causan sus creencias. En gran medida el debate entre

el internismo y el externismo se ha enfocado en si el sujeto debe acceder [9] o no a los factores y condiciones que determinan la justificación de una creencia (Steup, 2005). Esta característica parece en cierto modo encontrarse en la teoría dual del razonamiento, en especial en la versión de Evans y Over (1996) al enfatizar a S1 como un sistema implícito y a S2 como uno explícito. Sin embargo, esta distinción parece desdibujarse por varias razones:

a) No todos los defensores de la teoría dual han visto en la distinción implícito-explícito (inconsciente-consciente) el rasgo distintivo entre S1 y S2. Por ejemplo, Sloman (1996) sostiene que la conciencia que el sujeto puede tener sobre los procesos y/o los *outputs* originados por dichos procesos no determina la diferencia entre los dos sistemas. Como expuse en las críticas a la teoría dual del razonamiento, no existen criterios completamente claros para diferenciar a S1 de S2 (Evans, 2006). Si lo anterior es el caso y la distinción central entre los dos sistemas de razonamiento no es el acceso que se puede tener de los procesos que subyacen a éstos, entonces la relación entre S1 y el externismo, y entre S2 y el internismo parece desdibujarse.

b) Relacionado con lo anterior, es preciso recordar que una de las críticas a la teoría dual de sistemas es el tipo de acceso (consciente o no) y del control que se puede tener de los sistemas de razonamiento (Evans, 2006). Esto es, aun cuando la conciencia pueda ser un rasgo que determine cuándo un mecanismo pertenece a S1 o S2, se dice que hay mecanismos de los que el sujeto no es consciente pero que pueden pertenecer a S2 o que puede controlar hasta cierto punto y que subyacen a S1. El hecho de que la noción de acceso o control que se tiene sobre los procesos puede subdeterminar su inclusión en un sistema de razonamiento, va en contra de la idea de que los productos de S2 ofrecen creencias que están justificadas según el criterio internista de justificación, puesto que una de las razones para sostener una propuesta como la de Eraña (en proceso b) es que S2 es guiado por principios explícitamente representados a los cuales el sujeto tiene control consciente. Sin embargo, si el control o acceso que se tiene a los procesos no es lo que determina cuándo uno de éstos es parte de S1 o S2, entonces el control consciente que puede tener un sujeto sobre sus creencias tal y como lo defiende el internista no puede responder siempre a los productos de S2.

c) El hecho de que haya mecanismos que fueron conscientes y que se automatizaron de modo tal que el sujeto no tiene acceso a los procesos sino a los productos de los mismos, no tiene implicación alguna con la idea de que si dichos procesos son *eficientes*, entonces pueden ofrecer creencias justificadas desde la perspectiva externista. Sin embargo, si se sostiene que los productos de S2 están justificados por los criterios internistas, en tanto que se tiene control consciente de las relaciones entre sus creencias, entonces las creencias que en un momento estarían justificadas se convertirían en creencias injustificadas después de que el proceso se automatizara. En otras palabras, supóngase que *P* es un proceso consciente perteneciente a S2 y el control de un sujeto sobre la relación de las creencias causadas por *P* junto con otras creencias es lo que hiciese que tales creencias estuviesen justificadas. Si el proceso *P* se automatizara, entonces el sujeto perdería el control consciente de *P*, además *P* se volvería inaccesible, por lo que los *outputs* de *P* dejarían de estar justificados desde la perspectiva internista. Así, la idea de que los productos de S2 están justificados, según el criterio internista, tendría como consecuencia la idea contraintuitiva de que algo que está justificado deje de estarlo debido a los procesos automatizados.

Se podría sostener que no es contraintuitivo pensar que los sujetos pueden estar justificados en creer *p* en un tiempo t1 y dejar de estarlo en un tiempo t2. Esto se podría deber a dos razones: *i.)* que nueva evidencia en t2 haga que el sujeto considere que su creencia *p* no esté justificada, por ejemplo al creer que *q implica no-p* y creer también *q*; o *ii.)* que el sujeto ha olvidado las razones que hacían que su creencia esté justificada, por ejemplo, cuando las razones para creer *p* dependía de la creencia que *(p o r)* y *no-r* y ha olvidado tales creencias. La objeción que aquí hago a la idea de que los *outputs* de los procesos pertenecientes a S2 no están justificados, porque hacen que una creencia justificada deje de estarlo, no se fundan en las razones *i.)* y *ii.)*. La crítica que hago a la idea de que S2 no produce creencias justificadas, según el criterio internista, consiste en que aun cuando no haya evidencia para dejar de creer que *p*, y que el sujeto no haya olvidado las razones para creer que *p*, *p* podría dejar de estar justificada simplemente por la automatización de un proceso perteneciente a S2.

Lo que estas críticas señalan es la oscuridad de la noción de control consciente y conciencia para distinguir entre dos sistemas de razonamiento distintos, así como para determinar si un mecanismo cognitivo particular pertenece a S1 o S2. Estos señalamientos socavan la conexión entre la teoría dual del razonamiento y dos nociones de justificación (una internista y otra externista). En particular, el hecho de que no sea claro que lo que distinga a S2 de S1 sea el acceso que uno tiene al sistema, o el hecho de que haya procesos conscientes en S1 e inconscientes en S2 y la existencia de procesos automatizados, socavan varias intuiciones internistas, por ejemplo, que cuando un sujeto está justificado puede dar y ofrecer razones, y la idea de que cuando un sujeto está justificado, éste cree justificadamente que lo está. De lo anterior, considero que la propuesta de que los productos de distintos sistemas de razonamiento están justificados, según distintos criterios de razonamiento (internista y externista), debe ser rechazada. A simple vista, la propuesta de una noción dual de justificación parece adecuada, pero una exploración más detallada de la misma hace patente que la relación entre los dos sistemas de razonamiento y las dos nociones de justificación es conceptualmente débil.

Un argumento que se podría ofrecer a favor de una propuesta como la de Eraña (en proceso b) es que la noción internista de la justificación no se preocupa por el origen causal de las creencias. Como expuse en el primer capítulo, Bonjour (1985) y Davidson (1983) hacen uso de la vieja distinción entre el modo en que las creencias son causadas del modo en que son justificadas. Así, según estos filósofos, podría ser que los procesos que causen las creencias de un sujeto sean tales que no pueda tener acceso a dichos procesos, pero sería suficiente para posibilitar la justificación internista que el sujeto tenga acceso a los *outputs* de S1 y S2, y a las relaciones entre éstos. El problema con un argumento que se apoye en esta idea es que para los internistas no basta que los sujetos cognoscentes tengan acceso a sus creencias, sino que sean conscientes de que las creencias están relacionadas de una manera particular con el sistema de creencias y que crean (justificadamente) que esa manera particular es la más adecuada. Una forma de entender lo anterior es expuesta en el capítulo I; para Bonjour (1985) la justificación no sólo depende de que las creencias estén relacionadas coherentemente en un sistemas de creencias, sino

que el sujeto debe creer (justificadamente) que la coherencia es una propiedad que conduce a la verdad [10]. Son estas características las que integran la intuición de que para que los sujetos cognoscentes estén justificados, éstos deben ser epistémicamente responsables y racionales [11]. Las críticas *a), b)* y *c)* expuestas más arriba precisamente señalan que estas intuiciones internistas no pueden encontrarse en los productos de S2, es decir, S2 no produce creencias justificadas, según el criterio internista, porque *a), b)* y *c)* muestran que hay diferentes sentidos en que el sujeto cognoscente no pueden tener control consciente sobre distintos procesos que subyacen a S2. Sin tal control consciente, el sujeto no puede ser epistémicamente responsable, lo que contradice la idea de que S2 ofrece creencias justificadas de manera internista.

VI.1.2. LA TEORÍA DUAL DE SISTEMAS Y EL INTERNISMO

Antes de defender que los procesos postulados por la teoría dual de sistemas ofrecen resultados que satisfacen distintos criterios de justificación, Eraña (en proceso a) sostiene que en realidad lo único que la teoría dual de sistemas muestra es que la noción de racionalidad y justificación no son equivalentes. Con mayor precisión, lo que ella defiende en el artículo "Dual system theory. A way to bridge the gap between Internalism and Externalism" es que si bien todas las creencias justificadas son creencias racionales, no todas las creencias racionales están justificadas. Veamos.

Eraña (en proceso a) sostiene que los procesos de S2, es decir, procesos que son conscientes, que siguen reglas que pueden ser usados por el sujeto para decidir qué creer, etc., ofrecen productos que están justificados —por las razones expuestas en la sección anterior— bajo los criterios internistas de justificación. Si bien, según ella, los productos de S1 parecían estar justificados, según la propuesta externista, el análisis filosófico de la noción de justificación muestra que el concepto de justificación siempre implica un tipo de responsabilidad epistémica. Es importante señalar que la idea de que la noción de justificación supone un tipo de responsabilidad epistémica no está apoyada por la teoría dual de sistemas, sino que Eraña (en proceso a) lo hace de manera independiente y por consideraciones meramente analíticas.

> Asumamos, además, que nuestros juicios intuitivos acerca de la justificación requieren que el sujeto de la creencia no sólo explique su creencia, sino también que ofrezca razones para sostenerla (Eraña, en proceso a, p. 14, la traducción es mía).

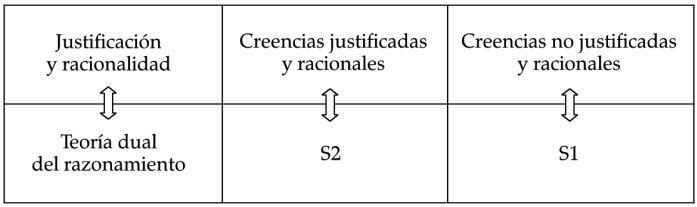
Eraña está partiendo, así, de cierta noción intuitiva de justificación, en donde lo importante es que la noción recoja lo que intuitivamente se espera de un sujeto que está justificado. Sin embargo, las intuiciones que Eraña (en proceso a) recoge son las que usualmente defienden los internistas. Entre las razones que ella brinda para defender esta idea es que ofrecer razones parece requerir un tipo de reflexión, en donde el sujeto cognoscente de alguna manera debe acceder a sus creencias y verlas como "razones", o debe ser capaz de establecer explícitamente una liga entre sus creencias y el proceso causal que las produce (Eraña, en proceso a). Según Eraña, cualquiera de estas posibilidades requieren el punto de vista reflexivo del sujeto, que le permite reconocer la probabilidad de que una creencia justificada sea verdadera [12].

> Para que una creencia esté justificada del todo... es necesario que el sujeto de la creencia *S* provea una defensa de la misma, y esto no significa otra cosa sino que *S* tome un punto de vista reflexivo al considerar las razones que apoyan la creencia que está siendo evaluada. Así, de esto tenemos, que la única noción de justificación epistémica disponible —i.e., la única que satisface nuestros juicios intuitivos acerca de la justificación— implica la reflexión (Eraña, en proceso a, p. 15, la traducción es mía).

Si se aceptan los argumentos de Eraña (en proceso a) fundados en intuiciones internistas, entonces se tiene que aceptar que la reflexión y la responsabilidad epistémica son condiciones necesarias para la justificación. Si lo anterior es correcto y además se tiene en cuenta que S1 es un sistema no reflexivo, entonces se tendría que concluir que S1 no es un sistema que produce creencias justificadas de algún tipo (Eraña, en proceso a, p. 16). De donde, la justificación debe necesariamente provenir de la operación del sistema S2 (Eraña, en proceso a, p. 16).

Con una conclusión como esta es posible entender por qué Eraña sostiene que la racionalidad y la justificación no van de la mano. Los defensores de la teoría dual sostienen que los dos

sistemas de razonamiento ofrecen creencias que en su mayoría son racionales, pero según los argumentos de Eraña (en proceso a) sólo los productos de S2 pueden estar justificados, entonces S1 ofrecería creencias que son racionales y que no estarían justificadas. Por su parte, S2 ofrecería en su mayoría creencias racionales y justificadas. De este modo, Eraña sostiene que debemos rechazar la idea de que la noción de justificación y racionalidad están estrechamente ligadas, pues si su argumento fuese correcto todas las creencias justificadas serían racionales, mientras que no todas las creencias racionales estarían justificadas. El siguiente esquema puede ilustrar la postura particular de Eraña (en proceso a).



Lo que es importante subrayar es que Eraña (en proceso a) no está sosteniendo que la noción internista de justificación sea conceptualmente compatible con todos los productos de la teoría dual de sistemas. En términos generales, para ella la noción de justificación no sería compatible con la teoría dual en sí misma, sino sólo con los productos procedentes de S2. La teoría internista de la justificación, desde su perspectiva, no puede ser compatible con los productos provenientes de S1, pues la naturaleza de los procesos que subyacen a este sistema, según Eraña (en proceso a), no pueden albergar la idea de control y responsabilidad epistemológica que ella considera están necesariamente detrás de la noción de justificación.

Una de las formas en que se puede poner en duda esta propuesta de Eraña (en proceso a) es apelando simplemente a las críticas particulares que se han hecho al internismo. En el capítulo I presenté varias críticas a la propuesta internista de Bonjour que pueden ser usadas para objetar la propuesta de Eraña aquí expuesta. En particular las críticas al internismo, revisadas en el

primer capítulo, hacen ver que la noción de justificación no implica necesariamente el punto de vista reflexivo, que sugiere que cuando un sujeto está justificado, cree justificadamente que lo está. No repetiré aquí las críticas al internismo hechas anteriormente, pero considero que es difícil aceptar, como lo hace Eraña (en proceso a), que la única noción de justificación, esto es, "la única que satisface nuestros juicios intuitivos acerca de la justificación— implica la reflexión" (Eraña, en proceso a, p. 15, la traducción es mía). Precisamente las críticas a la postura internista de la justificación es que dicha noción deja de lado otras intuiciones de lo que significa estar justificado (en particular la idea de que detrás de la noción de justificación se encuentran la noción de verdad y los procesos de formación de creencias) (King, 2000).

Si es posible dudar de la noción particular de justificación que defiende Eraña (en proceso a), no habría razones fuertes para sostener que se debe rechazar, como ella lo hace, la estrecha relación que existe entre las nociones de racionalidad y de justificación. Eraña tendría que mostrar por qué es adecuado separar la íntima relación que muchas veces se ha tomado como punto de partida entre la noción de racionalidad y justificación, sin presuponer una idea internista de justificación. Según lo expuesto anteriormente, se tendría que aceptar que la noción de justificación se reduce a los productos de S2 (Eraña, en proceso a), sólo si se presupone que la única noción que recoge las intuiciones de lo que es una creencia justificada es el internismo. Sin embargo, en el primer capítulo argumenté que no existe un único conjunto de intuiciones que la noción de justificación deba recoger, de donde no hay razones para sostener que la noción de justificación necesariamente presupone un cierto tipo de intuiciones. El argumento de Eraña (en proceso a) precisamente busca explicar que la noción de justificación implica reflexión y el de racionalidad no necesariamente implica reflexión, de donde todas las creencias justificadas son racionales, pero no todas las creencias racionales están justificadas. Pero si es verdad que no hay un único conjunto de intuiciones detrás de la noción de justificación, no se puede sostener que esta noción y la de racionalidad no están íntimamente conectadas.

De hecho, estas objeciones en contra de que los productos de S2 están justificados, según criterios internistas, pueden ser refor-

zadas por los argumentos expuestos en la sección anterior, en donde muestro que difícilmente se puede aceptar que los productos de S1 están justificados según estándares externistas, y los productos de S2 están justificados según criterios internistas. Por ejemplo, entre las objeciones que presenté frente a esta postura dual de la justificación es que en S2 hay procesos conscientes que con el paso del tiempo pueden hacerse inconscientes, esto es, procesos automatizados. Si lo anterior es el caso, una creencia que está justificada, según el internismo, dejaría de estarlo debido simplemente a los procesos automatizados, ya que éstos dejarían de ser procesos reflexivos propios de la justificación internista. Si lo anterior es correcto, entonces se estaría frente a la idea contraintuitiva de que una creencia justificada deje de estarlo por la simple automatización de los procesos cognitivos, lo que da pie al rechazo de que los productos de S2 responden a un tipo de justificación internista. Esto, junto con las objeciones señaladas anteriormente, permiten sostener que la teoría dual de sistemas no puede ser compatible con los criterios internistas de justificación.

VI.1.3. LA TEORÍA DUAL DE SISTEMAS Y EL EXTERNISMO

Si he argumentado de manera correcta, no es posible sostener que los *outputs* ofrecidos por el sistema S2 estén justificados según los criterios internistas, ni tampoco es posible sostener que cada sistema de razonamiento brinde creencias justificadas según criterios internistas o externistas. Los argumentos usados en las dos secciones anteriores están dirigidos, principalmente, en contra de la idea de que los criterios internistas de justificación no son conceptualmente compatibles con la teoría dual de sistemas. A mi juicio, lo anterior sugiere que una alternativa viable es defender que la teoría dual de sistemas es compatible únicamente con criterios de justificación externistas. En esta sección brindaré argumentos que apoyan la postura de que los dos sistemas de razonamiento postulados por la teoría dual de sistemas ofrecen *outputs* justificados, según estándares externistas. Expongamos esto.

El externismo defiende una noción de justificación en la cual una creencia está justificada si es producto de un *proceso cognitivo adecuado*. Para el externismo, un sujeto puede estar justificado, aun cuando no crea justificadamente que su creencia sea produc-

to de un proceso adecuado. Como es fácil notar, el externista no niega que los sujetos puedan ser conscientes de que sus creencias sean producto de un proceso confiable. Lo que el externismo sí niega es que siempre que un sujeto está justificado, el sujeto puede acceder por mera introspección a las condiciones que justificaron su creencia. Por su parte, la teoría dual de sistemas sostiene que los seres humanos cuentan con dos sistemas cognitivos distintos. Los sujetos son conscientes sólo de los productos de uno de los sistemas, si son causados por S1, o de los productos y de los procesos que los originaron si pertenecen a S2. Esto ofrece evidencia empírica para sostener que los sujetos cognoscentes no pueden tener acceso a todos los procesos racionales que dieron origen a sus creencias, en particular, no pueden tener acceso a los procesos que pertenecen a S1. Así, una primera manera de pensar que el externismo puede ser compatible con la teoría dual es que existen procesos que subyacen a los dos sistemas, muchos de los cuales son racionales y no son accesibles conscientemente a los sujetos (especialmente a los que pertenecen a S1). Lo que intento mostrar aquí, sin embargo, es que no sólo la teoría dual y una noción de tipo externista son lógicamente compatibles, sino que los productos de los dos sistemas de razonamiento están justificados según criterios externistas. Para ello es necesario mostrar que los procesos de los dos sistemas son procesos adecuados o confiables.

Para mostrar que en S1 subyacen procesos confiables basta con subrayar algunas de las características de este tipo de procesos. Entre los rasgos que hay que destacar es que estos procesos son rápidos, automáticos, ayudan a la supervivencia, son eficaces para resolver problemas a los que los sujetos se enfrentan, además éstos no son conscientes de la operación y control de los mismos. Por este conjunto de características —y por los argumentos ofrecidos anteriormente (Eraña, en proceso b)— considero que los procesos de S1 son adecuados (o confiables en el sentido de que ofrecen respuestas eficientes a problemas particulares), por lo que sus productos están justificados desde una noción externista de la justificación.

Las posturas expuestas anteriormente por Eraña (en proceso b y en proceso a) han intentado mostrar que S2 ofrece creencias justificadas según el criterio internista, apoyándose en la idea de

que el sujeto es consciente y puede tener control sobre los procesos de S2. En general, estas posturas se apoyan en la idea de que S2 es un proceso reflexivo y, en ese sentido, suficiente para la justificación de tipo internista. Si se desea mostrar que los procesos de S2 ofrecen *outputs* que están justificados según criterios externistas, basta con probar que los *outputs* provenientes de S2 son producto de procesos confiables, independientemente de que los sujetos tengan control consciente de algunos de esos procesos. Para mostrar lo anterior es necesario observar que los procesos de S2 son confiables o adecuados, y dado que para una noción externista es suficiente la confiabilidad con la que se producen las creencias para que una creencia esté justificada, entonces S2 produce creencias justificadas, según los criterios externistas. Hay varias razones para apoyar este argumento:

1. Si se toma en cuenta que S2 es un sistema reflexivo, consciente, lento, propiamente humano, se tiende a pensar que estos rasgos asocian a los productos de S2 con el internismo (Eraña, en proceso a y b). Sin embargo, se olvida que desde la postura externista no se niega que la reflexión o los procesos cognitivos conscientes ofrecen creencias que están justificadas. La diferencia entre la postura externista e internista a este nivel es que mientras que para el internismo lo que justifica a las creencias no es el proceso que originó a las creencias, para el externismo si tal proceso es adecuado (o confiable, desde la posición confiabilista) entonces por sí mismo produce creencias justificadas. Las características de los procesos que pertenecen a S2 sugieren que son confiables en distintos sentidos, uno de ellos es que tales procesos siguen reglas explícitas —por ejemplo, fundadas en la lógica— que suponen que sus productos son en su mayoría verdaderos. Otro ejemplo, proveniente de Stanovich y West (2003) es que los procesos pertenecientes a S2 son lo suficientemente eficaces para ayudar a los sujetos a resolver problemas complejos en la sociedad contemporánea. De lo anterior, hay motivos para creer que los procesos que subyacen a S2 son confiables. Es la confiabilidad de tales procesos lo que hace que los productos de los procesos de S2 estén justificados desde la postura externista y no el "punto de vista reflexivo" que es enfatizado por los internistas (Eraña, en proceso a).

2. S2 es un sistema influido culturalmente y difiere entre individuo e individuo a partir de la educación formal. El sujeto aprende a guiar su pensamiento por métodos adquiridos culturalmente, y aquí Stanovich y West (2003) están pensando en la instrucción recibida en las sociedades occidentales contemporáneas. Muchos de estos métodos tienen el propósito explícito de preservar la verdad de las creencias así como de generar mayor número de creencias, verdaderas que falsas. Por ejemplo, cuando los sujetos aprenden estadística o teoría de las decisiones, para poder resolver determinado tipo de problemas, aprenden entre otras cosas reglas que le ayudan a guiar cierto tipo de razonamiento que se encuentra en S2. En ese sentido los procesos de S2 están influenciados por esos métodos y son confiables en tanto que dichos métodos generan mayor número de creencias verdaderas que falsas. Si lo anterior es correcto, es posible afirmar que los productos de S2 están justificados porque son producto de procesos confiables y no por la reflexión que los sujetos tengan acerca de tales procesos, métodos o reglas que los guían.

3. La naturaleza del sistema S2 es despersonalizar y descontextualizar los problemas de razonamiento. Justo la postura externista se ha visto como una postura que tiende a juzgar los criterios de justificación desde una perspectiva objetiva o desde la tercera persona (Eraña, en proceso a [13]). Si la postura externista establece los criterios que determinan la justificación de una creencia desde la tercera persona, entonces la naturaleza del sistema S2, en tanto que descontextualiza y despersonaliza los problemas, toma una actitud desde la tercera persona y con ello sus productos son evaluados desde un acercamiento objetivo, y no subjetivo o de la primera persona como lo hace el internismo. Así, si es adecuado entender el debate entre el internismo y el externismo como un debate entre la evaluación epistémica desde la primera y tercera persona, y los procesos que subyacen a S2 toman una perspectiva objetiva o de la tercera persona, entonces es verosímil sostener que los productos que ofrece S2 están justificados, según los criterios externistas de justificación.

Dadas las características anteriores los procesos de formación de creencias de S2 son adecuados o confiables y dado que para la postura externista la *confiabilidad* es suficiente para producir

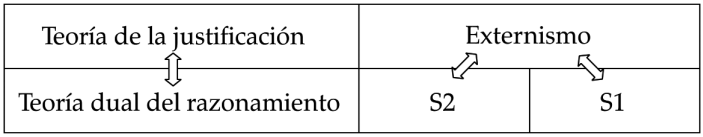
creencias justificadas, entonces las creencias de S2 son creencias justificadas, según criterios externistas. Así, si lo argumentado anteriormente es correcto, tanto los procesos de S1 como de S2 ofrecen creencias justificadas según los estándares externistas.

Además de las razones anteriores para mostrar cómo los productos de lo dos sistemas ofrecen creencias justificadas según los estándares externistas, es preciso notar que los defensores de la teoría dual defienden que en los dos sistemas de razonamiento subyacen procesos racionales [14]. Lo que hay que resaltar aquí es que muchas de las creencias producto de S1 son racionales, aun cuando el sujeto no tenga acceso a los procesos que subyacen a este sistema. Por su parte, la mayoría de creencias producto de S2 también son racionales. Si se tiene en cuenta la idea de que la noción de justificación y la noción de racionalidad están íntimamente ligadas, entonces tiene sentido pensar que las creencias de los dos sistemas son racionales y están justificadas. Esta idea nuevamente sugiere que los procesos que subyacen a los dos sistemas de razonamiento son lo suficientemente adecuados o confiables para producir creencias racionales.

Bajo esta perspectiva, más que una noción de justificación no excluyente, la teoría dual de sistemas ofrece *outputs* justificados solamente bajo criterios externistas. Una noción externista de la justificación compatible con la tesis de que existe una distinción entre los sistemas S1 y S2 podría darle apoyo empírico a la idea de que existe un conocimiento pre-reflexivo (o "conocimiento animal", para usar la terminología de Sosa) en donde para saber no es necesario que el sujeto ofrezca razones que muestren los fundamentos de su saber, y el conocimiento reflexivo en donde el sujeto señala las razones que tiene para sostener el conocimiento que posee (Sosa, 1992 [15]). El conocimiento pre-reflexivo está ligado íntimamente a S1, mientras que el conocimiento reflexivo es muchas veces producto de S2. Teniendo todo lo anterior en cuenta, es preferible esta nueva relación entre la teoría dual del razonamiento y la noción externista de justificación, por sobre las dos propuestas exploradas en los dos primeros apartados de esta sección.

A diferencia de lo expuesto anteriormente, en donde se sostiene que la teoría dual de sistemas supone una noción no excluyente de la justificación, y la propuesta de que sólo los productos de

S2 ofrecen creencias justificadas y los productos de S1 ofrecen creencias racionales y no justificadas, en esta sección he argumentado que los dos sistemas de razonamiento propuestos por la teoría dual de sistemas brindan respuestas que están justificadas según criterios externistas. El siguiente esquema ejemplifica la relación que defiendo en este apartado ente la teoría dual de sistemas y la noción externista de justificación.



VI.2. RACIONALIDAD Y JUSTIFICACIÓN
DESDE LA TEORÍA DUAL DE SISTEMAS

He sostenido que la teoría dual de sistemas en su nivel descriptivo tiene mayor afinidad teórica con el externismo que con una postura híbrida (internista para S2 y externista para S1) o con solamente la teoría internista de la justificación. Además, he sostenido que los productos de los sistemas S1 y S2 están justificados según estándares externistas. Esta conclusión es importante porque, si lo revisado en los otros capítulos es correcto, la tradición de heurística y sesgo, y la psicología evolucionista —en un nivel descriptivo— son igualmente compatibles con las nociones internista y externista de la justificación. En las conclusiones de los capítulos II y III argumenté que las dos teorías psicológicas antes mencionadas no pueden decir nada significativo con relación al debate en torno a la noción de justificación. No obstante, si lo argumentado en la sección anterior es correcto, hay buenas razones para sostener que los *outputs* ofrecidos por los dos sistemas de razonamiento están justificados según los criterios externistas. En otras palabras, la noción de justificación más adecuada con la descripción del razonamiento, según la psicología cognitiva actual, es una teoría de tipo externista.

Ahora bien, como he expuesto en el capítulo anterior, los defensores de la teoría dual de sistemas consideran que el razonamiento debe ser evaluado de al menos dos modos distintos. Evans y Over (1996) postulan que lo que ellos denominan "racio-

nalidad 1" o "racionalidad personal" evalúa el funcionamiento de los procesos subsumidos en S1, y la "racionalidad 2" o "racionalidad impersonal" evalúa el funcionamiento de los procesos que pertenecen a S2. Por su parte, Stanovich y West (2000, 2003) sostienen que, lo que ellos denominan "racionalidad evolucionista" evalúa los procesos de S1 y la "racionalidad normativa" evalúa los procesos de S2. Tal y como cada conjunto de psicólogos define a sus teorías de la racionalidad, éstas implican *i.)* en el caso de Evans y Over, dos acercamientos distintos de racionalidad, a saber, una racionalidad deontológica y otra consecuencialista, y *ii.)* en el caso de Stanovich y West, conducen a un acercamiento consecuencialista de la racionalidad de tipo pragmatista y biológica.

Estos psicólogos han pensado que la teoría dual a nivel descriptivo, es decir, la distinción entre dos tipos de razonamiento conduce a la distinción entre dos distintos tipos de racionalidad. Sin embargo, desde un punto de vista lógico, distinguir entre dos teorías del razonamiento no implica necesariamente la distinción entre racionalidad 1 y racionalidad 2, como lo creen Evans y Over, o entre la racionalidad normativa y la racionalidad evolucionista, como afirman Stanovich y West. No obstante, considero que lo valioso de la teoría dual de sistemas a nivel normativo es que esta teoría propone la existencia de dos estándares normativos distintos para evaluar al razonamiento. En otras palabras, la teoría dual de sistemas se está comprometiendo con la idea de que el razonamiento humano no puede ser evaluado con un único estándar normativo. Uno de los argumentos para sostener esta idea proviene de lo visto en el capítulo IV donde presenté críticas importantes en contra del acercamiento deontológico de la racionalidad, así como en contra de la visión estándar de la racionalidad. Si estas críticas son correctas, no todo el razonamiento puede ser evaluado únicamente con la visión estándar de la racionalidad, sino con acercamientos de tipo consecuencialista [16].

A mi juicio, otra razón para sostener que debe haber al menos dos teorías de la racionalidad para evaluar el razonamiento es a partir de la naturaleza misma de los procesos cognitivos. Según los defensores de la teoría dual de sistemas, la construcción de tareas propias del sistema S1 están altamente contextualizadas y personalizadas, mientras que los procesos pertenecientes a S2 sirven para descontextualizar y despersonalizar los problemas

(Stanovich y West, 2000). Si se sostiene el principio de "el deber implica el poder", defendido en el capítulo IV, difícilmente se puede sostener que los procesos de razonamiento que sirven para lidiar con ciertas tareas, deben ser evaluados con estándares que no tomen en cuenta la tarea en cuestión, ni tampoco con estándares que vayan más allá de las limitaciones de procesos que hayan sido seleccionados para resolver cierto tipo de tareas. Esto se refuerza con la idea de los defensores de la teoría dual de que los procesos de S1 son "antiguos" evolutivamente hablando, mientras que los procesos de S2 son "modernos". Muchos de los ejemplos usados por Stanovich y West (2003) de los procesos "antiguos" son aquellos que se seleccionaron en la etapa de cazadores-recolectores humanos (que servían para resolver los problemas con los que típicamente se enfrentaban). En contraste, los procesos que subyacen a S2, "modernos" evolutivamente hablando, sirven para lidiar con problemas propios de las sociedades contemporáneas, en donde la satisfacción de ciertas reglas formales puede ser importante. Dada esta distinción y tomando en cuenta el principio de que "el deber implica el poder", no es adecuado evaluar con el mismo estándar normativo a procesos que no fueron seleccionados para resolver problemas que de hecho no pueden resolver, esto es, no se debe evaluar con el mismo estándar normativo a los procesos de S1 y a los procesos de S2. Por lo tanto, si los procesos deben ser evaluados con los estándares que pueden satisfacer y se acepta que los procesos que pertenecen a S1 y S2 realizan tareas diferentes (porque tienen un origen evolutivo distinto), entonces deben existir al menos dos estándares normativos para evaluar el razonamiento humano.

La defensa de la distinción entre dos tipos de teorías normativas para evaluar al razonamiento que se ofrece en la teoría dual es que sin tal distinción no se tiene una respuesta a la "paradoja de la racionalidad". La supuesta paradoja surge de la idea intuitiva de que los sujetos son seres racionales y la numerosa evidencia proveniente de los laboratorios psicológicos de que son irracionales (Evans y Over, 1996). Según los defensores de la teoría dual de sistemas, la postulación de dos sistemas de razonamiento y dos teorías de la racionalidad es lo que brinda una explicación de la "paradoja de la racionalidad". Dicha explicación puede ser

usada como un tercer argumento a favor de que el razonamiento no puede ser evaluado por un único estándar normativo.

Si los tres argumentos anteriores son correctos, los defensores de la teoría dual acertadamente sostienen que no hay una única teoría normativa con la que se pueda evaluar el razonamiento humano. No obstante, los argumentos anteriores no demuestran que sólo deban ser dos las teorías normativas, ni tampoco que sean las defendidas por Evans y Over (1996) o las defendidas por Stanovich y West (2000), pero estos argumentos sí sugieren que se necesita de al menos dos modos de entender la racionalidad. ¿Cómo se relacionan estas ideas con la postura que he defendido, según la cual los productos de los dos sistemas de razonamiento postulados por la teoría dual están justificados según los estándares externistas? En tanto no existe consenso entre la teoría dual de sistemas con relación a los estándares normativos que deben evaluar el razonamiento, revisaré por separado tanto la propuesta de Evans y Over, como la propuesta de Stanovich y West. Veamos.

VI.2.1. EXTERNISMO Y LA DISTINCIÓN ENTRE RACIONALIDAD 1 Y RACIONALIDAD 2

Para Evans y Over, la racionalidad personal o racionalidad 1 es la responsable de evaluar a S1, mientras que la racionalidad impersonal o la racionalidad 2 es responsable de evaluar a S2. Según Evans y Over (1996), S1 permite que procesamientos complejos de información se lleven a cabo de manera rápida. Si bien S1 ofrece respuestas rápidas, éstas no siempre son correctas según la visión estándar de la racionalidad. Sin embargo, dado que estas respuestas son producto de procesos que han probado ser eficientes para alcanzar las metas personales de los sujetos, ellas no son consideradas irracionales. Así, desde la racionalidad 1, los resultados de los procesos implícitos pueden ser considerados racionales si éstos son medios eficaces para alcanzar metas personales. S2 es un sistema lento, pero capaz de producir *outputs* que, a juicio de Evans y Over, sistemáticamente satisfacen los requisitos normativos estipulados por la visión estándar de la racionalidad. En cierto sentido, aun cuando los dos sistemas siguen reglas de razonamiento, sólo las creencias que son producto del sistema S2 son sistemáticamente correctas desde la perspectiva de las teorías formales.

Con relación a la racionalidad 1 se puede decir que algunas veces los sujetos cognoscentes son conscientes de las metas que tienen y de los pasos que siguen para alcanzar esas metas. Pero generalmente son inconscientes e incapaces, según Evans y Over, de describir los procesos que los ayudan a alcanzar sus fines o metas. Asimismo, los sujetos cognoscentes no tienen conocimiento acerca de cómo dichos procesos funcionan. En otras palabras, lo sujetos generalmente no tienen conocimiento ni control consciente de muchos de sus procesos cognitivos.

> Los sujetos podrían saber, por ejemplo, que van a salir a comprar el periódico al puesto de revistas, pero no sabrán casi nada acerca de cómo funciona su sistema visual para permitirles tal desplazamiento. El trabajo de este sistema está más allá de su conocimiento o control, y así difícilmente se puede decir que tienen buenas razones para hacer que el sistema produzca representaciones del mundo de un tipo en lugar de otro (Evans y Over, 1996, pp. 9-10, la traducción es mía).

En este caso parece claro que lo que hace que un proceso cognitivo sea racional 1 es la eficiencia de tal proceso para ayudar a los sujetos a conseguir sus metas personales. Este rasgo tiene una fuerte afinidad teórica con el externismo, puesto que para el externismo, aun cuando un sujeto no tenga control sobre sus procesos cognitivos, esto no implica que si dichos procesos son eficientes, entonces el sujeto pueda contar con creencias racionales. De este modo, tanto el externismo como la racionalidad 1 sugieren que detrás de las nociones de justificación y racionalidad 1 están los procesos cognitivos eficientes para alcanzar ciertas metas, aun cuando el sujeto puede no tener control o conocimiento del proceso en cuestión. Esto sugiere que existe compatibilidad teórica entre el externismo y la racionalidad 1.

El que el sujeto no pueda tener control sobre algunos de sus procesos y aun así tales procesos generen creencias racionales, no significa que el sujeto nunca puede tener control sobre los mismos. En el caso de la racionalidad impersonal o la racionalidad 2, ser racional no puede ser producto de un accidente o resultado de la buena o mala suerte (Evans y Over, 1996). Para la racionalidad 2, los sujetos cognoscentes deben tener buenas razones para

actuar o elaborar juicios de la manera en que lo hacen, y ello es parte de la explicación que dan para su acción o su modo de razonar. Esto es ilustrado por Evans y Over de la siguiente manera:

> Los sujetos tendrían buenas razones para concluir que lloverá siguiendo una regla de inferencia, el *modus ponens*, que les permite inferir la conclusión de las proposiciones: "el barómetro está cayendo", y "lloverá, si eso ocurre". Si los sujetos están siendo racionales 2, esperamos que la existencia de tal razonamiento pueda ser indicada en sus protocolos, en donde se les solicita dar un reporte verbal de su pensamiento o razonamiento mientras lo realizan (1996, p. 9. la traducción es mía).

En el caso de la racionalidad 2, un sujeto es racional 2, si sigue un conjunto de reglas y, en principio, satisface tales reglas. Generalmente se cree que para las posturas externistas no es importante la noción de reglas de razonamiento (Eraña, en proceso b). Sin embargo, en sí mismo el externismo puede ser compatible con la idea de que satisfacer ciertos estándares normativos es lo que puede hacer de un proceso o método de formación de creencias un proceso confiable o adecuado. Además, si bien la idea de reglas explícitas de razonamiento pueden sugerir una conexión entre el internismo y la racionalidad 2, Evans y Over (1996) no están afirmando que siempre que alguien es racional 2, el sujeto debe tener acceso consciente a las reglas y además, como lo hacen algunos internistas (Bonjour, 1985), creer justificadamente que las reglas que está siguiendo son reglas que conducen a la verdad.

La idea que quiero subrayar aquí es que la distinción entre racionalidad 1 y racionalidad 2 defendida por Evans y Over (1996) puede ser compatible con la postura de que los productos de los dos sistemas de razonamiento ofrecen resultados justificados según los criterios externistas de justificación. Profundizando un poco en la relación entre la noción externista de justificación y la distinción entre racionalidad 1 y racionalidad 2 propuesta por Evans y Over, en el capítulo anterior he argumentado que, tal y como postulan la distinción Evans y Over, la racionalidad 1 se compromete con una postura de tipo consecuencialista pragmatista, mientras que la racionalidad 2 se compromete con la visión estándar de la racionalidad. Si bien el externismo parece compa-

tible con estos dos acercamientos indagaré qué tan fuerte es la afinidad teórica entre tales teorías.

Los productos provenientes de S2, según Evans y Over, satisfacen en su mayoría el criterio provisto por la racionalidad 2 y con ello satisfacen también los criterios provenientes de la visión estándar de la racionalidad. En este sentido, si los productos provenientes de S2 satisfacen la visión estándar de la racionalidad y, como he argumentado, los productos de este sistema están justificados por criterios externistas, entonces parecería que el criterio externista debería ser compatible con la visión estándar de la racionalidad. Se ha argumentado anteriormente que la visión estándar de la racionalidad se encuentra enmarcada dentro del acercamiento deontológico, por lo que la conexión se ampliaría a que como los productos de los procesos pertenecientes a S2 satisfacen la racionalidad 2 y están justificados según el externismo, entonces debería ser el caso que las creencias justificadas según el criterio externista que provengan de S2 sean racionales según la racionalidad 2, según la visión estándar de la racionalidad y según un cierto tipo de acercamiento deontológico.

El argumento anterior, sin embargo, es lógicamente incorrecto, puesto que tiene una estructura del tipo: si A implica B y si A implica C, entonces B debe implicar C, o C debe implicar B. En otros palabras, de que los productos de S2 estén justificados según criterios externistas y estos productos sean en su mayoría racionales 2, no se sigue que el criterio externista se comprometa necesariamente con una visión de la racionalidad 2, esto es, con una visión estándar de la racionalidad, o que las creencias que satisfacen la visión estándar de la racionalidad deban estar justificadas según la noción externista. He tratado de argumentar que el criterio externista de justificación puede ser compatible con la visión estándar de la racionalidad, pero de ahí no se puede sostener que tal teoría de la racionalidad conduzca a dicha teoría de la justificación y viceversa.

¿Pero qué sucede con la noción de justificación externista y la racionalidad 1? Para que una creencia producto de S1 sea racional no es necesario satisfacer las normas provenientes de la visión estándar de la racionalidad, sino satisfacer un tipo de acercamiento consecuencialista de la racionalidad. Este acercamiento a la racionalidad, como se ha explicado anteriormente, considera que

ser racionales consiste en actuar de manera tal que probablemente se alcancen ciertas metas o resultados [17]. Si los procesos de S1 satisfacen los criterios de la racionalidad 1 en el que subyace un acercamiento consecuencialista de la racionalidad y los productos de S1 están justificados según criterios externistas, entonces el externismo parecería —a primera vista— una postura compatible con un acercamiento consecuencialista de la racionalidad. De hecho, el confiabilismo de Goldman, por ejemplo, es una manera de entender al consecuencialismo, pues sostiene que un proceso de razonamiento adecuado es aquel que tiende a producir mayor número de creencias verdaderas que falsas (Goldman, 1986). Esto significaría que una creencia producida por S1 que está justificada, según criterios externistas, es una creencia racional según el acercamiento consecuencialista de la racionalidad.

Es difícil sostener que en tanto S1 ofrece respuestas que bajo la noción externista estarían justificadas, y dado que S1 es evaluado por la racionalidad 1, entonces la noción externista se comprometería con la racionalidad 1 [18]. La idea que parece apoyar esta postura es que una propuesta externista como la de Goldman se compromete con el consecuencialismo. De ahí, se puede falsamente creer que es adecuado pensar que el vínculo entre la noción externista y S1 conducen necesariamente a un vínculo entre el externismo y la racionalidad 1, por lo tanto, entre el externismo y el consecuencialismo. Sin embargo, la propuesta particular que Goldman defiende es un consecuencialismo de tipo veritista y la racionalidad 1 que Evans y Over tienen en mente es un consecuencialismo pragmatista. Estos psicólogos consideran que la importancia de la racionalidad 1 es evaluar a los sistemas de razonamiento a partir de qué tanto acercan a la obtención de metas y deseos personales, y no de nociones epistemológicas como la verdad. Partiendo de la teoría dual, no se podría sostener que una creencia que está justificada según el externismo es una creencia que es racional creer según la racionalidad 1 o el consecuencialismo pragmatista. Si los *outputs* de S1 están justificados por criterios externistas y si además se comparte la idea de los estándares normativos propuestos por Evans y Over, entonces la teoría dual de sistemas en su nivel normativo y descriptivo no puede ser compatible con cualquier tipo de externismo, a saber, no podría ser compatible con los externismos veritistas.

En el apartado anterior sostuve que los productos de los dos sistemas de razonamiento ofrecen productos justificados según los estándares externistas. Los argumentos para apoyar esta idea descansan en la naturaleza misma de los dos sistemas de razonamiento y en las críticas a las posturas "duales" e internista de conectar a la teoría dual de sistemas con sus nociones de justificación. Las propuestas más importantes de la teoría dual de sistemas están acompañadas de una teoría dual de la normatividad, en el caso de Evans y Over una teoría consecuencialista y otra deontológica para evaluar el razonamiento. A cierto nivel, la teoría externista es compatible con estas dos posturas, pero es difícil sostener que cualquier tipo de teoría externista aceptaría la propuesta normativa particular de Evans y Over. En la versión de estos psicólogos, la teoría dual defiende una teoría dual de la racionalidad, donde se postula una racionalidad 1, entendida como un tipo de consecuencialismo pragmatista, que evalúa a S1, y una racionalidad 2, entendida como la visión estándar de la racionalidad y un tipo de acercamiento deontológico de la racionalidad, que evalúa a S2. Si lo argumentado en esta sección es correcto, no todo tipo de externismo puede ser compatible con esta teoría dual de la racionalidad. Si se desea fundar una teoría externista en la propuesta normativa defendida por Evans y Over se tendría que dar lugar a una teoría externista que distinga dos tipos de justificación que den cabida a la racionalidad 1 y a la racionalidad 2, y además se tendría que defender un externismo no veritista.

VI.2.2. EXTERNISMO Y LA DISTINCIÓN ENTRE RACIONALIDAD EVOLUCIONISTA Y RACIONALIDAD NORMATIVA

Si bien los defensores de la teoría dual de sistemas comparten un grupo central de ideas (a un nivel descriptivo), una de sus divergencias se encuentra en la manera de concebir la racionalidad. Para Stanovich y West (2000, p. 660), existe una relación entre los dos sistemas cognitivos y dos maneras de entender la racionalidad: por un lado, postulan una "racionalidad normativa" que es usada para evaluar a S2, en la que el producto de un proceso cognitivo es racional si conduce de manera eficaz a la obtención de las metas de un individuo y, por otro lado, postulan una "racionalidad evolucionista", que vinculada a S1, sostiene que el

producto de un proceso cognitivo es racional, si optimiza la metas adaptativas de los genes. Estas dos teorías de la racionalidad se encuentran subsumidas en un acercamiento consecuencialista de la racionalidad de distintos tipos: la racionalidad evolucionista se compromete con un consecuencialismo de tipo biológico y la racionalidad normativa se compromete con un consecuencialismo de tipo pragmatista.

Indagaré ahora si es posible crear un "puente" entre la noción externista de justificación y las visiones de racionalidad defendidas por Stanovich y West a través de la teoría dual de sistemas. Si la mayoría de los productos resultado de los procesos que subyacen a S1 y S2 son racionales, según el acercamiento consecuencialista, y los productos de tales procesos están justificados, según el criterio externista de la justificación, entonces parece que se tiene que aceptar que una creencia que es racional desde la postura consecuencialista debe estar justificada según criterios externistas. Si lo anterior fuese correcto, la teoría dual apoyaría la idea de que una creencia que está justificada, según los estándares externistas, es una creencia que es racional creer desde un acercamiento consecuencialista. Es necesario hacer diversas precisiones. En primer lugar, si bien he argumentado que las dos nociones de racionalidad que defienden Stanovich y West son de tipo consecuencialista, lo que estos psicólogos denominan "racionalidad evolucionista" se compromete con un consecuencialismo biológico, mientras que lo que ellos denominan "racionalidad normativa" puede entenderse como un consecuencialismo de tipo pragmatista. Si los productos de S1 y S2 satisfacen los criterios externistas de justificación y tales productos son evaluados por dos formas de consecuencialismo distinto, entonces el externismo conduciría a un consecuencialismo biológico y pragmatista vía la teoría dual de sistemas. ¿Es adecuado este vínculo entre el externismo y el consecuencialismo hecho a partir de la teoría dual de sistemas? La respuesta a esta pregunta es *no*. En la sección anterior he señalado que ese tipo de relación parte de un argumento lógicamente incorrecto. De que los productos de S1 y S2 están justificados según criterios externistas y que los productos de estos sistemas sean racionales según distintos tipos de consecuencialismo, no se sigue que los productos justificados de modo

externista sean racionales desde acercamientos consecuencialistas (biológicos o pragmatistas) y viceversa.

¿Acaso los criterios que los externistas defienden son de tipo consecuencialista pragmatista o biológico? ¿Una creencia que está justificada, según la visión externista de la justificación, es una creencia que es racional creer bajo un acercamiento consecuencialista pragmatista o biológico? Las respuestas a estas preguntas son nuevamente *no*. La propuesta confiabilista de Goldman, expuesta con cierto detalle en el primer capítulo de este libro, es un buen ejemplo de cómo una teoría externista se puede comprometer con un acercamiento consecuencialista de la racionalidad, sin que ello suponga el tipo de consecuencialismo que Stanovich y West tienen en mente. El criterio general defendido por Goldman (1979) es que una creencia está justificada si es producto de un proceso cognitivo confiable, entendiendo por "confiable" aquel que produce mayor número de creencias verdaderas que falsas. En cambio, los procesos que subyacen a S1 y S2 son racionales en tanto que, según Stanovich y West (2003), *a)* conducen frecuentemente a las metas que los sujetos particulares buscan, o *b)* conducen a la satisfacción de metas tales como la reproducción, supervivencia o la propagación de los genes. Claramente *a)* y *b)* son evaluados de modo consecuencialista, pero no del modo en que Goldman u otros externistas lo conciben [19]. Esto muestra, en primer lugar, que existen teorías externistas que no empatan con la tríada externismo-dos sistemas de razonamiento-racionalidad evolucionista y normativa. En segundo lugar, la teoría dual de sistemas ofrece productos justificados según la noción externista, pero si a la teoría dual de sistemas le agregamos la teoría dual de la normatividad que Stanovich y West defienden, nos percatamos que no cualquier tipo de externismo puede ser compatible con su teoría de la racionalidad. En particular el tipo de teoría externista que puede ser compatible con la teoría dual de sistemas en su nivel descriptivo y normativo defendido por Stanovich y West tendría que eliminar la idea de que la verdad es el tipo de consecuencia usado para evaluar el razonamiento humano y debe, al menos, distinguir entre dos tipos de justificación que rescaten los dos tipos de racionalidad conscencuencialista defendidos por Stanovich y West.

En suma, he explorado dos propuestas normativas de los defensores de la teoría dual: la propuesta de Evans y Over, y la propuesta de Stanovich y West. Como he señalado anteriormente, no existe una implicación lógica entre la postulación de dos sistemas de razonamiento y las teorías normativas particulares que han propuesto los defensores de la teoría dual. Sin embargo, sí hay motivos suficientes para aceptar al menos dos teorías normativas para evaluar el razonamiento humano. En este apartado he tratado de conectar la propuesta particular de Evans y Over, así como la de Stanovich y West con la idea de que los productos de S1 y S2 están justificados según criterios externistas. Tanto la propuesta defendida por Evans y Over, como la de Stanovich y West, son compatibles con mi tesis de que los *outputs* de los dos sistemas de razonamiento están justificados según el externismo. Sin embargo, si se desea que la teoría externista de la justificación se funde en las teorías normativas particulares que se han defendido junto con la teoría dual de sistemas, se tendría que sostener que *i.)* las creencias productos de S1 y S2 están justificadas bajo un tipo de externismo no veritista, esto es, no toda postura externista puede ser compatible con la teoría dual de sistemas en su nivel descriptivo y normativo, y *ii.)* una teoría externista que comparta los presupuestos normativos de la teoría dual tendrá que —de alguna manera— distinguir entre al menos dos niveles de justificación, cada uno de ellos dependiente de los estándares normativos con los que se evalúa el razonamiento.

VI.3. UNA NOCIÓN EXTERNISTA COGNITIVISTA DE LA JUSTIFICACIÓN

Una de las preguntas que motiva este capítulo es, "¿puede la teoría dual de sistemas decir algo acerca del debate entre el internismo y el externismo?" Si lo argumentado aquí es correcto, los *outputs* ofrecidos por los dos sistemas de razonamiento postulados por la teoría dual de sistemas no satisfacen los criterios internistas de justificación, ni tampoco es concebible que tales *outputs* satisfagan una noción no excluyente de justificación (en donde internismo y externismo responden a distintos sistemas de razonamiento). La postura que defiendo es que los productos de los dos sistemas de razonamiento postulados por la teoría dual están justificados por estándares externistas. Ahora bien, he seña-

lado que no existe consenso acerca de cuáles son las teorías normativas que postulan los defensores de la teoría dual de sistemas. Una primera propuesta (la defendida por Evans y Over) sugiere que el razonamiento debe ser evaluado por un acercamiento consecuencialista y deontológico de la racionalidad. Otra propuesta (la defendida por Stanovich y West) postula que el razonamiento debe ser evaluado únicamente por criterios consecuencialistas. Si se desea que una teoría externista de la justificación se funde en la teoría dual de sistemas y se acepta, como lo han hecho los defensores de esta teoría, la existencia de al menos dos teorías de la racionalidad, entonces dicha noción externista de justificación debe aceptar dos niveles distintos de justificación. Además, si lo esbozado aquí es correcto, estas nociones de racionalidad no están comprometidas con la verdad, por lo que una teoría externista de la justificación que quisiera apoyarse en la teoría dual de sistemas tendría que ser no veritista. En lo que resta de este capítulo, esbozaré las líneas generales de una teoría externista apoyada en la teoría dual de sistemas y en los postulados normativos defendidos por Stanovich y West (2000, 2003).

La idea de que una teoría de la justificación debe apoyarse en una teoría del razonamiento surge del debate revisado en el primer capítulo de este trabajo, donde las intuiciones que defienden las teorías internistas y externistas en torno al debate sobre la justificación hacen referencia a la racionalidad y el razonamiento. Si es importante que las teorías de la justificación se apoyen en las teorías del razonamiento, y la mejor teoría del razonamiento dada la evidencia actual es la teoría dual de sistemas, entonces es verosímil hacer uso de esta teoría para forjar una noción de justificación. Por lo argumentado en este capítulo, considero que el tipo de teoría que se debe construir es de tipo externista. Como otras teorías externistas, mi propuesta considera que lo que justifica a una creencia es el proceso que dio origen o causa una creencia. En este sentido, muchas veces los sujetos pueden tener creencias justificadas, sin saber que están justificados.

Mi propuesta externista, que denominaré "externismo cognitivista" (EC), parte de la idea defendida por la teoría dual de que existen dos sistemas de razonamiento que son evaluados por dos teorías de la racionalidad distintas. *Grosso modo,* una creencia estará justificada si es producto de un proceso racional. Un pro-

ceso racional es aquel proceso que conduce a aquello que los individuos valoran o que conduce a metas de tipo biológico, como la supervivencia y la reproducción. En tanto un proceso racional puede serlo de dos modos, distinguiré entre dos tipos de justificación:

— Justificación 1 es la creencia que es producto de un proceso racional bajo un criterio de tipo consecuencialista biológico.
— Justificación 2 es la creencia que es producto de un proceso racional bajo un criterio de tipo consecuencialista pragmatista.

Así una creencia está justificada 1 si es una creencia racional desde el consecuencialismo biológico, esto es, será racional o estará justificada en el sentido de que el proceso que lo produjo es eficiente para alcanzar las metas biológicas, tales como la reproducción y la supervivencia. En contraste, una creencia está justificada 2 si es una creencia racional según el consecuencialismo pragmatista, i.e., el proceso que la produjo es un medio eficaz para alcanzar metas personales. Las nociones de justificación 1 y justificación 2 no son excluyentes. Habrá creencias que estén justificadas de las dos maneras, muchas creencias estarán justificadas sólo de un modo, y habrá creencias que no estarán justificadas de ninguna manera.

Generalmente los procesos que subyacen a S1 darán lugar a creencias justificadas 1, es decir, creencias que son producto de procesos que ayudan a alcanzar ciertas metas biológicas. Este tipo de procesos, como lo indica la teoría dual, suelen ser automáticos, no se pueden controlar, se comparten con otros animales, son rápidos, etc. Los procesos que subyacen a S2 muchas veces darán lugar a creencias justificadas 2, es decir, creencias que son producto de procesos que ayudan a alcanzar metas individuales. Los procesos que pertenecen a S2, como repetidamente he señalado, son cercanos a la inteligencia analítica, explícitos, conscientes, lentos, propiamente humanos, etc. Es con relación a las creencias justificadas 2 donde el sujeto puede ofrecer y dar razones, ser crítico, seguir reglas de un cierto tipo e incluso ser epistémicamente responsable. Sin embargo, el ser crítico o la responsabilidad epistémica no es lo constitutivo del buen razonamiento, sino que se puedan alcanzar las metas que el sujeto valora. Con relación a

las creencias justificadas 1, si son producto de S1, el sujeto no tiene otra cosa que aceptarlas. Con respecto a S1, los sujetos no pueden decidir qué creer, la automaticidad de estos procesos y la rapidez con que ofrecen resultados impiden que muchas veces el sujeto pueda evaluar sus resultados críticamente.

Lo distintivo de EC con relación a las teorías de la justificación expuestas en el primer capítulo es que la teoría que propongo se compromete con el consecuencialismo, pero no con el consecuencialismo de tipo veritista. El confiabilismo de Goldman (1986) es una teoría veritista de la justificación, ya que es con relación al número de respuestas verdaderas que arroja un sistema cognitivo como dicho sistema pueda ser considerado confiable y por ende sus resultados justificados. Del mismo modo, el internismo de Bonjour (1985) se compromete con la idea de que la noción de justificación es instrumental y es adecuada si puede mostrarse que los criterios que determinan la justificación de una creencia conducen a la verdad. Según Bonjour (1978), el conocimiento requiere de la justificación y el rasgo característico de este tipo de justificación es su relación interna o esencial con la verdad. El tipo de externismo que defiendo, denominado EC, no es veritista, donde, lo importante de la noción de justificación no es su relación con la verdad sino con la racionalidad.

Dos argumentos por los cuales considero que la noción de justificación no puede evaluarse a partir de la eficacia para conducir a la verdad son:

1. Una postura de tipo veritista sostiene que el criterio adecuado de justificación es aquel que demuestre que conduce a la verdad. En otras palabras, para un consecuencialismo de tipo veritista, T es el criterio de justificación correcto si es posible demostrar que T conduce a la verdad. Sin embargo, para poder demostrar que T conduce a la verdad se requiere de un argumento para demostrarlo. Si efectivamente se desea que tal argumento demuestre algo, tal argumento debe estar justificado o valorado adecuadamente por un metacriterio. ¿Bajo qué metacriterio debe estar justificado tal argumento? No se puede sostener que dicho metacriterio es T, pues se cometería petición de principio al suponer la validez de T como metacriterio para mostrar que T conduce a la verdad. Pero si se sostiene que el metacriterio no requiere del

criterio T, entonces no se comete petición de principio, pero se supondría la adopción de otro criterio que determina cuándo algo es el principio adecuado de justificación, por lo tanto, tal criterio tendría que demostrarse nuevamente y esto conduciría a un regreso al infinito. Por lo tanto, demostrar que un criterio es el adecuado a partir de su conducción a la verdad corre el peligro de cometer una petición de principio o conducir a la adopción de otro criterio de racionalidad que a su vez requiera de demostración.

2. Laudan (1984) ha argumentado, en un contexto diferente, que existe un tipo de metas que pueden definirse con precisión, pero que no existen criterios que determinen cuándo se ha alcanzado la meta y cuándo no. La verdad, a juicio de este filósofo, es el tipo de metas que puede ser definida adecuadamente —al parecer, cree Laudan (1984, p. 53), hay cierto consenso en torno a la definición de verdad— aunque no se ha podido especificar un criterio para determinar cuándo se ha alcanzado y cuándo no. Evaluar a un criterio que determine cuándo una creencia esté justificada requiere no sólo de una noción de verdad, sino de un criterio que determine cuándo se satisface tal criterio. Con todo, no es claro que exista tal criterio, por lo que no es operacionalmente posible evaluar un criterio a partir de la conducción a la verdad.

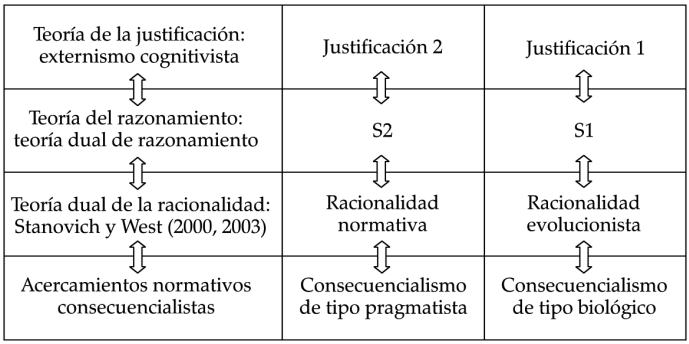
A mi juicio, los argumentos 1 y 2 sugieren que existen dificultades teóricas al aceptar que la conducción a la verdad debe ser el criterio que se debe usar para evaluar los estándares normativos que determinan cuándo una creencia está justificada. Estos problemas a los que se enfrenta un acercamiento veritista de la racionalidad dan pie a explorar acercamientos consecuencialistas no veritistas dentro de la noción de justificación.

Por otro lado, si se desea apoyar una teoría de la justificación en una teoría del razonamiento, se debe indagar si la noción de verdad juega un papel importante en el modo en que los psicólogos cognitivos evalúan los procesos y mecanismo de razonamiento humano. Frente a esto, ni los defensores de la tradición de heurística y sesgo ni los psicólogos evolucionistas comúnmente apelan a la noción de verdad como último criterio para juzgar la ejecución del razonamiento humano. Lo mismo ocurre con la teoría dual de sistemas. No es claro que los procesos que subyacen a S1 o S2 sean procesos que lleven a la verdad; si bien podrían

conducir a ella, la racionalidad de tales procesos no es evaluada en tanto conduzca a esa meta epistémica. Los defensores de la teoría dual de sistemas no sostienen que S1 o S2 están compuestos de mecanismos eficaces, en el sentido de que conducen sistemáticamente a la verdad [20]. La EC es una teoría consecuencialista de tipo pragmatista y biológica, cuyos criterios para determinar los estándares adecuados para justificar una creencia no pertenecen a un ámbito veritista. La crítica que se puede hacer a una postura como ésta es que no habría nada propiamente epistemológico en una noción de justificación que deje de lado la verdad [21]. De momento considero que este problema rebasa los alcances de este libro, sin embargo creo que si las teorías del razonamiento contemporáneas no han integrado la noción de verdad en sus teorías normativas, y además, se sostiene que una noción de justificación debe apoyarse en las teorías psicológicas empíricas, entonces se debe concebir nociones de justificación en donde la noción de verdad (o la conducción a la verdad) no juegue el papel más importante.

En suma, la EC es una teoría que se apoya en la teoría dual de sistemas, en donde se distinguen dos modos de estar justificado. Una creencia está justificada 1 si es producto de un proceso racional evolucionista, esto es, un tipo de acercamiento consecuencialista biológico. Una creencia está justificada 2 si es producto de una proceso racional en términos de la racionalidad normativa, esto es, un tipo de acercamiento consecuencialista pragmatista. Generalmente el buen funcionamiento de S1 generará creencias justificadas 1, mientras que el buen funcionamiento de S2 producirá creencias justificadas 2. Esta teoría es externista en tanto que lo que justifica a las creencias es el que hayan sido causada de manera adecuada. Según lo revisado en este capítulo, la teoría dual en su nivel descriptivo y normativo no puede apoyar cualquier tipo de externismo, sino uno que asuma un acercamiento no veritista y dos niveles de justificación distintos, como lo hace EC. Considero que el esbozo de una teoría externista como el hecho en estás páginas es insuficiente, pero a mi juicio muestra el tipo de teoría que podría hacer uso de las teorías psicológicas para ofrecer una visión de lo que significa que una creencias esté justificada, así como una explicación de la relación entre las no-

ciones de justificación y racionalidad. Una visión esquemática de la teoría EC podría ser el siguiente.



VI.4. CONCLUSIONES

El hecho de que la teoría dual de sistemas postule dos sistemas de razonamiento en donde subyacen procesos con ciertas características (por ejemplo, procesos conscientes y procesos inconscientes), y que haga uso de dos teorías normativas para evaluar el razonamiento humano, hace pensar que la aceptación de la teoría dual de sistemas tiene implicaciones en el debate entre el internismo y el externismo estudiado en este trabajo. Por este y otros motivos, en este capítulo he indagado lo que la teoría dual de sistemas puede decir acerca de la noción de justificación.

En el capítulo reviso tres propuestas particulares que tratan de relacionar la teoría dual de sistemas con distintas nociones de justificación. La primera de ellas sostiene que los *outputs* ofrecidos por S1 responden a la noción externista de la justificación, mientras que los *outputs* ofrecidos por S2 responden a la noción internista de justificación (Eraña, en proceso b). La segunda propuesta defiende una noción particular de justificación, en la que dicha noción implica necesariamente tomar un punto de vista reflexivo. En tanto que S1 no es un sistema reflexivo, sólo los *outputs* ofrecidos por S2 pueden ser racionales y estar justificados. Según esta propuesta, los *outputs* de S1 serían racionales, pero no estarían

justificados (Eraña, en proceso a). La tercera propuesta que he argumentado es la más adecuada en tanto defiende la idea de que los dos sistemas de razonamiento ofrecen *outputs* que están justificados según criterios externistas. En este capítulo he ofrecido razones para aceptar esta última postura.

Así, he mostrado que la teoría dual de sistemas no es neutral con relación al debate entre el internismo y el externismo en teoría de la justificación, puesto que la naturaleza de los procesos que subyacen a los sistemas de razonamiento sugieren fuertemente que los productos causados por tales procesos están justificados según criterios externistas. Esto contrasta con los capítulos anteriores, en donde había mostrado que las teorías del razonamiento expuestas a lo largo del trabajo resultaban —al menos en un nivel descriptivo— igualmente compatibles con las nociones de justificación aquí estudiadas. Si lo argumentado en este capítulo es correcto, la teoría dual de sistemas sí puede ser usada para resolver el debate entre las teorías internistas y externistas de la justificación.

Ahora bien, en una segunda sección de este capítulo he revisado cómo la idea de que los productos de los dos sistemas de razonamiento están justificados según estándares externistas se conecta con las teorías normativas con las que se comprometen los defensores de la teoría dual. He hecho notar que los dos sistemas de razonamiento no implican necesariamente la distinción que Evans y Over (1996) o que Stanovich y West (2000, 2003) hacen. Además, ofrecí argumentos para señalar que se deben postular al menos dos teorías normativas para evaluar el razonamiento. Revisando las dos propuestas particulares (la de Evans y Over y la de Stanovich y West) es posible percatarse que la noción externista de justificación es compatible con las propuestas normativas defendidas por la teoría dual de sistemas. Aun así, la afinidad teórica que existe entre los criterios externistas y sus propuestas normativas particulares es débil. En primer lugar, ninguna de las propuestas normativas defendida por la teoría dual de sistemas sostendría que la verdad es el criterio determinante para evaluar el razonamiento humano. Debido a que la mayoría de teorías externistas se han comprometido con un consecuencialismo veritista, los criterios normativos que aceptan los defensores de la teoría dual no pueden ser compatibles con cualquier tipo de externismo. En segundo lugar, las propuestas nor-

mativas postuladas para evaluar el razonamiento son duales, por lo que si el criterio de justificación se funda en estos acercamientos, se tendría que sostener que tal criterio debe distinguir al menos dos maneras de estar justificado.

En suma, en este capítulo he defendido que los productos de los dos sistemas de razonamiento postulados por la teoría dual están justificados por estándares externistas. Además, si se asumen las teorías normativas que han defendido quienes postulan la teoría dual de sistemas, entonces se tiene que sostener un criterio externista de justificación no veritista que distinga al menos dos niveles de justificación distintos. Al final de este capítulo he esbozado una teoría externista (que he denominado "externismo cognitivista") de la justificación inspirada en la propuesta normativa defendida por Stanovich y West (2000, 2003) y en la existencia de dos sistemas de razonamiento en donde operan procesos de naturaleza distinta.

VII.
CONCLUSIONES.
RACIONALIDAD, RAZONAMIENTO Y LA NOCIÓN DE JUSTIFICACIÓN EPISTÉMICA

Casi han pasado tres décadas desde la publicación de dos artículos en teoría de la justificación que, en gran medida, han marcado el debate entre las nociones internistas y externistas hasta nuestros días: uno es "The internalist conception on justification", escrito por Goldman (1980), el otro, escrito por Bonjour (1980) "Externalists theories of empirical knowledge". Estos textos son paradigmáticos, porque en ellos se pueden observar claramente las intuiciones que cada noción de justificación defiende. En este trabajo he partido de este debate y he propuesto una manera de abordarlo. Mi argumento se apoya en que las dos nociones defendidas por los internistas y externistas hacen referencia, directa o indirecta, a la racionalidad y al razonamiento, por lo que un estudio exhaustivo en dichos campos puede echar luz sobre el debate contemporáneo en torno a la noción de justificación.

A lo largo de este libro hice uso, como lo han hecho algunos filósofos, de una distinción entre tres distintas maneras de estudiar a la razón: el estudio descriptivo, el normativo y el evaluativo (Samuels, Stich, Faucher, 2004). El primero consiste en describir, generalmente de modo empírico, cómo es que el ser humano de hecho razona, así como descubrir los mecanismos y procesos psicológicos que subyacen en los patrones de razonamiento observados. Este tipo de estudios es llevado a cabo principalmente por psicólogos cognitivos, antropólogos cognitivos y científicos computacionales, y aquí se pueden ubicar a las diversas teorías del razonamiento que he explorado. El estudio normativo de la razón intenta dar cuenta, a diferencia del estudio descriptivo, no tanto de cómo los seres humanos razonan, sino de cómo deben razonar. Una de las metas centrales en este tipo de estudios es

descubrir las reglas o principios que indiquen qué es razonar correctamente. A este estudio pertenecen los dos acercamientos a la racionalidad revisados en este trabajo, es decir, los acercamientos deontológicos y el consecuencialista, así como las distintas maneras de entenderlos. Por último, el estudio evaluativo de la razón busca determinar el alcance que tiene el razonamiento humano de acuerdo con un estándar normativo específico. Dados ciertos criterios de lo que cuenta como buenos razonamientos, quienes realizan este tipo de estudio evalúan cuándo el razonamiento cumple con tales criterios. Por ejemplo, el proyecto evaluativo de la razón podría indicar cuándo el ser humano ha actuado racionalmente con respecto a ciertas normas y principios. Aquí se encuentra, por ejemplo, la interpretación pesimista de la racionalidad humana, que surge de la descripción del razonamiento hecha por la tradición de heurística y sesgo, y de la aceptación de la visión estándar de la racionalidad.

En este trabajo he explorado tres distintas teorías del razonamiento, a saber, la tradición de heurística y sesgo, la psicología evolucionista y la teoría dual de sistemas. Cada una de estas teorías descriptivas, de manera explícita o implícita, tiene vínculos con una teoría evaluativa y normativa de la racionalidad. La primera teoría del razonamiento expuesta en este libro es la tradición de heurística y sesgo. La teoría evaluativa, con la que muchos psicólogos y filósofos se comprometen en esta tradición, es con la interpretación pesimista de la racionalidad. Para llegar a dicha interpretación estos teóricos parten de un criterio normativo particular, i. e., la visión estándar de la racionalidad, que como señalé en el capítulo IV, se funda en el acercamiento deontológico de racionalidad. Si lo argumentado en este trabajo es correcto, la tradición de heurística y sesgo no puede ser usada para decidir qué noción de justificación es la más adecuada. Es este sentido, tanto la noción internista como la externista pueden ser compatibles con el aspecto descriptivo de esta teoría. Lo que sí tiene implicaciones importantes en la teoría de la justificación es la interpretación pesimista, que puede apoyarse en la tradición de heurística y sesgo. La versión más fuerte de esta interpretación (que denominé interpretación pesimista [II]) supondría que no existen creencias racionales y en ese sentido no se podría hablar de justificación. Esta interpretación del razonamiento humano,

según lo visto en el capítulo II, tiene implicaciones escépticas, pues si no es posible ofrecer una noción de justificación debido a que los seres humanos son sistemáticamente irracionales, y la noción de conocimiento requiere de la noción de justificación, entonces no se podría sostener que los sujetos saben.

La segunda teoría del razonamiento que exploré es la psicología evolucionista. Esta teoría descriptiva fue contrastada con la tradición de heurística y sesgo, en tanto que los defensores de la misma la han presentado como una rival a tal tradición psicológica. Tres de sus tesis centrales fueron destacadas, esto es, la tesis de que la mente tiene una estructura masivamente modular, la hipótesis frecuentista y la hipótesis de la detección del mentiroso o tramposo. Con este conjunto de ideas centrales y la aceptación de la visión estándar de la racionalidad, los psicólogos evolucionistas ofrecen una interpretación optimista de la racionalidad humana. Así, argumenté que a nivel descriptivo, la psicología evolucionista sostiene que los mecanismos de razonamiento poseen un carácter modular; a nivel normativo, los psicólogos evolucionistas asumen la visión estándar de la racionalidad, que los conduce a un acercamiento de tipo deontológico y, por último, a un nivel evaluativo, los psicólogos evolucionistas defienden una interpretación optimista de la racionalidad humana. Si bien dicha interpretación puede ser entendida de distintas maneras, una evaluación de las mismas permite seleccionar cuáles de ellas tiene mayor respaldo empírico. Según lo explorado en el tercer capítulo de este libro, aun concediendo que las tesis centrales de la psicología evolucionista sean correctas, su posición optimista es compatible con una de las interpretaciones pesimistas de la tradición de heurística y sesgo. La importancia de la evidencia proporcionada por la psicología evolucionista radica en que con ésta es posible descartar la interpretación pesimista más radical. Como ocurre con la tradición de heurística y sesgo, la descripción del razonamiento hecho por la psicología evolucionista es neutral con respecto al debate entre el internismo y el externismo en teoría de la justificación, en el sentido de que la descripción de la mente como modular y las hipótesis de la detección del mentiroso y del razonamiento frecuentista pueden ser lógicamente compatibles, tanto con la noción internista de la justificación, como con la noción externista. Una de las conclusiones a las que he llegado en esta parte del libro

es que estas dos teorías psicológicas no pueden ser usadas para resolver el debate en teoría de la justificación presentado en el primer capítulo de este trabajo.

La tercera teoría del razonamiento revisada en este libro es la teoría dual del razonamiento. Diversos psicólogos cognitivos (Evans, 2003) y filósofos (Samuels, Stich y Bishop, 2002) han visto en la teoría dual de sistemas la teoría que puede integrar de manera adecuada las diversas teorías del razonamiento en pugna, esto es, parece que la teoría dual de sistemas se encuentra en el punto medio entre el optimismo defendido por la psicología evolucionista y el pesimismo proveniente de la tradición de heurística y sesgo. A grandes rasgos, la teoría dual sostiene que los diversos procesos de razonamiento están subsumidos en dos sistemas. Uno de estos sistemas, denominado S1, es un sistema rápido, automático, que el ser humano comparte con otros animales, inconsciente, etc. En contraste, el ser humano cuenta con otro sistema, denominado S2, que es lento, reflexivo, del cual el sujeto es consciente, muchas veces fundado en reglas, etc. En este libro muestro que los psicólogos han desarrollado al menos dos posturas de cómo la teoría dual de razonamiento se une con la racionalidad. Una de ellas —la defendida por Evans y Over— lleva a dos acercamientos normativos distintos de racionalidad, a saber, un acercamiento de tipo consecuencialista y un acercamiento de tipo deontológico; la otra postura —defendida por Stanovich y West— sugiere que esos dos modos de entender a la racionalidad pueden ser integrados en dos acercamientos consecuencialistas entendidos de modo pragmatista y biológico.

En el último capítulo de este trabajo, indago cuáles son los vínculos teóricos entre las dos nociones de justificación y los dos sistemas de razonamiento postulados por la teoría dual. Un primer intento por relacionar las nociones de justificación con la teoría dual es argumentando que los productos de S2 responden a la noción internista de justificación, mientras que los productos de S1 responden a la noción externista (Eraña, en proceso b). Una segunda manera de relacionar la teoría dual de sistemas con la noción de justificación es reduciendo la noción de justificación a procesos causados por S2. Eraña (en proceso a) defiende que la noción de justificación implica necesariamente el punto de vista reflexivo, por lo que sólo S2 podría ofrecer creencias justificadas

y racionales. Bajo esta perspectiva, el sistema S1 generalmente produciría creencias racionales pero no justificadas. Parte del capítulo VI me dedico a mostrar cómo estas dos propuestas no son viables. La propuesta que considero más adecuada es aquella en la que la noción externista por sí misma puede dar cuenta de los productos de los dos sistemas de razonamiento. Si mis argumentos son correctos, los productos de los procesos de razonamiento pertenecientes a los dos sistemas de razonamiento no pueden estar justificados por el criterio internista de la justificación ya sea por sí misma, o en conjunción de una teoría externista. ¿Cómo se relaciona esto con la idea sostenida por defensores de la teoría dual, de que debe haber dos teorías normativas para evaluar el razonamiento? Los defensores de la teoría dual de sistemas se han comprometido con una teoría dual de la normatividad para evaluar los dos sistemas de razonamiento. Sin embargo, no existe consenso con relación a cuáles deben ser las teorías normativas particulares que se deben defender. En tanto que la aceptación de dos sistemas de razonamiento en sí misma no implica lógicamente alguna distinción particular entre teorías normativas, analizo si es compatible la postura de que los *outputs* ofrecidos por la teoría dual están justificados, según criterios externistas, con las teorías normativas que han postulado los psicólogos de la teoría dual.

Por un lado, Evans y Over (1996) distinguen entre un acercamiento consecuencialista de tipo pragmatista que ellos denominan "racionalidad 1" y un tipo de acercamiento deontológico comprometido con la visión estándar de la racionalidad que ellos denominan "racionalidad 2". Estos dos tipos de racionalidad pueden ser compatibles con ciertos criterios externistas, pero no con cualquier tipo de externismo. Si se acepta que la teoría dual de sistemas en su nivel descriptivo debe ser evaluado por la racionalidad 1 y la racionalidad 2, postuladas por Evans y Over (1996), y se sostiene que los productos de los dos sistemas están justificados por la teoría externista, entonces se tiene que defender un externismo no veritista que haga una distinción entre dos niveles normativos. Por otro lado, la propuesta normativa que defienden Stanovich y West (2000, 2003) consiste en un consecuencialismo de tipo pragmatista que evalúa a S2 y un consecuencialismo de tipo biológico que evalúa a S1. Los consecuencialismos que estos

psicólogos postulan no son de tipo veritista, como los formulados por algunos epistemólogos interesados en la noción de justificación. De este modo, el tipo de externismo que se puede defender a partir de la teoría dual de sistemas es un externismo no veritista que, además, debe distinguir al menos dos niveles de justificación.

En suma, según lo revisado en el capítulo VI, los criterios externistas de justificación con el que están justificados tanto los productos de S1 como de S2 pueden ser compatibles con las teorías normativas que los defensores de la teoría dual han propuesto. Sin embargo, si se aceptan tales propuestas normativas y se desea conservar la afinidad teórica entre los criterios externistas y los dos sistemas de razonamiento postulados por la teoría dual, se tiene que sostener un tipo de externismo no veritista y la idea de que la noción de justificación debe distinguir al menos dos niveles de justificación.

Así, a lo largo de este trabajo he explorado diversas maneras de entender la relación entre racionalidad y la noción de justificación *vía* distintas teorías del razonamiento. El supuesto naturalista de que las explicaciones filosóficas deben ser compatibles así como estar apoyadas en la evidencia empírica, conduce al problema de cuál es la teoría científica en el campo de la psicología del razonamiento que brinda la evidencia empírica más sólida. Según lo revisado en esta exposición, la teoría dual de sistemas parece ser la mejor teoría empírica con que de momento se cuenta, en tanto que *i*) es el punto medio entre la tradición de heurística y sesgo y la psicología evolucionista; *ii*) ofrece una respuesta que lidia con las interpretaciones optimistas y pesimistas de la racionalidad humana, y *iii*) los compromisos normativos que son asumidos por esta teoría no se reducen a una única teoría de la racionalidad. Esta última característica es importante si se tiene en cuenta que la visión estándar de la racionalidad se enfrenta a diversos problemas. Si la teoría dual de sistemas es la mejor teoría del razonamiento humano y si mi argumentación en el último capítulo es correcta, entonces esta teoría en su nivel descriptivo sugiere fuertemente que la noción de justificación debe ser entendida de manera externista. Esta teoría de la justificación, si se aceptan los postulados normativos de los defensores de la teoría dual, no es de corte confiabilista *à la* Goldman, sino una teoría externista no veritista. El esbozo de la teoría externista cognitivista (EC) es un

primer intento por fundar una teoría externista de la justificación en la teoría dual de sistemas, tanto en un nivel descriptivo como normativo. Esta teoría tiene la virtud de tomar en cuenta los estudios psicológicos desarrollados hoy en día y ofrece una explicación clara de la relación entre la justificación y la racionalidad.

En este trabajo he partido de dos grandes supuestos epistemológicos. El primero de ellos es que la noción de racionalidad y la de noción justificación están estrechamente unidas, en el sentido de que una creencia justificada para un sujeto *S* es una creencia que para *S* es racional creer. El segundo de los supuesto del que parto es una perspectiva naturalizada, entendiendo a ésta como la idea de que la ciencias cognitivas, y en especial la psicología cognitiva del razonamiento, pueden explicar y echar luz sobre las nociones epistemológicas, sobre los métodos y modos en que se adquiere el conocimiento. Estos dos supuestos se han mezclado a lo largo de la exposición.

Los supuestos naturalistas en epistemología se pueden entender de diversas maneras. Aun así, he conservado una "perspectiva débil" de naturalismo, que puede ser contrastada con aquella "perspectiva fuerte" en donde los problemas filosóficos centrales pueden ser resueltos por las ciencias cognitivas. Bajo esta perspectiva, se podría creer que la psicología cognitiva puede por sí misma brindar una noción de justificación. Si bien comparto la preocupación que motiva un principio naturalizado tan fuerte, no acepto que la psicología cognitiva del razonamiento por sí misma puede responder a los problemas centrales de la epistemología, dentro del cual se encuentra la noción de justificación. Como señalo desde el primer capítulo, gran parte de este trabajo está motivado por la búsqueda de una teoría del razonamiento que pueda dar respuesta al aparente *impasse* que actualmente existe en el análisis filosófico en torno a la noción de justificación. Muy a grandes rasgos, la pregunta a responder se podría plantear como "¿pueden las teorías del razonamiento ayudar a salir del debate actual en torno a la noción de justificación epistémica?". El análisis hecho a lo largo de este trabajo muestra que no hay una respuesta inmediata u obvia a la anterior pregunta. En primer lugar, porque en la psicología cognitiva del razonamiento no se ha establecido un consenso total acerca de cuál de las distintas teorías en pugna es la más adecuada. De momento, por ejemplo, todo parece

indicar que la teoría dual de sistemas puede convertirse en un nuevo paradigma en la psicología del razonamiento, pero esta teoría todavía se enfrenta con algunos problemas empíricos y de claridad conceptual que podrían frenar el apoyo a esta teoría. En segundo lugar, las conexiones y distinciones hechas en epistemología contemporánea no siempre tienen correlatos en la psicología cognitiva. Por ejemplo, la psicología del razonamiento no parece apoyar una noción fundacionista de la justificación, y ni siquiera puede sugerir si los términos en que se ha dado el debate entre fundacionismo y coherentismo son los adecuados.

A pesar de este tipo de consideraciones, en este trabajo muestro que las teorías del razonamiento ofrecen una explicación a los procesos y mecanismos de razonamiento, y sus supuesto normativos —en donde muchas veces entran en juego consideraciones filosóficas— proporcionan la base de lo que constituye una creencia racional. Una vez que se tienen estos *datos* se puede explorar cuáles teorías de la justificación se acomodan o se apoyan en dichos datos. Los vínculos que uno puede hacer entre razonamiento, racionalidad y la noción de justificación se asemejan a una búsqueda por encontrar las conexiones más coherentes entre estas posturas. Si lo anterior es correcto y lo argumentado en el último capítulo también lo es, entonces los criterios con los que están justificadas las creencias productos de los distintos sistemas de razonamiento son de tipo externista. Si además se considera que la teoría dual de sistemas en su nivel descriptivo debe aceptar las teorías normativas postuladas por los defensores de la teoría dual, entonces se tiene que aceptar un externismo particular, a saber, un externismo no veritista que distinga dos niveles dentro de la justificación.

NOTAS

INTRODUCCIÓN

1 La definición tradicional del conocimiento, proveniente del *Tetetes* de Platón, sostiene que un sujeto *S* sabe que *p*, si y sólo si *S* cree que *p*, "p" es verdadera y *S* está justificado o cree racionalmente que *p*.

2 El gran número de literatura generada a partir de los contraejemplos tipo Gettier señala que las condiciones con las que se definió tradicionalmente el conocimiento no son suficientes para brindar una definición adecuada (Dancy, 1986, Villoro, 1982, Sosa, 1994). Dada la manera en que se presentaron estos contraejemplos se consideró que el problema principal radicaba en la noción de justificación, por lo que, por un lado, se ofrecieron definiciones alternativas de conocimiento enfocadas a reforzar el tipo de justificación que requieren las creencias verdaderas para ser conocimiento (Lehrer, 1965, Clarck, 1963, Goldman, 1978) y, por otro lado, se inició un estudio exhaustivo de la noción misma de justificación. Para un modo distinto de entender los contraejemplos tipo Gettier véase Tomasini (2001).

3 Expondré las críticas a la visión estándar de la racionalidad y al acercamiento deontológico de la racionalidad en el capítulo IV.

4 Al final del primer capítulo explicaré con cierto detalle en qué consiste cada una de estas intuiciones epistémicas, así como el motivo por el cual estas intuiciones conducen al estudio de la racionalidad y el razonamiento.

5 Evans (1991), por ejemplo, sostiene que existen tres distintos objetos de estudio empíricos en el razonamiento humano: *a*) explicar la competencia deductiva de razonamiento, *b*) explicar los distintos errores y sesgos en los estudios de razonamiento, y *c*) explicar la dependencia del razonamiento con respecto a su contenido y su contexto.

6 En los capítulos II, III y V de este libro se hacen explícitos los presupuestos normativos de diversas teorías del razonamiento.

7 Muchas veces la diferencia entre dos teorías del razonamiento en torno a uno de estos aspectos va acompañado de (o conduce a) la diferencia en otro aspecto. Por ejemplo, una teoría del razonamiento puede diferir de otra con respecto al tipo de fenómeno psicológico que debe

explicar, en parte porque no comparten los mismo supuestos metodológicos (Elio, 2002).

8 Este modo de entender al naturalismo se podría ubicar dentro de lo que Kim (1994) ha denominado la epistemología "centrada en torno a la justificación" —en donde ubica propuestas como la de Kornblith o Goldman, por ejemplo— para distinguirla de proyectos teóricos más radicales como el de Quine.

9 En el capítulo V expongo parte de los argumentos para apoyar esta idea.

10 En el capítulo V expondré con cuidado lo que aquí se hace de manera breve.

11 Por motivos de claridad evitaré presentar aquí los argumentos para apoyar esta tesis y me limitaré únicamente a explicarla brevemente.

12 En el capítulo V se exponen las diferencias entre estas propuestas.

13 En el capítulo IV haré una distinción entre tipos de consecuencialismo: veritista, biológico y pragmatista.

I. DE LA NOCIÓN DE JUSTIFICACIÓN AL ESTUDIO DE LA RAZÓN

1 En otras palabras, la definición recursiva de la justificación toma al inciso "a)" de la definición de fundacionismo como cláusula base y al inciso "b)" como cláusula recursiva.

2 La concepción holista o sistémica no debe entenderse como una postura bidireccional. Una postura bidireccional de la justificación, a diferencia de una postura unidireccional, sostiene que una creencia puede estar justificada por un grupo de creencias, pero a su vez, esta creencia junto con otras puede justificar a aquellas creencias de las que ella misma había sido justificada.

3 Generalmente la justificación de una creencia se debe al apoyo de un conjunto de creencias, de modo que una formulación más precisa de la primera premisa del argumento del regreso de las justificaciones sería "la creencia *p* está justificada si se ofrece un conjunto de creencias *q*, *q'*, *q''*, etc. que la fundamenten". Sin embargo, por motivos de claridad supondré que una creencia puede estar justificada por una única creencia.

4 Una versión de este argumento es presentado por Davidson (1983, pp. 471-489) quien a grandes rasgos sostiene que para que un sujeto pueda considerar a una creencia como básica necesita de una creencia *extra*, a saber, aquella creencia que le indica que la creencia básica de hecho es una creencia básica. Sin embargo, con esta creencia *extra* la creencia básica dejaría de serlo, pues no estaría justificada sin apelar a otras creencias.

5 En contraste, el internismo sostiene *grosso modo* que el sujeto epistémico debe tener acceso a las condiciones internas que determinan si sus creencias están justificadas (Steup, 2005). Más adelante caracterizaré con mayor detalle al externismo y al internismo. Para una revisión general sobre el debate entre el internismo y el externismo véase Kornblith (2000), Steup (2005) y Pappas (2005).

6 La distinción que separa el *origen* o *causa* del conocimiento del cómo se *justifica* o *valida* es una distinción clásica que se puede rastrear hasta la obra kantiana. Recuérdese que en la *Crítica de la razón pura* —y más precisamente en la primera sección de la "Deducción trascendental de los conceptos"— Kant presenta su famosa distinción entre la cuestión de derecho (*quid iuris*) y las cuestión de hecho (*quid facti*) como dos tareas epistemológicas distintas.

7 Pappas (2005) sostiene que en epistemología se ha defendido también un internismo que él denomina deontológico. El internismo deontológico, según este filósofo, sostiene que una creencia justificada es aquella que satisface los deberes o reglas epistemológicas adecuadas. Es importante hacer esta distinción porque más adelante (capítulo IV) abordaré un acercamiento a la racionalidad conocido como deontológico, que deseo distinguir de la noción internista de justificación.

8 De las críticas anteriores, Goldman afirma que el estudio de las capacidades de razonamiento son importantes para saber hasta qué medida se pueden alcanzar los ideales epistemológicos propuestos por los filósofos. De ahí, Goldman sugiere que las "ciencias cognitivas podrían ser relevantes para establecer los estándares de racionalidad..." (Goldman, 1993, p. 22, la traducción es mía). En el capítulo IV abordaré este tipo de propuestas.

9 En general hago uso de la expresión *proceso cognitivo adecuado* o *epistemológicamente adecuado* para no identificar a las diferentes propuestas externistas con el confiabilismo de Goldman. Un proceso adecuado podría ser un proceso confiable *à la* Goldman, pero también podría ser un proceso que nomológicamente conecta al mundo con las creencias de los sujetos como sostiene Armstrong (Kornblith, 2000), o una facultad "virtuosa" como sostiene Sosa (1992).

10 Si bien existen distintas versiones de confiabilismo, la revisada en este libro es —como ya he mencionado— la posición defendida por Alvin Goldman (1979, 1986) que modifica (de modo parcial) posteriormente (1992).

11 De hecho, Kornblith (1994) considera que en este punto la epistemología de Goldman tiene más coincidencias con la epistemología tradicional que con la epistemología naturalizada. A este respecto Kornblith afirma:

> El rompimiento de Goldman con la epistemología tradicional en *Epistemology and Cognition*, y los rasgos de la descripción que lo hacen una variedad del naturalismo, no deben ser encontrados aquí. Más bien, éstos son encontrados en la insistencia de Goldman de que eso no es toda la epistemología (Kornblith, 1994, p. 260, la traducción es mía).

12 Esto sugiere una noción gradual de justificación, pues es posible que la confiabilidad de un proceso cognitivo pueda ser mayor o menor a partir del número de verdades que produzca.

13 Nótese la similitud entre esta definición y la definición recursiva de la justificación emparentada con el fundacionismo.

14 Los contraejemplos que se presentan en esta sección son versiones de los expuestos en Bonjour (1980). Lehrer (2000) ofrece otros ejemplos que ponen en duda la noción externista de justificación.

15 Esta dificultad está relacionada con el denominado "problema de la generalidad". Véase Conee y Feldman (1998).

16 Generalmente se considera que la confiabilidad de un proceso cognitivo, como en el caso del confiabilismo de Goldman, consiste en que tal proceso conduce a la verdad. Sin embargo, no quiero comprometerme con tal postura, pues las teorías externistas no tienen que necesariamente aceptar proyectos de tipo veritistas. En el capítulo VI esbozo un tipo de teoría externista no veritista fundado en una teoría psicológica particular.

17 Según lo revisado anteriormente, para un internista como Bonjour cuando un sujeto está justificado, éste *debe creer* justificadamente que lo está. En cambio, para un internista que sostenga que el acceso se puede entender como accesibilidad (Pappas, 2005) tendría que enunciarlo como "cuando un sujeto está justificado, éste *puede creer* justificadamente que lo está". La caracterización que generalmente usaré para definir al internismo será la de Bonjour, teniendo en mente que la noción de acceso puede entenderse también como accesibilidad.

18 Eraña (en proceso a, y en comunicación personal) ha sostenido, en el plano epistemológico, que el debate entre las nociones internista y externista de la justificación puede entenderse como un debate entre una noción de justificación personal (o subjetiva) y una noción de justificación en tercera persona (u objetiva). Considero que esta distinción entre justificación en primera y tercera persona puede hacerse tanto en la justificación epistemológica como en la justificación moral.

II. LA TRADICIÓN DE HEURÍSTICA Y SESGO
Y LA RACIONALIDAD

1 Las relaciones entre cada una de estas maneras de estudiar a la razón serán revisada posteriormente; por el momento, basta indicar que existen implicaciones entre cada una de estas tareas.

2 La idea de que los estándares normativos tienen su origen en el modo que los seres humanos de hecho razonan puede ser entendido como un acercamiento psicologista a la racionalidad (Elio, 2002). Dos acercamientos a la racionalidad pueden ser contrastados con el psicologismo. Uno de ellos sería defender un acercamiento en el que las normas de racionalidad se encuentran en un reino de verdaderas fuera del ámbito humano. Este segundo acercamiento es denominado por Elio (2002) como "platonismo". Un acercamiento diferente a la racionalidad sería aquel que sostuviese que los estándares normativos pueden ser entendidos como invenciones que pueden ser usadas para producir propósitos particulares. Este último acercamiento es denominado por Elio (2002) como "constructivismo".

3 Algunos psicólogos evolucionistas han argumentado que estos resultados pueden ser explicados (esta explicación se encuentra en al próximo capítulo). Para una exposición de cómo se ha tratado de explicar el éxito que los sujetos tienen con algunos tipos de tarea de selección, mientras que presentan dificultades con versiones de la tarea de selección como la recién presentada véase Evans (1991) y Stein (1996).

4 Posteriormente presentaré los vínculos que esta teoría descriptiva tiene con una teoría evaluativa de la razón y una teoría normativa de la racionalidad.

5 Véase Gilovich y Griffin (2002), y Kahneman (2002) para conocer el trabajo que este grupo de psicólogos ha llevado a cabo durante más de dos décadas.

6 En la siguiente sección analizaré la distinción entre competencia y ejecución de razonamiento.

7 Diferentes psicólogos han tratado de clasificar las reacciones que se han tenido frente a la evidencia empírica que, muchos han sostenido, apoya a una versión pesimista de la racionalidad humana. Evans (1993), por ejemplo, sostiene que los argumentos en contra de esta interpretación se dividen en aquellos que señalan *i.)* que existe un problema con el sistema normativo, ya que es posible que el sujeto del experimento esté razonando por medio de algún sistema distinto al que el experimentador está razonando; *ii.)* que existe un problema de interpretación, pues es posible, señala esta crítica, que el sujeto del experimento malinterprete la prueba; y, *iii.)* que existe un problema de "validación externa", en el sentido de que los psicólogos apoyan

una visión pesimista de la racionalidad humana porque los estudios de sesgo responden a una "moda" en psicología cognitiva y a una forma en que los investigadores pueden ganar renombre profesional, además esta crítica sugiere que los estudios de la tradición de heurística y sesgo se fundan en experimentos artificiales y poco representativos del razonamiento humano. En los capítulos III y IV analizaré las diversas reacciones que se han tenido en contra de la interpretación pesimista de la racionalidad humana.

8 "El hablante está constreñido por sus intenciones, su conocimiento del mundo y del interlocutor, sus estados mentales, sus limitaciones atencionales y de memoria, el contexto, etc. Es decir, que la actuación verbal depende tanto de la competencia como de un conjunto de variables psicológicas" (de Vega, 1998, p. 30).

9 Después de una minuciosa explicación de la distinción entre la competencia-ejecución en la lingüística y el razonamiento, en *Without Good Reason*(1996), Stein presente cuáles son las diferencias importantes entre estas distinciones.

10 Cohen combina varias críticas en sus escritos, es decir, piensa que no todos los experimentos pueden ser reducidos a meros errores de ejecución. (Cohen, 1981, p. 323). En este capítulo, como señalé anteriormente, me centraré sólo en aquella postura que defiende Cohen en la que los experimentos en psicología cognitiva sólo muestran errores de ejecución.

11 Stein ejemplifica esta postura con los experimentos hechos por los psicólogos evolucionistas, en los que se muestra que los sujetos pueden resolver tareas de selección de manera adecuada, aun cuando el marco en donde quedan descritas dichas tareas de selección no son nada familiares a los sujetos (Stein, 1996, p. 86 ss).

12 Para apoyar esta posición, Stein expone el trabajo de Manktelow y Evans, en donde se presenta a los sujetos cuatro cartas donde se hace referencia a los que cada experimentador come y bebe (las cartas son "comer puerco", "comer pescado", "tomar vino tinto" y "tomar cerveza"), y la regla que tienen que probar es "Si como cerdo, entonces tomo vino tinto". Los resultados de este experimento son muy similares a las versiones abstractas de la tarea de selección presentadas por Wason (Stein, 1996, p. 88).

13 En otro trabajo Piatelli-Palmarini afirma:

> debemos desconfiar siempre de las interpretaciones estadísticas hechas "a primera vista", debemos desconfiar de los cálculos de probabilidades nuestros y ajenos, a menos que hayan sido elaborados muy cuidadosamente por expertos *en ciencias estadísticas*. Fíjense bien, no estoy simplemente (y superficialmente) alertando

> contra el uso mistificador de las estadísticas, a veces por parte de autoridades no del todo autorizadas, o de expertos intencionadamente manipuladores. Hay algo peor. Estoy poniendo en guardia *también* ante nuestra lectura intuitiva ingenua de estadísticas inapelables, a prueba de bomba. Existen numerosos ejemplos, y hemos aportado aquí algunos, de conclusiones equivocadas que todos sacamos, *con absoluta buena fe*, a partir de datos estadísticos perfectamente fiables (Piatelli-Palmarini, 2005, p. 132).

Algunos filósofos creen que una postura tan radical como la que sostiene Piatelli-Palmarini nunca es defendida por la tradición de heurística y sesgo (Samuels y Stich, 2004), por lo que se ha argumentado que estas posturas son parte de los excesos retóricos que no representan las afirmaciones centrales de dicha tradición.

14 La visión estándar de la racionalidad, que sostienen los defensores de la interpretación pesimista, tiene tres virtudes importantes, *i.)* provee un criterio preciso con el cual se puede evaluar el razonamiento humano; *ii.)* es coherente con la intuición de que la lógica y la teoría de la probabilidad tienen una estrecha relación con cuestiones acerca de cómo se debe razonar; y *iii.)* si los principios de razonamiento son derivados de las matemáticas y la lógica, éstos parecen ser los principios que se deben adoptar para capturar la idea de que las normas de razonamiento son principios universales (Gigerenzer, 1993).

15 La irracionalidad depende de la aceptación de un estándar particular de racionalidad, que puede ser puesto en duda. En el capítulo IV exploraré un estándar normativo alternativo al aquí expuesto. Sin embargo, de momento supondré que la visión estándar de la racionalidad es un criterio adecuado para evaluar el razonamiento humano.

16 Eraña (en proceso a) ha puesto en duda que la noción de justificación y racionalidad se encuentren tan íntimamente ligadas. Su propuesta es que toda creencia justificada es racional, pero que hay creencias racionales que no están justificadas. En el capítulo VI de este trabajo examinaré esta propuesta y ofreceré razones para rechazarla.

III. LA PSICOLOGÍA EVOLUCIONISTA Y LA RACIONALIDAD

1 Un teórico emparentado muchas veces con Cosmides y Tooby es Gigerenzer. Si bien Gigerenzer y su grupo de investigación nunca se han definido a sí mismos como psicólogos evolucionistas, gran parte de sus trabajos —principalmente aquel realizado en los años noventa— comparten muchas afinidades. Sin embargo, hay tesis importantes de Cosmides y Tooby que Gigerenzer y su grupo no han aceptado. En su momento indicaré cuáles son las ideas que Gigerenzer no

comparte con Cosmides y Tooby, pero muchas veces citaré su trabajo como apoyo a la psicología evolucionista.

2 A este respecto Tooby y Cosmides afirman:

> la razón por la cual tenemos un conjunto de circuitos (neuronales), en lugar de otro, es que el circuito que tenemos fue mejor para resolver los problemas a los que nuestros ancestros enfrentaron durante la historia evolutiva de nuestra especie (Cosmides y Tooby, 1997, la traducción es mía).

3 Gigerenzer no acepta la hipótesis de la modularidad masiva, tal y como lo hacen Cosmides, Tooby, Sperber, Buss u otros psicólogos evolucionistas. Sin embargo, muchas veces su trabajo parece acercarse mucho a la visión modular de la mente, en tanto que las heurísticas rápidas y frugales que postula este psicólogo comparten algunas características con la noción de módulo de los psicólogos evolucionistas (Gigerenzer y Todd, 1999).

4 Es posible distinguir tres maneras de entender el computacionalismo (Horst, 2008): *a)* sostener que la mente literalmente es un computadora y que el pensamiento es un tipo de cómputo; *b)* sostener que algunos procesos mentales pueden ser modelados por técnicas de modelaje computacional; y *c)* pensar a la mente como una computadora puede ser una metáfora que nos sirve para entender ciertos procesamientos de información. Los psicólogos evolucionistas no son del todo claros con relación al tipo de computacionalismo que aceptan, pero generalmente sus posturas se ubicarían entre *a)* y *b)*.

5 En esta cita aparece otro rasgo importante de la noción de módulo, a saber, la especificidad de dominio que será explicado posteriormente.

6 De Vega (1998) sostiene que uno de los rasgos que distinguen a la psicología cognitiva de la psicología conductista es el computacionalismo con la que los cognitivistas se comprometen.

7 Nótese que lo que defienden los psicólogos evolucionistas es que la estructura o arquitectura de la mente es innata, tesis que podría ser defendida por un chomskyano. Pero de ahí, no se sigue que los psicólogos evolucionistas acepten la noción chomskyana de módulo. Si bien los psicólogos evolucionistas podrían aceptar que la mente cuenta con módulos chomskyanos, la noción de módulo que los psicólogos evolucionistas desean defender es principalmente la idea de un módulo como mecanismo. Para conocer la diferencia entre la noción chomskyana y mecanicista de módulo véase Carruthers (2006).

8 Muchas veces se ha definido a la especificidad de dominio como una restricción en el tipo de contenido que puede ser tomado como insumo. Esta noción de especificidad de dominio, según Carruthers (2006), es la manera "fodoriana" de definir especificidad de dominio,

pero no la manera en que los psicólogos evolucionistas lo definen. Para este filósofo, los psicólogos evolucionistas entienden por especificidad de dominio la función que los módulos pueden realizar (Carruthers, 2006, p. 19). Distinciones entre la especificidad de dominio (dominio actual, dominio propio y dominio cultual) se encuentran en Sperber (1994).

9 Por "informacionalmente encapsulado" se entiende que otras partes de la mente no pueden influir o tener acceso al trabajo interno de un módulo solamente a sus salidas o *outputs*. A su vez, los módulos sólo tienen acceso, según Fodor, a etapas de procesamiento de niveles inferiores y no al procesamiento de información de niveles superiores, esto es, lo que la mente sabe y cree no puede afectar el trabajo de los módulos. Este rasgo de la modularidad ha sido fuertemente criticado por algunos psicólogos, puesto que niega la importancia del desarrollo cognitivo (Kamiloff-Smith, 1992). Para conocer diversas maneras de entender la noción de encapsulamiento véase Carruthers (2006).

10 Al sostenerse la tesis de la modularidad masiva los psicólogos evolucionistas modifican la noción fodoriana de módulo (Carruthers, 2006).

11 Es preciso indicar que dentro de la psicología evolucionista no siempre se han compartido los mismos criterios para considerar que algo cuenta como "evidencia" para la existencia de un módulo cognitivo. Un ejemplo de ello, es el debate entre Cosmides y Tooby (1992) y Sperber y Girotto (2003).

12 Stanovich y West (2000) denominan a los defensores de la interpretación optimista de la racionalidad emparentados con la psicología evolucionista, los "panglosianos". En clara referencia al doctor Pangloss, de la novela *Cándido,* de Voltaire, quien afirmaba que "todo sucede para bien".

13 Cosmides y Tooby reconocen que el primer psicólogo en proponer dicha hipótesis fue Gigerenzer (1994).

14 Samuels, Stich y Bishop (2004) ilustran lo anterior haciendo una analogía con una calculadora electrónica. Una calculadora electrónica estándar está diseñada para recibir como insumos problemas matemáticos que están presentados en una notación con base 10. Si la calculadora está bien diseñada, la calculadora tendrá un mecanismo confiable que resolverá problemas que estén formulados en el formato adecuado, esto es, que estén formulados en una notación arábiga con base 10. Si se desea usar la calculadora para resolver problemas establecidos en numerales romanos, debido a que simplemente no hay botones para "X", "V", "L", etc., no habría manera de que la calculadora lidie con ese tipo de problemas. La única manera de resolverlos adecuadamente sería traduciendo dichos problemas a una notación

arábiga (Samuels, Stich y Bishop, 2004). Así como el buen trabajo de la calculadora dependerá de la manera en que el problema sea presentado, siendo bueno en el caso de la notación arábiga, en el caso de la mente humana, el buen trabajo de los mecanismos de razonamiento probabilístico dependerá del formato en que se provee la información presentada, teniendo buenos resultados al procesar formatos en modo de frecuencia.

15 En el capítulo anterior distinguí dos versiones de la interpretación pesimista de la racionalidad humana, por lo que un análisis adecuado tiene que tomar en cuenta al menos estas dos versiones de la interpretación pesimista. Más adelante se lleva a cabo este análisis.

16 Dawkins (1985) presenta algunos comentarios críticos a la postura de Trivers.

17 Los esquemas de razonamiento pragmático, según sus defensores, son estructuras de conocimiento abstracto que los seres humanos usan para razonar. Dichos esquemas son inducidos por las experiencias de la vida cotidiana tales como "obligaciones" y los "permisos" (Cheng y Holyoak, 1985, p. 395).

18 Se ha llegado a sostener que una postura tan optimista de la racionalidad humana no es parte de las afirmaciones centrales de la psicología evolucionista, por lo que este tipo de afirmaciones hechas por los psicólogos evolucionistas deben verse como un exceso retórico (Samuels, Stich y Bishop, 2002).

> Mucha de la disputa entre los psicólogos evolucionistas y los defensores de la tradición de heurística y sesgo es en sí mismo una ilusión. Los "fuegos artificiales" generados por cada grupo de psicólogos al concentrarse en los excesos retóricos del bando contrario ha distraído la atención de lo que realmente están afirmando (Samuels, Stich y Bishop, 2002, p. 237, la traducción es mía).

19 Como mostraré más adelante, los psicólogos evolucionistas están aceptando implícitamente, al igual que los defensores de la tradición de heurística y sesgo, la visión estándar de la racionalidad.

20 La numeración del trabajo de Samuels, Stich y Faucher proviene del documento electrónico que puede ser descargado en formato .pdf del sitio: http://ruccs.rutgers.edu/ArchiveFolder/Research%20Group/Publications/pubs.html.

21 Una manera de conciliar la postura pesimista (I), defendida por la tradición de heurística y sesgo, con la interpretación optimista (A), defendida por los psicólogos evolucionistas, es postulando teorías duales de procesamiento. Revisaré estas teorías en el capítulo V de este trabajo. En dicho capítulo defenderé que la teoría dual de sistemas es

la teoría que actualmente explica de mejor de manera el razonamiento humano.

22 Algunas veces los defensores mismos de la tradición de heurística y sesgo han sostenido que (al menos en parte) la psicología evolucionista es compatible con su proyecto teórico:

> Aun una de las metáforas más populares del acercamiento evolucionista con relación al razonamiento —aquél en donde la mente es como una navaja suiza— es enteramente consistente con el acercamiento de la tradición de heurística y sesgo (Gilovich y Griffin, 2002, p. 10, la traducción es mía).

23 Es posible que el internista argumente, en contra de esta crítica, que la noción de justificación no tiene relación alguna con el modo en que las creencias son originadas o causadas, sino con la relación lógica entre las creencias (Davidson, 1983). Considero que si bien dentro de la visión internista de justificación el origen de las creencias no es importante, sí es importante para el internismo el control consciente que los sujetos deben tener –en gran parte en esto consiste la responsabilidad epistémica. Por lo que esta crítica sí sería pertinente para un internista. En el capítulo VI ofreceré varios argumentos en contra del internismo.

24 Por otro lado, los argumentos de Carruthers (2006) a favor de un tipo de encapsulamiento e inaccesibilidad como parte de la noción de módulo defendida por los teóricos de la modularidad masiva han sido puestos en duda por otros filósofos Samuels (2006) y Prinz (2006).

IV. LA INTERPRETACIÓN PESIMISTA DE LA EVIDENCIA EMPÍRICA Y LAS VISIONES DE LA RACIONALIDAD

1 Las opciones "f)" y "h)" son tomadas de la presentación del problema de razonamiento "Linda, la cajera feminista" expuesta en el capítulo II de este libro.

2 Gigerenzer (1996), al igual que Adler, sostiene que los defensores de la tradición de heurística y sesgo imponen principios estadísticos como normas sin examinar su contenido, es decir, sin indagar cómo es que los sujetos entienden las tareas que se les solicita ejecutar en un experimento. Este psicólogo sostiene que no se puede presuponer que los sujetos entienden las distintas tareas experimentales de la misma manera en que los psicólogos del razonamiento lo han hecho. Como sostendría Adler, Gigerenzer afirma que no se puede asumir que en el caso de "Linda, la cajera feminista" los sujetos entiendan el término "probable" con exactamente el mismo significado que lo tiene dentro del cálculo de probabilidades o el conyunto "y" como un operador

lógico " ^ ". Dependiendo del contexto, estas palabras pueden ser entendidas de manera distinta; "probable", por ejemplo, puede significar plausible, verosímil, que aparenta ser verdadera o que dada la presente evidencia es razonable esperar que suceda.

> Una variedad de inferencias semánticas y pragmáticas son esbozadas a partir de las instrucciones en el problema de Linda, y esas inferencias afectan radicalmente la interpretación de las instrucciones... Por ejemplo, algunas [personas] parecen esbozar que "T $\rightarrow$ T & no F", esto es, "Linda es cajera en un banco" implica que "Linda es cajera en un banco y no es activista en un movimiento feminista" (Gigerenzer y Regier, 1996, p. 25, la traducción es mía).

Si este es el caso, de acuerdo con Gigerenzer, de los experimentos de razonamiento no se podría concluir que los sujetos cometen sistemáticos errores de razonamiento, puesto que la manera en que los sujetos entienden las pruebas no implicaría que los principios normativos están siendo violados.

3 Algunos filósofos creen que uno de los debates centrales entre los defensores de la tradición de heurística y sesgo, y la psicología evolucionista es precisamente decidir cuál es el modo adecuado en el que se debe interpretar la teoría de la probabilidad (Samuels, Stich y Bishop, 2002).

4 Una alternativa para responder a estas preguntas en una visión psicologista de la lógica. Para una defensa de la misma véase Jaquette (2003).

5 Una idea similar es defendida, en el medio filosófico de habla hispana, por Alejandro Tomasini, quien sostiene que no creemos todo aquello que es inferido de nuestras creencias.

> Decir que Juan cree una proposición implicada por *p* es simplemente decir que *formalmente* la acepta. Pero no se sigue de ello que Juan *haga uso* de ella, que de hecho la emplee o la aplique. La proposición *p* que Juan cree puede implicar *q, r, s*, etc., pero no se sigue que Juan crea *q, r, s* y demás (Tomasini, 2001, p. 43).

Estas ideas son usadas por Tomasini (2001) para mostrar que, en sentido estricto, los contraejemplos tipo Gettier no ponen en duda la noción tradicional del conocimiento.

6 Nótese la estrecha relación de este acercamiento a la racionalidad con lo que Pappas (2005) denomina "internismo deontológico" (capítulo I sección "Internismo y externismo"). Para este filósofo, el internismo deontológico es aquel que considera que una creencia está justificada si satisface los deberes epistémicos adecuados. Sin embargo, esto no significa que cualquier tipo de internismo se compromete necesariamente con un acercamiento deontológico. En las conclusiones de este

capítulo mostraré cómo una postura internista podría defender un acercamiento consecuencialista.

7 Un acercamiento similar en ética parece defender Skyrms cuando pone de manifiesto la relación que hay entre los estudios en biología y evolución con relación a otras disciplinas filosóficas.

> ¿Qué tiene que ver todo esto con la ética y la filosofía política?... No he dicho nada acerca de cómo los seres humanos deberían vivir sus vidas o cómo la sociedad debería organizarse. La ética es un estudio de las posibilidades de cómo uno podría vivir. La filosofía política es el estudio de las posibilidades de cómo las sociedades podrían organizarse. Si la posibilidad es construida generosamente, tendríamos una teoría utópica. Aquellos que lidian con "los hombres como son" necesitan trabajar con un sentido más restrictivo de posibilidad... Aun aquellos que buscan cambiar al mundo deben primero aprender a cómo describirlo (Skyrms, citado en Danielson, 2004, p. 436, la traducción es mía).

8 Gigerenzer y Selten (2002) han distinguido entre diversas nociones de racionalidad acotada, en parte, porque sostienen que muchas veces erróneamente se considera que la racionalidad dependiente de recursos es equivalente a la idea de racionalidad como optimización bajo restricciones.

9 Este argumento es defendido también por Goldman (1993). Véase la sección "Internismo y externismo" del capítulo I de este libro.

10 La racionalidad dependiente de recursos precisamente hace énfasis en que los estándares de razonamiento no pueden dejar de lado las capacidades cognitivas y el medio ambiente (Gigerenzer y Selten, 2002). Una propuesta similar es la defendida por Goldman (1993) y Jaquette (2003).

11 La teoría dual de sistemas, como argumentaré en el próximo capítulo, puede ser vista como ese entramado empírico y conceptual que integra, no sólo parte de la evidencia empírica proveniente de la tradición de heurística y sesgo, sino de distintas teorías normativas con las cuales debe ser evaluado el razonamiento humano.

12 La versión original se encuentran en Goldman 1986, p. 84.

13 Esto es lo que Bonjour denomina "la metajustificación de una teoría de la justificación". Véase capítulo I sección "Internismo y externismo".

V. LA TEORÍA DUAL DE SISTEMAS

1 Las versiones compatibles son la versión pesimista (I): muchas veces en que los juicios se desvían de las normas apropiadas de racionalidad son explicadas por el hecho de que los juicios y decisiones hechos por

los sujetos se fundan en heurísticas que algunas veces conducen a sistemáticos errores de razonamiento; y, la versión optimista (A): muchos de los casos en los cuales los juicios de los sujetos cognoscentes concuerdan con las normas de racionalidad son explicadas por el hecho de que los seres humanos activan módulos darwinianos que hacen un buen trabajo cuando son provistos de los insumos adecuados.

2 En los últimos años, el mismo Kahneman (2002) ha defendido una postura similar a la teoría dual de razonamiento, en la que se distinguen aquellos mecanismos cognitivos reflexivos de aquellos mecanismos automáticos de los que no se tiene control.

3 El trabajo de Gigerenzer y Regier (1996) pone en duda esta sugerencia.

4 La idea de que la teoría dual de sistemas tiene como propósito resolver la aparente paradoja de la racionalidad es compartida por Stanovich y West (2000) en tanto estos psicólogos aceptan que la teoría dual es capaz de unir la separación existente entre las teorías descriptivas y normativas de la razón.

5 Este sistema se compone de diferentes mecanismos implícitos, por lo que muchas veces estos psicólogos hacen referencia al mismo en plural, esto es, "sistemas tácitos o implícitos". Posteriormente expondré una crítica a la falta de precisión con la que los defensores de la teoría dual han definido este sistema.

6 Como mostraré en la siguiente sección, estos procesos son pieza importante para la explicación de la evidencia usada por la tradición de heurística y sesgo para apoyar la idea de que los seres humanos cometen errores sistemáticos de razonamiento.

7 La idea de que los seres humanos comparten los procesos tácitos con otros animales, mientras que no ocurre lo mismo con los procesos explícitos se apoya en que estos procesos están íntimamente conectados con el lenguaje. Por ello, Evans y Over (1996, p. 154) consideran que es altamente probable que la posesión del lenguaje sea un requisito para los procesos reflexivos conscientes. Recientemente se ha puesto en duda que el sistema explícito sea exclusivo de los seres humanos (Evans, 2006).

8 Antes de *Rationality and Reasoning* el mismo Evans defendió teorías duales del razonamiento, en donde la relación entre los sistemas era paralela o secuencial; en una sección de dicho libro menciona cuáles son, a grandes rasgos, las diferencias entre estas posiciones (Evans y Over, 1996, p. 144 ss.). Para una propuesta particular de la relación entre los dos sistemas de razonamiento véase Carruthers (2009).

9 Al igual que Evans y Over, Stanovich y West afirman que este sistema podría consistir en diferentes subsistemas que podrían ser relativamente independientes (Stanovich y West, 2003, p. 8).

10 Stanovich y West sostienen que nombrar a uno de los sistema "S1" y al otro "S2" es un modo de defender una versión general de la teoría dual y no una postura particular de la misma (Stanovich y West, 2003, p. 8). Lo anterior parece subrayar que para estos psicólogos cognitivos existe un núcleo central de afirmaciones que comparten los defensores de la teoría dual.

11 La teoría dual del razonamiento se ha visto enriquecida por reveladoras ideas provenientes de la literatura psicológica en el aprendizaje implícito, donde también se ha defendido la existencia de un sistema explícito y otro implícito. Algunos defensores de la teoría dual citan el trabajo de Reber, quien sostiene que los procesos implícitos tienen menos variabilidad entre los individuos y son en gran medida independientes de la edad o el IQ. Reber, sostienen Evans y Over (1996), hace uso de estas ideas para defender lo que él denomina la primacía de la inconciencia cognitiva. Al igual que Stanovich y West, Reber afirma que los procesos tácitos evolucionaron anteriormente y que son compartidos con otros mamíferos superiores. Sin embargo, a diferencia de los psicólogos evolucionistas, los psicólogos defensores de la teoría dual de sistemas no sostienen que

> el pensamiento de nivel superior involucrado en el razonamiento y toma de decisiones pueda ser explicado de modo verosímil en términos de mecanismo evolutivos específicos (Evans y Over, 1996, p. 147, la traducción es mía).

12 La prueba SAT, denominada actualmente como "prueba de razonamiento SAT", es una prueba usada en los Estados Unidos para la admisión en universidades de dicho país. La prueba fue implementada a principios del siglo pasado, pero ha sido reformulada con el paso del tiempo. Actualmente consta de tres grandes secciones: matemáticas, lectura crítica y escritura.

13 Sloman considera que un sistema basado en reglas es aquel que sigue explícitamente una regla y no aquel que puede "entrar en conformidad" con ciertas reglas. Según Sloman, se debe aceptar la distinción hecha por varios psicólogos entre un

> sistema que *sigue* reglas de otro que simplemente *está en conformidad* con las reglas... Una computadora sigue las reglas que están escritas en un programa, mientras una pelota cayendo al suelo sólo está en conformidad con la leyes de la física... limito el término *basado en reglas* a los sistemas que explícitamente siguen reglas (Sloman, 1996, p. 5, la traducción es mía).

La propuesta de Sloman (1996) ha sido criticada por Gigerenzer y Regier (1996), quienes argumentan que en los dos sistemas de razonamiento hay mecanismos que pueden ser explicados a través de reglas.

Varios defensores de la teoría dual (Evans, 2003) parecen ponerse del lado de Gigerenzer al sostener que la conformidad con las reglas no es un criterio que separe claramente entre los dos sistemas de razonamiento propuestos por la teoría dual.

14 A diferencia de Evans y Over, para Sloman (1996, p. 6) la conciencia que se puede tener de los procesos de razonamiento es sólo una heurística falible, no una condición necesaria o suficiente, para identificar los diferentes sistemas.

15 Según Sloman, un caso especialmente ejemplar en donde se cumple con el criterio S es en la ilusión Müller-Lyer, donde la percepción y el conocimiento proveen distintas creencias. La idea de que las dos líneas horizontales son de diferente tamaño lo proporciona la percepción, mientras que la idea de que las dos líneas son del mismo tamaño es proporcionado por el conocimiento. Sloman considera que este ejemplo ilustra claramente la satisfacción del criterio S (1996 p. 11).

16 Dado que parte de la evidencia empírica para mostrar que los seres humanos cuentan con dos sistemas de razonamiento depende del criterio *S*, Sloman afirma que

> el fracaso de un problema de razonamiento para satisfacer el Criterio S no es evidencia en contra de los dos sistemas de razonamiento. El sistema asociativo y el sistema basado en reglas podrían converger en la misma respuesta (Sloman, 1996, p. 11, la traducción es mía).

17 Estos psicólogos no ofrecen un argumento para sostener esta afirmación. No obstante, como argumentaré más adelante, la postulación de dos sistemas de razonamiento por parte de la teoría dual sí sugiere que el razonamiento humano debe evaluarse con al menos dos estándares normativos.

18 Lo anterior muestra que no existe una relación completamente paralela entre la racionalidad 1 y el sistema tácito, y la racionalidad 2 y el sistema explícito. Evans y Over afirman

> Obviamente no podemos hacer una ecuación directa entre la racionalidad 1... y los procesos implícitos del tipo 1 exitosos, o entre la racionalidad 2 y procesos explícitos del tipo 2 correctos, bajo una teoría normativa (Evans y Over, 1996, p. 147, la traducción es mía).

19 Stanovich y West se comprometen directamente con la postura de Dawkins haciendo uso de su terminología y algunas de sus principales propuestas. En un intento por entender la propuesta de Stanovich y West, es prudente apuntar que para Dawkins un gen es una unidad biológica que forma parte de los cromosomas, el cual potencialmente permanece durante muchas generaciones y persigue sus propósitos de modo egoísta (Dawkins, 1985, pp. 40, 46 ss). Puede resultar extraño

sostener que un gen tiene objetivos o metas, pero Dawkins —al igual que Stanovich y West— consideran que hablar de esta manera es sólo metafórica. Dawkins afirma:

> he recalcado que no debemos pensar en los genes como agentes conscientes que persiguen un fin determinado. La ciega selección natural, sin embargo, los hace comportarse como si en realidad fuese así, y ha sido conveniente, como si empleásemos signos taquigráficos, referirnos a los genes en el lenguaje de la determinación. Por ejemplo, cuando decimos "los genes intentan aumentar su número en el futuro acervo génico", lo que realmente queremos decir es que "aquellos genes que se comportan de manera como para aumentar su número en los futuros acervos génicos tienden a ser los genes cuyos efectos percibimos en el mundo"... En ninguno... caso debemos atribuir a ello un sentido místico... la idea de la intención o propósito es sólo una metáfora... (Dawkins, 1985, pp. 291-92).

20 Dawkins, al igual que Stanovich y West, considera que los seres vivos son máquinas de supervivencia programados por los genes para preservarlos. Pero como los genes no pueden determinar directamente el comportamiento humano, construyeron el cerebro para llevar a cabo el control de las "máquinas de supervivencia". Lo distintivo de los cerebros del humano es que pueden rebelarse en contra de los genes, es decir, los seres humanos pueden seguir intereses que son propiamente humanos en detrimento de los intereses de los genes (Dawkins, 1985, pp. 88, 297-98). Esta propuesta de Dawkins es defendida por Stanovich en un libro de reciente aparición titulado *The Robot's Rebellion*, en clara alusión a los seres humanos como las máquinas que se rebelan a seguir los intereses de sus genes.

21 De hecho, al comparar la distinción entre racionalidad 1 y racionalidad 2 de Evans y Over y la distinción recién presentada Stanovich y West señalan:

> debido a que las teorías normativas se preocupan por las metas a un nivel personal, no a un nivel genético, las dos definiciones de racionalidad de Evans y Over (1996) pertenecen a lo que aquí ha sido denominado racionalidad normativa. Las dos se preocupan por las metas a nivel personal... su distinción entre dos conjuntos de mecanismos algorítmicos que pueden servir a la racionalidad normativa. La distinción que nosotros esbozamos es en término de *niveles* de optimización (a un nivel de los replicadores mismos —los genes— o a nivel del vehículo) (Stanovich y West, 2000, p. 665, la traducción es mía).

22 En general me centraré aquí en la posición de Evans y Over (1996).

23 Es importante señalar que el mismo Evans estudió junto con Wason muchos de los primeros experimentos en torno a la selección de tarjetas. El mismo Wason escribe el prefacio de *Rationality and Reasoning*.

24 En uno de estos experimentos, por ejemplo, se presenta a través de una computadora la tarea de selección tal y como comúnmente se ha formulado. Haciendo uso del *mouse* de la computadora se les pedía a los sujetos que apuntaran a las tarjetas que seleccionarían, pero que oprimieran el botón del *mouse* hasta estar seguros de su decisión. El tiempo en que se apuntaba una tarjeta antes de seleccionarla, así como el tiempo de inspección fue grabado por la computadora. En este experimento se probó que:

> Aquellas tarjetas que tienen una mayor frecuencia de selección tienen también una media mayor en tiempo de inspección... Para cualquier tarjeta dada, aquellos sujetos que escogen la tarjeta tendrán mayor tiempo de inspección que aquellas que no seleccionaron... Los sujetos apenas inspeccionan aquellas tarjetas que no seleccionaron (Evans y Over, 1996, p. 59, la traducción es mía).

25 Evans y Over (1996, p. 49) reducen el uso de la palabra sesgo a la desviación de un estándar normativo formal, es decir, a la racionalidad 2. En sentido estricto, según estos psicólogos, no hay sesgos de razonamiento desde la racionalidad 1.

26 Evans y Over revisan varios de los experimentos expuestos en el segundo capítulo de este libro.

27 Posteriormente argumentaré que los presupuestos normativos de la teoría dual no se reducen a la visión estándar de la racionalidad.

28 Evans y Over ni siquiera consideran que los seres humanos seamos siempre racionales desde la racionalidad 1:

> No estamos diciendo que las personas invariablemente son racionales 1 porque cualquier racionalidad está claramente acotada por constreñimientos cognitivos. Por supuesto, las personas podrían fracasar para alcanzar sus metas porque no han tenido la oportunidad de aprender la estrategia requerida o porque la tarea excede sus capacidades cognitivas de procesamiento. Son particularmente difíciles de alcanzar las metas muy altas y de largo alcance, concebibles sólo en el lenguaje y que requieren de mucho procesamiento explícito (Evans y Over, 1996, p. 142, la traducción es mía).

Nótese que Evans y Over están presuponiendo que la racionalidad 1 es un tipo de racionalidad dependiente de recursos.

29 Esta idea es llevada a la racionalidad, por ejemplo Evans y Over sostienen:

> aceptamos el argumento evolucionista de que el procesamiento tácito es compartido con los animales, quienes de este modo pueden ser racionales 1, pero sostenemos que la racionalidad 2 es únicamente humana (Evans y Over, 1996, p. 154, la traducción es mía).

30 Otra de las críticas de Gigerenzer y Regier a la propuesta de Sloman gira en torno a la idea de la existencia de un sistema asociativo, en donde las reglas no juegan un papel importante.

31 Es posible, según Gigerenzer y Regier (1996), apelar a consideraciones semánticas y pragmáticas, tales como las implicaciones conversacionales expuestas en el capítulo anterior, para poder entender cómo un sujeto puede sostener simultáneamente dos creencias contradictorias. Si embargo, como argumenté en el capítulo anterior, una dificultad que presenta la interpretación pragmatista es que no puede explicar cómo es que las versiones frecuentistas del problema "Linda, la cajera feminista", usadas por los psicólogos evolucionistas, hacen que los sujetos satisfagan los criterios provenientes de la visión estándar de la racionalidad.

32 Esta idea no implica que todos los resultados de S1 sean irracionales, según el estándar normativo de racionalidad, ni tampoco que los resultados de S2 sean siempre irracionales, según la racionalidad evolucionista. Habrá ocasiones en donde los resultados de cada uno de los sistemas convergen, así como la evaluación que se les da a los mismos. Esto se debe a que, según la teoría dual, el pensamiento humano se entiende como una interacción entre los dos sistemas de razonamiento (Stanovich y West, 2000, p. 662).

VI. LA TEORÍA DUAL DE SISTEMAS Y LAS NOCIONES DE JUSTIFICACIÓN Y DE RACIONALIDAD

1 Al final del capítulo II argumenté que a nivel descriptivo la tradición de heurística y sesgo no tiene implicaciones importantes acerca del debate en torno a la noción de justificación epistémica. Asimismo, en las conclusiones del capítulo III argumenté que a nivel descriptivo la psicología evolucionista tampoco puede ser usada para dirimir la discusión entre nociones internistas y externistas.

2 Uso la expresión *proceso cognitivo adecuado* para no identificar a las diferentes propuestas externistas con el confiabilismo de Goldman. Un proceso adecuado podría ser un proceso confiable *à la* Goldman, pero podría ser también un proceso que nomológicamente conecta al mundo con las creencias de los sujetos como sostiene Armstrong

(Bonjour, 1980), o una facultad "virtuosa" como sostienen Sosa (1992) y Greco (2006).

3 Todas las citas que hago de Eraña (en proceso a y b) provienen de trabajos en proceso, por lo que es posible que en versiones posteriores a los trabajos que aquí cito no sostenga exactamente lo mismo.

4 Problemas como salir a comprar el periódico, manejar el automóvil, etc., son el tipo de problemas que, según los defensores de la teoría dual, los seres humanos resuelven con facilidad.

5 La conexión particular que Eraña (en proceso b) desea hacer es entre lo que Evans y Over denominan "racionalidad 1", S1 y el criterio externista de justificación, así como entre lo que Evans y Over denominan "racionalidad 2", S2 y el criterio internista de justificación. Su argumento general para apoyar estas conexiones se funda en otras consideraciones tales como que S1 es evaluado por una teoría consecuencialista y la noción externista de justificación es una postura consecuencialista. Sin embargo, considero que esta conexión no es adecuada porque, como he expuesto en el capítulo anterior, no existe consenso dentro de los defensores de la teoría dual sobre cuáles deben ser las teorías normativas con las que se debe evaluar el razonamiento. En la siguiente sección "Racionalidad y justificación desde la teoría dual de sistemas" defiendo mi propia idea sobre la conexión entre justificación y las teorías normativas postuladas por los defensores de la teoría dual.

6 No es del todo claro por qué el rasgo (c) debería ser una característica que apoye la idea de que los productos de S2 están justificados según el internismo.

7 En contraste, la noción de justificación externista sólo estipula las propiedades que tiene una creencia justificada, pero no puede ser usada como guía o criterio que los individuos utilicen para decidir qué creer. Véase sección "La teoría confiabilista de la justificación" del capítulo I.

8 Para argumentar a favor de que S2 produce creencias justificadas, según el criterio internista de justificación, Eraña (en proceso a) afirma,

> es *suficiente* enfatizar la siguiente característica de S2... es un sistema de razonamiento cuyos procesos y *outputs* pueden ser evaluados apelando a estándares normativos de racionalidad, cuyas reglas pueden ser accesibles a nosotros, y cuya operación podemos ser completamente conscientes (Eraña, en proceso a, p. 8, la traducción es mía, el énfasis es mío).

9 Muchas veces se ha puesto en duda la noción misma de acceso interno (King, 2000) o la idea de *poseer evidencia* que justifica las creencias de un sujeto (Goldman, 1993). Por ejemplo, generalmente se considera

que un sujeto *S* puede racionalmente creer en *p*, si *p* es apoyada por el total de evidencia que *S* posee. Sin embargo, es poco claro si lo que cuenta como evidencia para *S* es todo aquello almacenado en la memoria de *S*, o bien, sólo aquello a lo que *S* puede fácilmente acceder. Cualquiera de las alternativas tiene dificultades, aun cuando la segunda alternativa, según Goldman (1993), parece más alentadora. Sin embargo, para Goldman (1993), la psicología de la memoria enseña que no existe una manera clara de definir "fácilmente accesible" por lo que es posible que la noción de "posesión de evidencia" o "acceso a evidencia" deba ser abandonada por principios muchos más complejos y sustanciales.

10 Goldman (1989) ha sostenido que esta propuesta de Bonjour (como la de la mayoría de los internistas) hace que la noción de justificación pueda aplicarse sólo a los individuos "epistemológicamente sofisticados", por lo que debe ser rechazada.

11 Eraña (en proceso a) comparte esta manera de entender la noción de justificación al afirmar que detrás de este concepto está siempre un punto de vista reflexivo.

12 Nótese la clara coincidencia de Eraña (en proceso a) con una postura como la de Bonjour (1985), en donde la justificación de una creencia para un sujeto *S* depende de la coherencia con un sistemas de creencias y de la justificación de que la coherencia conduce a la verdad.

13 En el capítulo I expuse una analogía que Bonjour (1985) y Goldman (1980) hacen con la ética señalando que la ética internista se puede entender como una ética subjetiva (o una ética desde la primera personas) y una ética externista se podría entender como una ética objetiva (o una ética desde la tercera persona).

14 Trataré este punto con más detalle en la siguiente sección.

15 Esta idea es compatible con la hipótesis defendida por los defensores de la teoría dual, según la cual, los seres humanos comparten S1 con algunos animales, mientras que S2 es un sistema propiamente humano.

16 Las críticas a la visión estándar de la racionalidad se encuentran en la sección "El rechazo de la visión estándar de la racionalidad", del capítulo IV. Precisamente en las conclusiones de tal capítulo subrayé la importancia de una teoría del razonamiento comprometida con diferentes acercamientos a la racionalidad.

17 Hay que recordar que la postura consecuencialista no está en contra de las reglas de razonamiento, sino que mantiene que razonar únicamente de acuerdo con algún conjunto de reglas no es constitutivo de un buen razonamiento (Foley, 1993).

18 El argumento, de hecho, es lógicamente incorrecto.

19 Goldman (1986), de hecho, considera que este tipo de metas, en tanto que no son epistemológicas, no pueden ser usadas para brindar criterios de justificación epistémica.

20 Esta idea parece clara en la siguiente afirmación:

> Primero que nada, es racional 1 para las personas buscar la realización de todas la metas, no sólo de las metas epistémicas. Podría ser el caso que otras metas no epistémicas entren en conflicto con metas epistémica. Por ejemplo... podría ser más importante para algunas personas, bajo ciertas circunstancias, mantener la confianza en sí mismos o la autoestima que satisfacer cualquier principio epistémico (Evans y Over, 1996, p. 42, la traducción es mía).

21 Haciendo críticas a la postura pragmatista de Stich (1990), Kornblith (1990) argumenta que la verdad es valiosas precisamente porque es la que conduciría al acercamiento de las metas que uno valora.

BIBLIOGRAFÍA

Adler, J. 1984. "Abstraction is uncooperative", *Journal for the Theory of Social Behavior*, 14, pp. 165-181.

Alston, W. 1976. "Has foundationalism been refuted?", *Philosophical Topics*, 29, pp. 287-305.

Barkow, J., Cosmides, L., Tooby, J., 1992. *The Adapted Mind. Evolutionary Psychology and the Generation of Culture*, New York: Oxford University Press.

Bonjour, L. 1976. "The coherence theory of empirical knowledge", *Philosophical Studies*, 30, pp. 281-312.

— 1978. "Can empirical knowledge have a foundation?", *American Philosophical Quarterly*, 15, pp. 1-13.

— 1980. "Externalists theories of empirical knowledge", *Midwest Studies in Philosophy*, 5, pp. 53-77.

— 1985. *The Structure of Empirical Knowledge*, Cambridge, Mass.: Harvard University Press.

— 2003. "A version of internalist foundationalism", en Bonjour, L. y E. Sosa (eds.), *Epistemic Justification: Internalism vs. Externalism. Foundations vs. Virtues*, Oxford: Blackwell.

Casscells, W., Schoenberger, A., y Grayboys, T. 1978. "Interpretation by physicians of clinical laboratory results", *New England Journal of Medicine*, 299, pp. 999-1000.

Carruthers, P. 2006. "The case for massively modular models of mind", en Stainton, R. (ed.), *Contemporary Debates in Cognitive Science*, Oxford: Blackwell Publishing.

— 2009. "An architecture for dual reasoning" en J. Evans y K. Frankish (eds.), *In Two Minds: Dual Processes and Beyond*. OUP.

Cheng, R., Holyoak, K. 1985. "Pragmatic reasoning schemas", *Cognitive Psychology*, 17, pp. 391-416.

Cherniak, C. 1986. *Minimal Rationality*, Cambridge, Mass.: MIT Press.

Chomsky, N. 1957. *Syntactic Structures*, The Hauge: Mouton.

Clarck, A. 1963. "Knowledge and grounds: A comment on Mr. Gettier's paper", *Analysis*, 24, 2, pp. 46-48.

Cohen, J. 1981. "Can human irrationality be experimentally demonstrated?", *The Behavioral and Brain Sciences*, 4, pp. 317-331.

— 1986. *The Dialogue of Reason*, Oxford: Claredon Press.
Conee, E., Feldman, R. 1998. "The generality problem for reliabilism", *Philosophical Studies*, 89, 1, pp. 1-29.
Cosmides, L., Tooby, J. 1992. "Cognitive adaptations for social exchange", en Barkow, J., Cosmides, L., Tooby, J., pp. 163-228.
— 1994. "Origin of domain specificity: The evolution of functional organization", en Hirschfeld, L., Gelman, S., *Mapping the Mind*, Cambridge, Mass.: MIT Press, pp. 85-116.
— 1996. "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty", *Cognition*, 58, pp. 1-73.
— 1997. "Evolutionary Psychology: A Primer", en el sitio electrónico http://www.psych.ucsb.edu/research/cep/primer.htm/.
Cummins, D. 1996. "Evidence for innateness of deontic reasoning", *Mind and Language*, 11, pp. 160-190.
Damasio, A. 2003. *El error de Descartes. La emoción, la razón y el cerebro humano*. Trad. Joandomènec Ros, Barcelona: Crítica.
Danielson, P. 2004. "Rationality and evolution", en Mele, A., Rawling, P., *The Oxford Handbook of Rationality*, Oxford: Oxford University Press, pp. 415-437.
Dancy, J. 1986. *An Introduction to Contemporary Epistemology*, Oxford: Blackwell Publishers.
Darwin, C. 2004. *The Origin of Species*, Edison, N. J.: Castle Books.
Davidson. D. 1983. "A coherence theory of truth and knowledge", en Henriech Dieter (ed.), *Kant oder Hegel?*, Stuttgart: Klett-Cotta, pp. 471-489.
Dawkins, R. 1985. *El gen egoísta*. Trad. Juana Robles, Barcelona: Salvat Editores.
Elio, R. 2002. "Issues in commonsense reasoning and rationality" en Elio, R. (ed.), *Common Sense, Reasoning and Rationality*, New York: Oxford University Press, pp. 3-36.
Eraña, A. (en proceso a) "Dual system theory. A way to bridge the gap between Internalism and Externalism".
— (en proceso b) "La noción de 'justificación', ¿un concepto dual?".
Evans, J. 1991. "Theories of human reasoning: The fragmented state of the art". *Theory and Psychology*, 1, pp. 83-115.
— 1993. "Bias and rationality", en Manktelow, K.I., Over, D. *Rationality: Psychological and Philosophical Perspectives*, London: Routledge and Kegan Paul.
— 2003. "In two minds: Dual-process accounts of reasoning", en *Trends in Cognitive Sciences*, vol. 7, no. 10.

— 2006. "Dual system theories of cognition. Some issues", Conference at Cognitive Science Society, Vancouver.
Evans, J., Over, D. 1996. *Rationality and Reasoning*, East Sussex: Psychology Press.
Fiedler, K. 1988. "The dependence of the conjunction fallacy on subtle linguistic factors", *Psychological Research*, 50, pp. 123-129.
Fodor, J. 1983. *The Modularity of Mind*, Cambridge, Mass.: MIT Press.
Foley, R. 1993. *Working Without a Net: A Study of Egocentric Epistemology*, New York: Oxford University Press.
García Campos, J. 2003. *El debate contemporáneo en torno a la noción de justificación epistémica*. Tesis de maestría, Facultad de Filosofía y Letras, UNAM.
— 2008. "Psicología cognitiva del razonamiento. Algunas expectativas y retos", *Ludus Vitalis*, XVI, num. 29, pp. 173-176.
Gigerenzer, G. 1991. "How to make cognitive illusions disappear: Beyond 'heuristics and biases' ", *European Review of Social Psychology*, 2, pp. 83-115.
— 1993. "The bounded rationality of probabilistic models", en Manktelow, K. y Over, D. (eds.), *Rationality: Psychological and Philosophical Perspectives*. London: Routledge and Kegan Paul.
— 1994. "Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa)", en Wright y Ayton (eds.), *Subjective Probability*, New York: John Wiley.
— 1996. "On narrow norms and vague heuristics: A reply to Kahneman and Tversky", *Psychological Review*, 103, pp. 592-596.
Gigerenzer, G., Hoffrage, U. y Kleinbölting, H. 1991. "Probabilistic mental models: A Brunswickean theory of confidence", *Psychological Review*, 98, pp. 506-528.
Gigerenzer, G., Hug, K. 1992. "Domain-specific reasoning: Social contracts, cheating and perspective change", *Cognition*, 43, pp. 127-171.
Gigerenzer, G,. Regier, T. 1996. "How do we tell an association from a rule? Comment on Sloman (1996)", *Psychological Bulletin*, 119, pp. 23-26.
Gigerenzer, G., Selten, R. 2002. "Rethinking rationality", en Gigerenzer, G. y Selten, R. (eds.), *Bounded Rationality. The Adaptive Toolbox*, Cambridge, Mass.: The MIT Press.
Gigerenzer, G., Todd, P., and the ABC Research Group. 1999. *Simple Heuristics That Make Us Smart*, New York: Oxford University Press.
Gilovich, T. y Griffin, D. 2002. "Heuristics and biases: Then and now", en Gilovich, T, Griffin, D y D. Kahneman, *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge: Cambridge University Press, pp. 1-18.

Goldman, A. 1978. "A causal theory of knowing", en Pappas y Swain (comps.), *Essays on Knowledge and Justification*, Ithaca: Cornell University Press, pp. 67–86.
— 1979. "What is justified belief?", in Pappas, G. (ed.). *Justification and Knowledge: New Studies in Epistemology*, Dordrecht: Reidel, pp. 1-23.
— 1980. "The internalist conception on justification", *Midwest Studies in Philosophy*, 5, pp. 27 – 52.
— 1986. *Epistemology and Cognition*, Cambridge, Mass.: Harvard University Press.
— 1989. "Bonjour's *Structure of Empirical Knowledge*", Bender, John (ed.) *The Current State of Coherence Theory*, Dordrecht: Kluwer Academic Publishers, pp. 105-114.
— 1992. *Liasons*, Cambridge, Mass.: MIT Press.
— 1993. *Philosophical Applications of Cognitive Science*, Boulder: West View Press.
Griggs, R., Cox, J. 1982. "The elusive thematic-materials effect in Wason's selection task", *British Journal of Psychology*, 73, pp. 407-420.
Greco, J. 2006. "Virtue, luck and the Pyrrhonian problematic", *Philosophical Studies*, 130, pp. 9-34.
Horst, S. 2008 "The computational theory of mind", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2008/entries/computational-mind/.
Jaquette, D. 2003. "Psychologism the philosophical *shibboleth*", en Jaquette, D. (ed.), *Philosophy, Psychology, and Psychologism*, Dordrecht: Kluwer Academic Publishers.
Kahneman, D. 2002. "Maps of bounded rationality: A perspective on intuitive judgment and choice", Nobel Prize Lecture.
Kahneman, D., Slovic, P., Tversky, A. 1982. *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.
Kahneman, D., Tversky, A. 1982a. "Introduction", en Kahneman, D., Slovic, P., Tversky, A.
— 1982b. "On the psychology of prediction", en Kahneman, D., Slovic, P., Tversky, A., pp. 48-68.
— 1996. "On the reality of cognitive illusions: A reply to Gigerenzer's critique". *Psychological Review*, 103, pp. 582-591.
Karmiloff-Smith, A. 1992. *Beyond Modularity. A Developmental Perspective on Cognitive Science*, Cambridge, Mass.: MIT Press.
Kim, J. 1994. "What is 'naturalized epistemology'?", en Kornblith, H. (ed.), pp. 33-56.
Kitcher, P. 1992. "The naturalists return", *The Philosophical Review*, 101, 1, pp. 53-114.

— 1993. *The Advancement of Science*: *Science without Legend*, *Objectivity without Illusion*, New York: Oxford University Press.
King, P. 2000. "Internalismo, externalismo y autoconocimiento", *Crítica*, 96, pp. 99-119.
Koehler, J. 1996. "The base-rate fallacy reconsidered", *Behavioral and Brain Sciences*, 19, p. 1-53.
Kornblith, H. 1982. "The psychological turn", *Australasian Journal of Philosophy*, 60, No. 3, pp. 238-253.
— 1983. "Justified belief and epistemically responsible action", *Philosophical Review*, XCII, pp. 33-48.
— 1993. "Epistemic normativity", *Synthese*, 94, pp. 357-376.
— (comp.). 1994. *Naturalizing Epistemology*, Cambridge, Mass.: The MIT Press.
— 2000. "Internalism and Externalism: A brief historical introduction", en, H. Kornblith (comp.), *Epistemology: Internalism and Externalism*, Malden, Mass.: Blackwell Publishing.
Larry, A., Moore, M. 2008. "Deontological ethics", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2008/entries/ethics-deontological/.
Laudan, L. 1984. *Science and Values*, Los Angeles: University of California Press.
Lehrer. K. 1965. "Knowledge, truth and evidence", *Analysis*, 25, 5, pp. 168-175.
— 2000. *Theory of Knowledge*, 2nd ed, Boulder: West View Press.
Nisbett, R. (ed.). 1993. *Rules for Reasoning*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.
Pappas, G. 2005. "Internalist vs. Externalist conceptions of epistemic justification", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2008/entries/justep-intext/.
Piatelli-Palmarini, M. 1994. *Inevitable Illusions: How Mistakes of Reason Rule Our Minds*, New York: John Wiley & Sons.
— 2005. *Los túneles de la mente. ¿Qué se esconde tras nuestros errores?*, trad. Maria Pons, Barcelona: Crítica.
Prinz, J. 2006. "Is the mind really modular?", en Stainton, R. (ed.) *Contemporary Debates in Cognitive Science*, Blackwell Publishing, pp. 22-36.
Quine, W. V. O. 1969. "Epistemology naturalized", *Ontological Relativity and Other Essays*, New York: Columbia University Press, pp. 69-90.
Samuels, R. 2000. "Massively modular minds: evolutionary psychology and cognitive architecture", en Carruthers, P. (ed.), *Evolution and the Human Mind*, Cambridge: Cambridge University Press.

— 2006. "Is the human mind massively modular?", en Stainton, R. (ed.) *Contemporary Debates in Cognitive Science*, Blackwell Publishing, pp. 37-56.

Samuels, R., Stich, S. 2004. "Rationality and psychology", en Mele, A., Rawling, P. (eds.), *The Oxford Handbook of Rationality*, Oxford: Oxford University Press, pp. 279-300.

Samuels, R, Stich, S., Bishop, M. 2002. "Ending the rationality wars: How to make disputes about human rationality disappear" en Elio, R. (ed.), *Common Sense, Reasoning and Rationality*, New York: Oxford University Press, pp. 236-268.

Samuels, R., Stich, S., Faucher, L. 2004. "Reason and rationality", en Matti Sintonen, et al., *Handbook of Epistemology*, Dordrecht: Kluwer, pp. 131-179.

Simon, H. 1957. *Models of Man: Social and Rational*, New York: Wiley.

Sinnott-Armstrong, W. 2006. "Consequentialism", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2008/entries/consequentialism/.

Sloman, S. 1996. "The empirical case for two systems of reasoning", *Psychological Bulletin*, 119, p. 3-22.

Slovic, P., Fischhoff, B., and Lichtenstein, S. 1976. "Cognitive processes and societal risk taking", en Carol, J., Payne, J. (eds.), *Cognition and Social Behavior*. Hillsdale, N. J.: Erlbaum.

Sosa, E. 1990. *Knowledge in Perspective*, New York: Cambridge University Press.

— 1992. *Conocimiento y virtud intelectual*, México: UNAM-FCE.

— 1994. "The analysis of 'knowledge that *p*' ", *Analysis*, 25, 1, pp. 1-18.

Sperber, D. 1994. "The modularity of thought and the epidemiology of representations", en Hirschfeld, L., Gelman, S., *Mapping the Mind*, Cambridge, Mass.: MIT Press, pp. 39-67.

Sperber, D., Girotto, V. 2003. "Does the selection task detect cheater-detection?", en Fitness, J. y Sterelny (eds.), *From Mating to Mentality*, New York: Psychology Press, pp. 197-226.

Stanovich, K. 1999. *Who Is Rational?* Mahwah, N. J.: Lawrence Erlbaum Associates.

— 2006. *The Robot's Rebellion. Finding Meaning in the Age of Darwin*, Chicago: The University of Chicago Press.

Stanovich, K., West, R. 2000. "Individual differences in reasoning: Implication for the rationality debate?", *Behavioral and Brain Science*, 23, pp. 645- 665.

— 2003. "Evolutionary versus instrumental goals: How evolutionary psychology misconceives human rationality", en Over, D. (ed.), *Evo-*

lution and the Psychology of Thinking: The Debate, East Sussex: Psychological Press.
Stein, E. 1996. *Without Good Reason*, Oxford: Claredon University Press.
Steup, M. 2005. "Jusification" , en Steup. M y Sosa, E. (eds.). *Contemporary Debates in Epistemology*, London, Blackewell, pp. 251-256.
Stich, S. 1990. *The Fragmentation of Reason*, Cambridge: MIT Press.
Tomasini, A. 2001. *Teoría del conocimiento clásica y epistemología wittgensteiniana*, México: Plaza y Valdés Editores.
Trivers, R. 1971. "The evolution of reciprocal altruism", *Quarterly Review of Biology*, 46, pp. 35-56.
Vega de, M. 1998. *Introducción a la psicología cognitiva*, Madrid: Alianza Editorial.
Villoro, L. 1982. *Creer, saber, conocer,* México: Siglo XXI Editores.
Wason, P. 1966. "Reasoning", en Foss, B. (ed.), *New Horizons in Psychology*, Harmondsworth: Penguin, pp. 135-151.
Williams, M. 2001. *Problems of Knowledge. A Critical Introduction to Epistemology*, Oxford: Oxford University Press.

COLECCIÓN
ESLABONES EN EL DESARROLLO DE LA CIENCIA

ESTUDIOS EN HISTORIA Y FILOSOFÍA DE LA BIOLOGÍA (2 vols.)
Editores:
Raúl Gutiérrez Lombardo; Jorge Martínez Contreras; José Luis Vera Cortés

Volumen 1. **HISTORIA**
Volumen 2. **FILOSOFÍA**

DEL A.D.N. A LA HUMANIDAD
HOMENAJE A FRANCISCO JOSÉ AYALA
Editora: Lucrecia Burges

LA MENTE ESTÉTICA
LOS ENTRESIJOS DE LA PSICOLOGÍA DEL ARTE
Gisèle Marty

LAS ANDANZAS DEL CABALLERO INEXISTENTE
REFLEXIONES EN TORNO AL CUERPO Y LA ANTROPOLOGÍA FÍSICA
José Luis Vera Cortés

NATURALEZA Y DIVERSIDAD HUMANA
Editores: Raúl Gutiérrez Lombardo; Jorge Martínez Contreras; José Luis Vera Cortés

SENDEROS DE LA CONSERVACIÓN Y DE LA RESTAURACIÓN ECOLÓGICA. EVALUACIÓN CRÍTICA Y ÉTICA
Editor: Jorge Martínez Contreras

METAPOCATÁSTASIS DE CIVILIZACIÓN
HUBO UN AYER, HOY UNA DIÁSPORA DE PEDAZOS ANTAGÓNICOS, EL MAÑANA ES LA ILUSIÓN
Flavio Cocho Gil

ARQUEOLOGÍA COGNITIVA PRESAPIENS
ACERCARSE A LA MENTE DE NUESTROS ANCESTROS. UNA APROXIMACIÓN HISTÓRICA AL CASO OLDOWAN
Aura Ponce de León

FUEGO Y VIDA
FUENTES DEL PENSAMIENTO QUÍMICO DE BUFFON
Violeta Aréchiga

Epistemología y psicología cognitiva propone un acercamiento naturalizado al estudio de la justificación. Partiendo de la encrucijada a la que han conducido las distintas nociones de justificación epistémica, en especial el debate entre internistas y externistas, se propone que el estudio del razonamiento y la racionalidad pueden ayudar a encontrar una solución a esta encrucijada. Así, en el libro se exploran distintas teorías sobre el razonamiento propuestas en la actual psicología cognitiva, para con éstas abordar el problema filosófico de la justificación. En el trabajo se revisan tres teorías del razonamiento: la tradición de heurística y sesgo; la psicología evolucionista, y la teoría dual de sistemas, así como los supuestos normativos y evaluativos que dichas propuestas asumen. Esta exploración concluye defendiendo que en psicología cognitiva del razonamiento, la teoría que de momento resulta más adecuada es la teoría dual de sistemas, y se argumenta que dicha propuesta psicológica apoya a un tipo de teoría externista de la justificación. De este modo, el libro ofrece una respuesta sugerente a un problema epistemológico central y crea un puente teórico entre las discusiones filosóficas y psicológicas que tienen como denominador común la cognición humana.

sep

ESLABONES EN EL DESARROLLO DE LA CIENCIA