



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO.

FACULTAD DE FILOSOFÍA Y LETRAS  
INSTITUTO DE INVESTIGACIONES FILOSÓFICAS

**HACIA UNA CONCEPCIÓN NATURALIZADA DE LOS  
PRINCIPIOS NORMATIVOS DE RAZONAMIENTO.**

*T E S I S*  
QUE PARA OBTENER EL TÍTULO DE  
MAESTRÍA EN FILOSOFÍA DE LA CIENCIA  
P R E S E N T A  
MARIA ELENA RAMOS ACEVES.

*DIRECTORA DE TESIS*  
DRA. CLAUDIA LORENA GARCÍA AGUILAR.

México D.F. 2007



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **AGRADECIMIENTOS.**

Quiero agradecer a la Dra. Claudia Lorena García Aguilar por haber aceptado recorrer este largo camino conmigo, su guía fue insustituible. Le agradezco infinitamente su paciencia y su trato siempre cordial.

Agradezco también a mis sinodales, los doctores: Ma. de los Angeles Eraña Lagos, Patricia King Dávalos, Axel Barceló Aspeitia y Sergio Martínez Muñoz, por el tiempo dedicado a la lectura de esta tesis y por sus valiosos comentarios.

# INDICE

<b>INTRODUCCIÓN.</b>	1
<b>CAPITULO UNO.</b>	
<b>PRIMEROS EXPERIMENTOS SOBRE RAZONAMIENTO EN PSICOLOGIA.</b>	9
1. Competencia y Ejecución.	10
2. Primeros Experimentos Sobre Razonamiento.	14
2.1 Visión Estándar de la Racionalidad.	14
2.2 La Tarea de Selección de Wason.	16
2.3 Investigaciones Subsecuentes Sobre la Tarea de Selección.	18
3. Efecto de Material Temático.	28
3.1 La Tarea de Selección.	29
3.2 El Problema THOG.	34
4. Conclusión.	39
<b>CAPITULO DOS.</b>	
<b>TEORIAS DEL RAZONAMIENTO.</b>	41
1. Teorías de Disponibilidad (TD).	41
2. Teorías de Esquemas de Razonamiento Pragmático (TERP).	42
3. Teoría del Contrato Social (TCS).	49
3.1 Contrastes: Teoría del Contrato Social vs. Teorías de la Disponibilidad	52
3.2 Contrastes: Teoría del Contrato Social vs. Esquemas de Razonamiento Pragmático.	55
4. Algunas Consideraciones Adicionales sobre TD, TERP y TCS.	57
5. Razonamiento Inductivo.	60
5.1 La Falacia de la Conjunción.	61
5.2 La Representación de la Información Probabilística.	69
6. Conclusión.	71

<b>CAPITULO TRES.</b>	
<b>ACERCA DE LA UNIVERSALIDAD DE LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO.</b>	74
1. El Argumento de Davidson.	75
1.1 ¿Por qué razón si hubiera sujetos que utilizaran principios normativos de razonamiento fundamentalmente distintos de los nuestros no podríamos interpretarlos?	75
1.2 ¿Por qué razón la idea de un sujeto que sea racional y al mismo tiempo no podamos interpretar es ininteligible?	85
2. En Contra de la Universalidad de los Principios Normativos de Razonamiento.	88
2.1 Racionalidad Mínima.	89
2.1.1 La Posición de Davidson.	91
2.1.2 En Contra de la Posición de Inferencias Esenciales.	93
2.2 El Argumento de Stephen Stich.	99
3. Conclusión.	105
<b>CAPITULO CUATRO.</b>	
<b>LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO EN EL MARCO DE UNA EPISTEMOLOGIA NATURALIZADA.</b>	107
1. La Justificación de los Principios Normativos de Razonamiento.	107
1.1 Epistemología Tradicional y Epistemología Naturalizada.	108
1.2 Los Principios Normativos de Razonamiento están “Indexados” a la Competencia Humana de Razonamiento.	113
1.3 Una objeción a la Concepción Naturalizada de los Principios de Razonamiento.	115
2. Los Principios Normativos de Razonamiento no son Necesariamente Universales.	118
2.1 Argumentos Conceptuales en contra de la Universalidad de los Principios Normativos de Razonamiento.	118
2.2 Argumentos Empíricos en contra de la Universalidad de los Principios Normativos de Razonamiento.	124
2.3 Relativismo.	131
3. Los Principios Normativos de Razonamiento no son, Necesariamente, Principios Formales.	134

<b>CONCLUSIÓN.</b>	137
<b>APENDICE A.</b>	141
<b>APENDICE B.</b>	148
<b>BIBLIOGRAFIA.</b>	150

## INDICE DE EXPERIMENTOS.

<b>CAPITULO UNO</b>		
Versión estándar de la tarea de selección.	Wason, 1966	16
Se explora la posibilidad de que los sujetos puedan estar interpretando la regla como estableciendo una conexión causal o temporal entre antecedente y consecuente.	Wason, 1969	18
Se explora la posibilidad de que los sujetos estén confundiendo ambos lados de las cartas, pensando en “el otro lado de la carta” como el lado que está hacia abajo.	Wason & Johnson-Laird, 1970 Goodwin & Wason, 1972	19
Se explora la posibilidad de que la falta de familiaridad de los sujetos con la regla sea el factor que da cuenta de su mala ejecución en la tarea de selección.	Wason & Shapiro, 1971	19
Se explora la posibilidad de que la tarea de selección produzca una sobrecarga cognoscitiva.	Goodwin & Wason, 1972	20
Las instrucciones hacen énfasis en la tarea de falsificación de la regla.	Hughes, 1966 Wason, 1968	20
La ejecución de los sujetos es afectada por la presencia de componentes negativos en la regla.	Evans, 1972	21
Evidencia a favor de un sesgo confirmatorio: El problema 2 4 6.	Wason, 1960	22
Evidencia a favor de un sesgo de positividad.	Clark & Chase, 1972	23
Otra versión del problema 2 4 6: DAX, MED	Tweney, et. al., 1980	23
Se explora la posibilidad de que una tabla de verdad defectiva subyazca en la competencia de razonamiento de los sujetos.	Wason, 1968 Johnson-Laird & Tagart, 1969	25
Se enfrenta a los sujetos a sus propias contradicciones.	Wason, 1969 Wason & Johnson-Laird, 1970	27
El problema de las ciudades y medios de transporte	Wason & Shapiro, 1971	29
Problema postal.	Johnson-Laird & Legrenzi & Legrenzi, 1972	29
El problema de las ciudades y medios de transporte.	Bracewell & Hidi, 1974	30
El problema de las materias académicas y universidades.	Van Duyne, 1974	30
“If I eat haddock, then I drink gin”	Manktelow & Evans, 1979	31
Fallas en replicar los resultados de Wason & Shapiro (1971), en el caso del problema de las ciudades y medios de transporte.	Manktelow % Evans, 1979 Brown, et. al., 1980 Griggs & Cox, 1982	31
El problema de locaciones en un campo universitario y medios de transporte.	Yachanin & Tweney, 1982	31
Falla en replicar los resultados obtenidos por Johnson-Laird et. al. (1972), en el caso del problema postal.	Griggs & Cox, 1982	31

Falla en replicar los resultados de Van Duyne, (1974).	Yachanin & Tweney, 1982	31
El problema de la edad y el tipo de bebida.	Griggs & Cox, 1982	32
Versión actualizada del problema postal.	Goldig, 1981	32
El problema del gerente de Sears.	Roy D'Andrade	33
El problema de la edad y el color de la ropa.	Cox & Griggs, 1982	33
Versión estándar del problema THOG.	Wason & Brooks, 1979	34
Problema Gastronómico.	Newstead, Griggs & Wamer, 1982	37
Problema THOG (Contexto Realista).	Giroto & Legrenzi, 1989	37

<b>CAPITULO DOS.</b>		
Se pone a prueba la hipótesis de que la evocación de un esquema de permiso puede ser facilitada por dos factores: 1) la provisión de una razón o propósito social para la regla condicional y, 2) la experiencia previa o la familiaridad con la regla. Se utilizan dos tipos de problemas: problema postal y problema del cólera. (experimento 1)	Cheng & Holyoak, 1985	44
Se contrasta la ejecución de los sujetos en dos tipos de problemas: 1) Un problema que describe una situación de permiso en términos abstractos. 2) Un problema arbitrario en términos concretos. (experimento 2)	Cheng & Holyoak, 1985	47
Se contrasta la ejecución de los sujetos en cuatro tipos de problemas: 1) Un problema de contrato social estándar no familiar (CSE-noF). 2) Un problema descriptivo no familiar (D-noF). 3) Un problema descriptivo familiar (D-F). 4) Un problema abstracto (ABS). (experimento 1)	Cosmides, 1989	52
Se contrasta la ejecución de los sujetos en cuatro tipos de problemas: 1) Contrato social invertido no familiar (CSI-noF). 2) Descriptivo no familiar (regla invertida), (D-noF). 3) Descriptivo familiar (D-F). 4) Abstracto (ABS). (experimento 2)	Cosmides, 1989	53
Se contrasta la ejecución de los sujetos en dos tipos de problemas: 1) Un problema de permiso con la estructura de un contrato social estándar (P-CSE). 2) Un problema de permiso sin la estructura de un contrato social (P-noCS). (experimento 3)	Cosmides, 1989	56



Se contrasta la ejecución de los sujetos en dos tipos de problemas: 1) Un problema de permiso con la estructura de un contrato social invertido (P-CSI). 2) Un problema de permiso sin la estructura de un contrato social, pero con la regla invertida (P-noCS). (experimento 4)	Cosmides, 1989	56
El problema de Linda.	Tversky & Kahneman, 1983	62
Se presentan una serie de manipulaciones al problema de Linda con el fin de explorar la posibilidad de que la comisión de la "falacia de la conjunción" pueda ser reducida.	Tversky & Kahneman, 1983	62-64
Se explora la posibilidad de que los sujetos estén entendiendo el término "probabilidad" de una manera alternativa. Problema del dado con cuatro caras verdes y dos caras rojas.	Tversky & Kahneman, 1983	67
Se explora la posibilidad de que la falacia de la conjunción pueda ser eliminada al presentar a los sujetos un problema cuyo planteamiento sea capaz de evocar en el sujeto una representación de la información en términos de frecuencia.	Gigerenzer, 1993	70

<b>CAPITULO CUATRO.</b>		
Se explora la posibilidad de que los sujetos puedan mejorar su ejecución en la tarea de selección como resultado de un micro-entrenamiento previo en el condicional material y sus reglas asociadas. (experimento 1)	Cheng, Holyoak, Nisbett & Oliver, 1993	126
Se explora la posibilidad de que los sujetos puedan mejorar su ejecución en la tarea de selección como resultado de un curso de un semestre en lógica estándar. (experimento 2)	Cheng, Holyoak, Nisbett & Oliver, 1993	128
Se explora la posibilidad de que los sujetos puedan mejorar su ejecución en la tarea de selección como resultado de un entrenamiento abstracto en las reglas asociadas con un "esquema de obligación". (experimento 3)	Cheng, Holyoak, Nisbett & Oliver, 1993	128

## INTRODUCCION.

Uno de los objetivos de la epistemología ha sido el establecimiento de criterios que nos permitan discernir el conocimiento genuino de la creencia no sustentada y de la mera opinión. De aquí el interés del epistemólogo en dilucidar el sentido de nociones como la de "creencia justificada". En tanto que, el que una creencia esté justificada depende, en parte, de que los procesos de razonamiento por medio de los cuales arribamos a ella sean correctos, la noción de "buen razonamiento" es también de fundamental interés para el epistemólogo.

¿Cuándo un razonamiento cuenta como buen razonamiento? ¿Qué es razonamiento correcto? Son preguntas a las que se busca responder con el fin de estipular los principios de razonamiento que han de normar nuestros procesos cognoscitivos y todo quehacer científico.

En la medida en que la epistemología se ha preocupado por dilucidar cuáles son los fundamentos últimos del conocimiento, se ha preocupado también de los fundamentos de su propio quehacer. Parece necesario que el epistemólogo deje en claro cuáles son los fundamentos sobre la base de los cuales él está en posición de decirnos cuáles son los principios normativos de razonamiento a los que todos nosotros debemos adherirnos. De no ser así ¿por qué habríamos de acatar las normas de buen razonamiento que el epistemólogo establece?

A lo largo del desarrollo de la epistemología, dos posiciones diferentes se han asumido respecto a los fundamentos de la normatividad epistémica: una de ellas es la posición asumida tradicionalmente, la otra es la posición que emerge a partir de W.V.O. Quine y que ha venido a conocerse como epistemología naturalizada.

En la concepción de la epistemología tradicional (no naturalizada), la justificación de todo estándar normativo de carácter epistémico, proviene de un análisis de nuestras intuiciones respecto de los significados y usos de nuestros conceptos epistémicos, tal como lo es el concepto de *racionalidad*. Por lo tanto, dicha justificación procede *a priori*, es decir, al margen de la experiencia. De modo que las maneras en que de hecho los seres humanos adquirimos y revisamos nuestros sistemas de creencias no son relevantes en la determinación de las maneras en que debemos hacerlo. Para la epistemología tradicional, esto último debe forzosamente ser así, de lo contrario la justificación de los estándares normativos sería circular, pues el epistemólogo estaría tratando de establecer los criterios que confieren justificación a nuestros procedimientos de adquisición y revisión de creencias mediante la

apelación a las maneras en que de hecho procedemos en la adquisición y revisión de creencias, es decir mediante la apelación a aquello mismo que quiere validar. Este modo de proceder –desde el punto de vista de la concepción tradicional- iría claramente en contra de los objetivos de la epistemología misma.

En contraste con la epistemología tradicional, para la epistemología naturalizada, la justificación de nuestros estándares normativos de carácter epistémico no puede proceder *a priori*. Las razones para asumir lo anterior se encuentran en los argumentos esgrimidos por Quine en contra de la distinción analítico-sintético, argumentos que llevaron finalmente a la disolución de tal distinción. Básicamente Quine sostiene que no hay enunciados analíticos, es decir, enunciados cuya verdad pueda ser establecida únicamente en virtud de los significados de las palabras que componen a dichos enunciados; por lo tanto, no hay enunciados que no sean susceptibles de ser revisados a la luz de la experiencia. Para Quine, todo enunciado puede ser revisado a la luz de consideraciones empíricas, incluido todo enunciado filosófico. De aquí que la línea de demarcación entre la filosofía y la ciencia empírica pierda su carácter nítido. Consecuentemente, la línea de demarcación entre epistemología y psicología también pierde nitidez. Si todo enunciado epistemológico puede ser revisado a la luz de la experiencia, entonces la epistemología debe echar mano de la psicología en aras de alcanzar su objetivo de establecer estándares de racionalidad. De aquí que una tesis central de una epistemología naturalizada sea que las maneras en que los seres humanos de hecho razonamos deben ser consideradas a fin de establecer las maneras en que debemos hacerlo.

El presente trabajo se circunscribe en el ámbito de una concepción naturalizada de la epistemología, por lo tanto los resultados de la investigación experimental en psicología del razonamiento serán aquí relevantes. Veremos qué pueden los experimentos sobre razonamiento sugerirnos en relación a la naturaleza de los principios normativos de razonamiento.

Lo primero de lo que podremos darnos cuenta es de que los resultados de dichos experimentos parecen no ser compatibles con una visión de la racionalidad ampliamente sustentada, tanto por filósofos como por psicólogos. Me refiero aquí a la que Stein (1996) llama *visión estándar de la racionalidad*.

Según la visión estándar de la racionalidad, ser racional es razonar de acuerdo con principios de naturaleza formal tales como los que se derivan de la lógica clásica, la teoría matemática de la probabilidad y la teoría de la elección racional (entre otros posibles cálculos), estándares que se consideran universales.

En la filosofía, una de las más prominentes concepciones de la racionalidad, que puede ser identificada como acorde con la visión estándar, es la propuesta por Donald Davidson.

Para Davidson, debe existir un núcleo importante de principios de razonamiento compartido por todos los seres humanos. Por supuesto, él identifica estos principios con los principios de la lógica, teoría de la probabilidad y teoría de la elección racional.

El argumento de Davidson para defender lo anterior es que, de no existir tal núcleo de racionalidad compartido, la interpretación no sería posible. Y la interpretación no sería posible simplemente porque un ser que no razonara de acuerdo con nuestros principios de razonamiento no podría ser considerado como un agente cognoscitivo.

La razón por la que Davidson afirma esto último está en la conexión que él establece entre las nociones de *racionalidad* y *creencia*. Davidson sostiene que una determinada creencia, pongamos por caso la creencia de Juan en que "llueve", sólo puede ser identificada como esa creencia dadas las relaciones lógicas que la creencia en cuestión mantiene con otras creencias y con los deseos e intenciones de Juan, y dadas las relaciones causales de dicha creencia con los objetos y los eventos en la situación de Juan. Si las relaciones lógicas y causales apropiadas no se sostienen, entonces la creencia de Juan no puede contar como la creencia de que "llueve". Así por ejemplo, si Juan dice creer que "llueve", pero al mismo tiempo él dice que si sale a la calle no se mojará y además no lleva paraguas ni gabardina, entonces nosotros tenemos buenas razones para poner en duda lo que Juan dice creer.

Vemos pues que las nociones de racionalidad y creencia están conectadas en el sentido de que si las creencias y demás actitudes proposicionales de un sujeto no están conectadas de maneras lógicas apropiadas, entonces resulta imposible atribuirle creencias. Y en tanto que no es posible atribuirle creencias, no puede entonces ser considerado como un agente cognoscitivo.

Sobre la base de lo anterior, Davidson afirma que ser racional<sup>1</sup> es una condición necesaria para contar como un agente cognoscitivo; por lo tanto, no pueden existir sujetos racionales que razonen de acuerdo con principios de razonamiento sustancialmente distintos de los nuestros. Consecuentemente, todo agente cognoscitivo comparte los mismos principios fundamentales de racionalidad, es decir, todo agente cognoscitivo debe razonar de acuerdo con principios que se desprenden de la lógica clásica, teoría de la probabilidad y teoría de la elección racional.

En contra de la visión estándar y de lo que era esperable dados los argumentos de Davidson acerca de la imposibilidad de un agente cognoscitivo cuyo razonamiento se desviara drásticamente de los principios normativos de racionalidad, los experimentos de razonamiento muestran que el razonamiento de los seres humanos se desvía

sistemáticamente de las formas prescritas que la visión estándar postula. Ante estos resultados, dos posiciones básicas pueden ser asumidas: o bien, los seres humanos somos irracionales, toda vez que violamos principios normativos ampliamente validados y establecidos; o bien, los estándares normativos que la visión estándar de la racionalidad postula no son, ni precondiciones para la adscripción de creencias, ni principios normativos apropiados para la evaluación del razonamiento humano.

La posición con la que me comprometo en el presente texto es con la segunda. Tratar de defender la tesis de que la visión estándar de la racionalidad es una concepción demasiado idealizada de la racionalidad como para poder ser alcanzada por los seres humanos, será entonces un objetivo central del presente escrito.

Para sustentar la tesis anterior apelaré aquí a los argumentos de Christopher Cherniak a favor de lo que él llama *racionalidad mínima*. Cherniak destaca el hecho de que los seres humanos somos criaturas con capacidades cognoscitivas finitas en el sentido de que tenemos una capacidad limitada de memoria, atención, disponemos de un tiempo finito, etc; por lo tanto, también tenemos una capacidad de razonamiento limitada. Por consiguiente, resulta absurda la pretensión de que los seres humanos razonemos de acuerdo con ciertos principios con los que, dadas nuestras capacidades cognoscitivas, simplemente *no podemos* hacerlo. Es por esto que Cherniak propone que: el *poder* razonar de acuerdo con ciertos principios de razonamiento, debe ser una condición necesaria para que esos principios puedan contar como principios normativos de razonamiento. En breves palabras, Cherniak propone que "*debe implica puede*".<sup>2</sup>

Ahora bien, si los experimentos de razonamiento nos muestran que la mayor parte de los sujetos participantes *no pueden* razonar de acuerdo con algunos de los principios normativos de razonamiento que la visión estándar postula, entonces -si aceptamos con Cherniak que *debe implica puede*- dichos principios no pueden contar como los estándares normativos correctos en la evaluación de dichos sujetos. Es así que la investigación psicológica sobre el razonamiento nos da elementos para sostener que la visión estándar es una concepción demasiado idealizada de la racionalidad y debe ser rechazada.

Pero la visión estándar, así como la concepción davidsoniana de la racionalidad, también pueden ser rechazadas sobre la base de otros argumentos. Dos presupuestos comunes implícitos tanto en la visión estándar como en las tesis de Davidson, son: primero, que los principios normativos de razonamiento deben ser principios universales, es decir, aplicables a la evaluación del razonamiento de todo sujeto en todo tiempo y lugar; segundo, que dichos principios son principios formales, es decir independientes de contenido.

A la luz de consideraciones filosóficas y empíricas sostendré que los dos presupuestos anteriores no se sustentan. Asentar esto será otro objetivo central en el presente trabajo.

En contra de la universalidad de los principios normativos de razonamiento pueden apuntarse dos argumentos filosóficos que nos llevan a cuestionar la idea de que deba haber un núcleo básico de principios de razonamiento forzosamente compartido por todos los sujetos para que la descripción intencional sea posible. Dichos argumentos abren la posibilidad para pensar que la diversidad cognoscitiva radical es plausible.

El primero de estos argumentos -que puede ser extraído desde las consideraciones que hace Cherniak en relación a cuáles serían las capacidades lógicas de un agente mínimamente racional- se apoya en un experimento mental que nos muestra la plausibilidad lógica de un agente racional que razone de acuerdo con principios de razonamiento diferentes de los nuestros. El segundo de estos argumentos, que debemos a Stephen Stich, da sustento a la idea de que la *similitud* en ciertos aspectos cognoscitivos entre el sujeto interpretado y el intérprete, es una condición necesaria para que la interpretación sea posible. Sin embargo, este argumento no pretende que haya similitud entre todos aquellos individuos que pueden ser considerados como agentes cognoscitivos. Por lo tanto, no requiere que haya un conjunto de principios de razonamiento compartidos por todo sujeto racional. Lo que se asume es simplemente que puede haber diferentes grados de similitud cognoscitiva entre los diferentes individuos, para individuos más similares entre sí la interpretación mutua será una tarea mucho más sencilla que para los individuos poco similares entre sí.

Además de los argumentos filosóficos recién esbozados, los resultados de los experimentos de razonamiento pueden también dar apoyo a la tesis de que los principios normativos de razonamiento no son universales. Puede subrayarse que la generalidad de los experimentos que se presentan en este trabajo muestran diferencias individuales en las capacidades de razonamiento de los sujetos. Algunos de estos experimentos muestran, además, diferencias individuales en lo que los sujetos *pueden* hacer en tareas de razonamiento, una vez que han sido sometidos a un proceso de aprendizaje en el que se intenta enseñarles a razonar de acuerdo con ciertos principios de razonamiento. A la luz de estos resultados experimentales y del compromiso asumido con la tesis de que "*debe implica puede*", es posible llegar a la siguiente conclusión: si no todos los sujetos *pueden* razonar de acuerdo con cierto principio de razonamiento, entonces ese principio no puede contar como un principio normativo de razonamiento para todos los sujetos, no obstante que pueda contar como un principio normativo de razonamiento para algunos sujetos. Por lo tanto, los principios normativos de razonamiento no son necesariamente universales.

En contra de la visión estándar que tiende a concebir a los principios normativos de razonamiento como principios formales -es decir, como un conjunto de reglas meramente sintácticas que pueden ser aplicadas a problemas de razonamiento con total independencia de los contenidos involucrados- puede argumentarse que: los resultados experimentales de la investigación en psicología del razonamiento muestran que, más que la estructura lógica o formal de los problemas de razonamiento planteados en los experimentos, es el contenido y el contexto de dichos problemas así como el conocimiento previo de los sujetos, lo que juega un papel determinante en las respuestas emitidas. Estos resultados han llevado a psicólogos, como Patricia Cheng y Leda Cosmides (entre otros muchos) a proponer teorías acerca de la competencia humana de razonamiento, que postulan principios de razonamiento, subyacentes a dicha competencia, dependientes de contenido o específicos a un dominio<sup>3</sup>. Si aceptamos lo que estas teorías están proponiendo, esto es, que la competencia humana de razonamiento está constituida, fundamentalmente, por principios de dominio específico; entonces, consecuentemente, hemos de considerar que gran parte de los principios normativos de razonamiento deben estar también relativizados a dichos dominios. Esto último debe ser así, toda vez que los principios conforme a los cuales los seres humanos *podemos* razonar son, fundamentalmente, de dominio específico.

En conclusión, el objetivo de este trabajo será mostrar, sobre la base de algunas consideraciones filosóficas, así como de los resultados de la investigación empírica en psicología del razonamiento humano, que la visión estándar de la racionalidad y con ella concepciones de racionalidad como la que Donald Davidson propone, deben ser rechazadas. Las razones para ello son básicamente las siguientes:

- 1) Asumen una concepción demasiado idealizada de la racionalidad, no aplicable a seres con capacidades cognoscitivas finitas como nosotros.
- 2) Asumen principios universales de racionalidad; no obstante, dados ciertos argumentos tanto filosóficos como empíricos, tenemos razones para creer que los principios normativos de razonamiento no son necesariamente universales, por lo tanto, no todo principio normativo de razonamiento puede ser aplicado a todo sujeto, en todo tiempo y lugar
- 3) Asumen principios formales de racionalidad, no obstante, tenemos aquí también razones de corte empírico para creer que gran parte de los principios normativos de razonamiento deben estar relativizados a dominios específicos.

En aras de mostrar lo anterior, el presente trabajo se organiza como sigue: en los dos primeros capítulos se presenta un panorama general de lo que ha sucedido en el ámbito de la investigación psicológica sobre el razonamiento humano en las últimas décadas. Veremos

entonces el papel clave que juegan factores como el contenido y el contexto de un problema de razonamiento, así como el conocimiento previo del sujeto, en la resolución del mismo. De este modo podremos darnos cuenta de que la estructura lógica de un problema no es lo único que cuenta para su solución. Veremos también que, aunque para algunos casos, ciertas tendencias en las respuestas de los sujetos pueden ser señaladas, suele haber también casos donde las respuestas son muy divergentes, lo cual da sustento a la tesis de que hay diferencias individuales sustantivas en la manera de razonar de los distintos sujetos. En el capítulo tres se hace una exposición de la concepción davidsoniana de la racionalidad y los argumentos esgrimidos por Davidson a favor de ésta. Veremos cómo es inherente y esencial a esta concepción la tesis de la universalidad de los principios normativos de razonamiento. El objetivo central del capítulo será atacar dicha tesis. Este ataque se hace sobre la base exclusiva de consideraciones filosóficas, echando mano de argumentos desarrollados por Christopher Cherniak y Stephen Stich. Al mismo tiempo irán quedando claras las razones por las cuales la concepción davidsoniana de la racionalidad es una concepción demasiado idealizada.

En el capítulo cuatro se presentan las dos concepciones fundamentales desarrolladas en el seno de la epistemología en relación a los fundamentos de la normatividad epistémica: la concepción tradicional y la concepción naturalizada. Se mostrará cómo los argumentos de Quine dieron lugar al surgimiento de un enfoque naturalizado en la epistemología el cual considera que los resultados de la investigación empírica en las ciencias particulares son relevantes para la indagación filosófica. Asumiendo un compromiso con un enfoque naturalizado de la epistemología proseguiré a analizar la repercusión de los resultados de los experimentos en psicología del razonamiento para la tesis davidsoniana de que los principios normativos de razonamiento deben ser universales. En este punto haré referencia a algunos experimentos psicológicos adicionales. Argumentaré que a la luz de los resultados de tales experimentos, la tesis de la universalidad de los principios normativos de razonamiento no puede sostenerse. Por otra parte también analizaré la repercusión de los resultados experimentales para la tesis de que los principios normativos de razonamiento son principios de naturaleza formal y por ende, aplicables a la evaluación del razonamiento sin importar los contenidos sobre los que versa el razonamiento ni los contextos en que éste se da. Argumentaré que a la luz de dichos resultados tenemos buenas razones para creer que los principios normativos de razonamiento no son necesariamente principios formales, toda vez que la gran mayoría de los principios a que hemos de apelar en la evaluación del razonamiento de seres humanos legos en materia de lógica y teoría de la probabilidad deben estar relativizados, no sólo a contenidos específicos, sino también a sujetos y contextos específicos.



## NOTAS.

---

<sup>1</sup> La afirmación de que un sujeto S es racional puede ser entendida de dos modos:

1) Por un lado, puede decirse que un sujeto S es racional en el sentido de que el sistema de creencias de S, en un tiempo específico T, constituye un conjunto coherente. 2) Por otro lado, puede decirse que un sujeto S es racional en el sentido de que las inferencias que S realiza son inferencias aceptables desde el punto de vista de los principios normativos de razonamiento que se consideran aceptables. En el primer sentido, la racionalidad puede ser entendida como estática; en el segundo sentido se trata de una racionalidad procedural o dinámica.

¿En cuál de estos sentidos usa Davidson el concepto de racionalidad cuando él afirma que, a fin de poder ser interpretado y contar como un agente cognoscitivo un sujeto debe ser racional? La respuesta a esta pregunta es controversial. No obstante, Davidson ha sido interpretado, en gran medida, como asumiendo este segundo sentido, y así es cómo se le interpreta aquí.

<sup>2</sup> Debe quedar claro aquí, que cuando se habla de que un cierto individuo “*puede*” razonar de acuerdo con un cierto principio de razonamiento, no se quiere decir que dicho principio se encuentra necesariamente en la competencia de razonamiento del individuo en cuestión, pues puede ser el caso que el principio de razonamiento no se encuentre en la competencia de razonamiento de tal individuo, pero que él pueda llegar a razonar de acuerdo con el principio tras un proceso de enseñanza-aprendizaje.

<sup>3</sup> La noción de “especificidad de dominio” ha sido entendida de muchas maneras: 1) para algunos autores, una regla o mecanismo inferencial es específico de dominio cuando hay restricciones en el tipo de información que dicho mecanismo puede aceptar como entradas (*input-specificity*); 2) para otros autores, un mecanismo inferencial es de dominio específico cuando ese mecanismo realiza una función que ningún otro mecanismo puede realizar (*function-specificity*). La noción que se usará en el presente texto es la primera, en tanto que ésta es la noción que, según propone García Aguilar, prevalece en el ámbito de la psicología del razonamiento humano. Sin embargo, la caracterización de la noción de especificidad de dominio que aquí se maneja, no estaría completa si no se hace la siguiente precisión: en el caso de una regla o mecanismo inferencial que es específico de dominio en el primer sentido (*input-specificity*), es posible aun distinguir dos posibilidades: 1) por un lado, puede ocurrir que un mecanismo inferencial admita sólo ciertos tipos de información como entradas, porque la regla subyacente en la operación de dicho mecanismo no está formulada de manera formal o completamente general; 2) por otro lado, puede ocurrir que un mecanismo inferencial admita sólo ciertos tipos de información como entradas, no por la especificidad de la regla inferencial subyacente (dicha regla puede estar formulada de manera completamente formal o general), sino más bien por razones que tienen que ver con la arquitectura cognoscitiva global y las consecuentes restricciones al flujo de información dentro del sistema, lo cual puede aislar al mecanismo inferencial en cuestión de ciertos tipos de información.

## **CAPITULO UNO.**

### **PRIMEROS EXPERIMENTOS SOBRE RAZONAMIENTO EN PSICOLOGÍA.**

El objetivo del presente capítulo es dar un panorama general del campo de la investigación psicológica sobre el razonamiento. Con este fin se hace una revisión más o menos exhaustiva de los primeros y más importantes experimentos sobre razonamiento en psicología, mismos que pueden ser ubicados en un periodo que va de la década de los 60's a la década de los 80's. Sin embargo, antes de entrar de lleno a esta revisión, es necesario dejar en claro algunas nociones fundamentales que se encuentran en la base de la investigación psicológica sobre el razonamiento.

Es trivial decir que la investigación empírica del razonamiento humano está esencialmente encaminada al desarrollo de una teoría del razonamiento que nos permita interpretar y organizar adecuadamente la gran cantidad de datos generados por los experimentos. No obstante, no es en modo alguno trivial dejar en claro y establecer desde un principio, cual es el tipo de teoría del razonamiento a que se pretende llegar.

La historia de la psicología ha estado marcada por una exacerbada oposición entre quienes creen que la conducta manifiesta -y por lo tanto observable- del ser humano, es de lo único que puede y debe dar cuenta una teoría psicológica (paradigma conductual); y quienes creen que la conducta observable es interesante en tanto que puede arrojar luz sobre leyes mentales subyacentes y que son éstas últimas de las que debe dar cuenta una teoría psicológica (paradigma cognoscitivo).

El tipo de teoría que los psicólogos del razonamiento intentan desarrollar se ubica claramente dentro de esta última postura. Por lo tanto, cuando ellos investigan el razonamiento humano no se interesan intrínsecamente en la habilidad para la solución de problemas que ciertas personas en particular puedan exhibir en situaciones concretas en la vida cotidiana. No obstante, sí se interesan en las habilidades manifiestas de las personas para enfrentar tareas de razonamiento, en tanto que éstas constituyen el único medio para dilucidar cuáles son las reglas o principios subyacentes que pueden dar cuenta de las capacidades exhibidas. En tanto que la habilidad para razonar y solucionar problemas, tal como ésta se manifiesta en el nivel conductual, es el resultado de la interacción de los mecanismos encargados del razonamiento con otros muchos mecanismos cognoscitivos tales como atención, memoria, afectividad, etc. que también entran en operación cuando un

sujeto enfrenta y resuelve un determinado problema de razonamiento; los juicios y las inferencias que la gente realiza a un nivel manifiesto no constituyen una vía directa para establecer cuáles son verdaderamente las reglas que subyacen a sus habilidades de razonamiento. Es por esto que los psicólogos del razonamiento, interesados en dilucidar cuáles son genuinamente las reglas o principios de razonamiento internamente representados que subyacen a la capacidad de los seres humanos para razonar, deben separar los efectos de estos otros mecanismos que pueden estar operando, de los efectos que resultan de la operación de él, o de los mecanismos encargados del razonamiento.

Una vez logrado lo anterior, los psicólogos estarían en condición de proponer una teoría general sobre la capacidad de los seres humanos para el razonamiento. La descripción de tal capacidad trascendería la habilidad que puede ser exhibida, o no, por seres humanos específicos en instancias particulares. Una analogía permitirá dejar más claro este punto. Si un estudioso de las capacidades perceptivas humanas intenta dar cuenta de las capacidades visuales del ser humano en contraposición, por ejemplo, con las de los felinos, él intentará describir una capacidad general que va más allá de las destrezas visuales de seres humanos particulares, algunos de los cuales pueden padecer miopía, ceguera o enfermedades transitorias del ojo que afectan una visión óptima. Lo que este investigador desea describir es más bien las capacidades visuales generales del ser humano normal en condiciones óptimas de visión. Esta distinción entre una capacidad que se exhibe en un momento determinado en una situación particular por un individuo particular, y aquella que puede ser caracterizada como una capacidad general propia de digamos la especie humana, es la antesala de una importante distinción que de manera sustancial ha orientado las investigaciones psicológicas sobre el razonamiento y delimitado la meta explicativa específica a que ha tendido la teorización sobre el razonamiento. Nos referimos a la distinción *competencia-ejecución* la cual abordaremos en seguida.

## **1. COMPETENCIA Y EJECUCIÓN**

La distinción competencia-ejecución ha sido traída al campo de estudio del razonamiento desde la Lingüística. Los lingüistas, desde Chomsky, suelen distinguir entre el conocimiento lingüístico subyacente al uso y entendimiento del lenguaje y la conducta lingüística efectiva.

El conocimiento lingüístico (competencia) es entendido como un conjunto de reglas abstractas o mecanismos que subyacen a la operación de las estructuras cognoscitivas encargadas del procesamiento del lenguaje. Se considera que estas reglas o estos mecanismos, aunque no accesibles a la conciencia, dan cuenta de la variedad de las

capacidades lingüísticas exhibidas por los seres humanos, tales como generar y comprender oraciones, detectar casos de falla gramatical, producción de paráfrasis, etc.

La conducta lingüística efectiva (ejecución) es la manifestación lingüística que resulta de la expresión de la competencia lingüística, y además, de otros mecanismos psicológicos como atención, memoria, afectividad o factores situacionales tales como estrés, ruido, uso de drogas, etc., que pueden interactuar con los mecanismos del lenguaje.

Considérese, por ejemplo, el caso de un comerciante mexicano que ofrece entusiastamente su mercancía pregonando que: “se venden calcetines para caballero de lana”. Aunque el comerciante ha cometido un claro error gramatical, a casi nadie se nos ocurrirá poner en duda su competencia para hablar el español, más bien tendemos a considerar que el comerciante no ha puesto suficiente atención a lo que está diciendo. No obstante, si este mismo error gramatical fuera cometido por un comerciante extranjero que apenas entendiera mínimamente el español seguramente atribuiríamos su error, no a una mera falla de atención, sino más bien a su deficiente competencia como hablante del español.

Este ejemplo puede darnos ya idea de una de las más importantes ventajas explicativas de la distinción competencia-ejecución, a saber, que nos permite –en ciertos casos- dar cuenta de fallas en la conducta lingüística, apelando a errores de ejecución que son externos a la competencia. De este modo, en determinadas situaciones, la producción de errores no compromete o pone en entredicho la competencia lingüística del hablante oyente. Se es así congruente con el hecho de que se puede ser competente en el manejo de una lengua aún cuando algunas veces se cometan errores.

La distinción competencia-ejecución es importante, además, porque permite dar explicaciones que van más allá de la conducta lingüística. Como ha sido destacado por Chomsky, la capacidad humana para entender y emitir nuevas expresiones del lenguaje es infinita, no obstante que, durante el proceso de aprendizaje, la exposición efectiva a contingencias lingüísticas es limitada y subdetermina las reglas que subyacen a la producción del lenguaje. De aquí que una explicación completa de la capacidad humana para el uso del lenguaje deba ir más allá de los datos (o inputs) a los que ha estado expuesto el individuo y más allá de su conducta efectiva, es decir, más allá de la ejecución lingüística exhibida por los seres humanos.

Se han aportado ya los ingredientes esenciales para comprender el concepto de *competencia lingüística*, sin embargo hace falta todavía mencionar una noción que es central en la caracterización chomskyana de dicho concepto. Esta noción central es esa de “*órgano mental*”. Chomsky considera que la competencia lingüística puede ser entendida como un órgano mental en tanto que, al igual que un órgano corporal como el corazón o el ojo, la

competencia humana para el lenguaje posee (entre otras) dos características esenciales:

1) Tiene una estructura básica genéticamente determinada, es decir, dicha competencia está innatamente especificada, y 2) constituye una unidad funcional. Esto quiere decir que las estructuras cognoscitivas encargadas del lenguaje sólo están comprometidas en esta función específica.

Una vez establecida la distinción competencia-ejecución y su importancia en el ámbito de la teoría lingüística veamos ahora como esta distinción puede ser transferida al ámbito del estudio del razonamiento humano.

En primer lugar, se puede considerar a la *competencia de razonamiento* como un conjunto de mecanismos o de reglas (no conscientes) que subyacen a la operación de las estructuras cognoscitivas encargadas del proceso de razonamiento. En segundo lugar, para ciertos casos en los que las personas puedan cometer “errores de razonamiento”, será posible en ocasiones, apelar a fallas de ejecución antes que suponer que ellas piensan o actúan irracionalmente, y antes que realizar cualquier cambio en nuestra teoría de la competencia de razonamiento. Así, por ejemplo, si Juan cree que ‘si hoy es lunes entonces tiene cita con el dentista’ y además cree que ‘hoy es lunes’, pero al mismo tiempo asevera que su cita con el dentista no es hoy, entonces nosotros, que sabemos de sus creencias y escuchamos su aseveración podríamos pensar que él ha cometido un error, una posible distracción u olvido, en vez de pensar que el problema es atribuible a la competencia de Juan para el razonamiento. Por supuesto, siempre que se apelé a un factor de ejecución para dar cuenta de fallas en el razonamiento, debe haber evidencia plausible -en términos de alguna teoría psicológica (o tal vez psicofisiológica)- a favor de dicho factor, de lo contrario, la falla debe ser imputada a la competencia de razonamiento de los sujetos.

Vemos pues que la ventaja explicativa de la distinción competencia-ejecución arriba señalada en relación a la competencia lingüística, se mantiene aquí: asumir que los seres humanos razonamos de acuerdo con ciertas reglas implica que nuestra conducta puede ser explicada por su ajuste sistemático a tales reglas aún cuando algunas veces nos desviemos de ellas, pues tales desviaciones pueden ser entendidas como meros errores atribuibles a factores interfirientes que no tienen nada que ver con nuestra competencia de razonamiento. Por último, se habrá de asumir que la *competencia de razonamiento* puede, del mismo modo que la competencia lingüística, ser entendida como un *órgano mental*. Esto implica que la habilidad de razonamiento humana emerge de la operación de un *órgano de razonamiento* que es funcionalmente especializado y que está innatamente especificado.

En relación a este último paso en la extrapolación de la distinción competencia-ejecución al ámbito de estudio del razonamiento humano permítaseme comentar lo siguiente: si bien

podemos estar de acuerdo en que la competencia de razonamiento constituye una unidad funcional en el sentido de que las estructuras cognoscitivas encargadas del razonamiento sólo están comprometidas en esta función, me parece que la tesis de una competencia de razonamiento innatamente especificada, es ampliamente discutible. Primero que todo cabría analizar qué significa que algo sea innato. Originalmente se tendió a identificar *innato* con lo que está *directamente especificado en los genes*. Sin embargo, avances en el entendimiento de la forma en que se expresan los genes, ha llevado a los teóricos del desarrollo a considerar que la visión *preformacionista* según la cual, los resultados finales del desarrollo están explícitamente codificados en los genes, es improbable. Ahora se conoce que la vía del genotipo al fenotipo puede ser muy indirecta y requiere de múltiples y sutiles interacciones entre genes y el ambiente, ya sea interno o externo al organismo. Por tal razón, la identificación de *innato* con *lo que está directamente especificado en los genes* parece ahora inapropiada, pues nos conduciría a la inaceptable consecuencia de que casi nada sería innato. Pero si esto es así, cabe ahora preguntarse ¿qué se debe entender cuando se afirma que una cierta característica es *innata*?

Johnson y Morton (1991) consideran que es útil distinguir entre los varios niveles de interacción de los genes con su ambiente ya sea molecular, celular, o el ambiente externo al individuo. Una vez establecidas estas distinciones, Elman et. al. sugieren, se puede reservar el término *innato* para referirse a “*cambios que surgen como un resultado de interacciones que ocurren dentro del mismo organismo durante la ontogenia. Esto es, interacciones entre los genes y sus ambientes molecular y celular sin recurso a información externa al organismo*”. (Elman, et. al., 1996. p. 22)

Ahora bien, sí a la luz de esta noción de *innatez*, se analizan las evidencias aportadas recientemente en el campo de la neurofisiología y la computación neural, tendremos buenas razones para considerar que el desarrollo de las capacidades cognoscitivas superiores de los seres humanos sea el resultado exclusivo de factores internos al organismo, es un hecho poco plausible. En *Rethinking Innateness*, Elman et.al. hacen un acopio de gran parte de la evidencia a que se ha aludido.

En el campo de las neurociencias, numerosas investigaciones han apoyado la tesis de que la corteza cerebral es fundamentalmente un *órgano de plasticidad* y que la especificidad funcional de ciertas áreas de la corteza es adquirida en el curso del desarrollo. Se ha mostrado, por ejemplo, que si la información de los ojos es reinervada y dirigida a la corteza auditiva en periodos tempranos del desarrollo, la corteza que se ha tomado prestada empezará a procesar estímulos visuales; o que cambios en la superficie corporal, debidos a amputación o parálisis, producen una reorganización de la corteza somato-sensorial en primates adultos (p. 3).

Por otra parte, en el campo de la computación neural, se han podido construir modelos de redes con propiedades de respuesta sumamente especializadas que emulan, por ejemplo, propiedades de la corteza visual: detección de bordes, de movimiento, etc., sin embargo, estas propiedades no tienen que ser pre-especificadas, *“ellas emergen naturalmente e inevitablemente de células las cuales no están inicialmente comprometidas, simplemente como una función de una simple regla de aprendizaje y exposición a la estimulación”*. (Elman, et.al. p. 5). Esto muestra que aún habilidades complejas pueden ser aprendidas sin ser pre-especificadas, es decir, sin ser innatas.

En resumen, se puede concluir que el asumir un compromiso con la tesis de que la competencia de razonamiento humana constituye una unidad funcional, no conlleva necesariamente un compromiso con la tesis de que esa competencia sea innata.

## **2. PRIMEROS EXPERIMENTOS SOBRE RAZONAMIENTO.**

### **2.1 VISIÓN ESTÁNDAR DE LA RACIONALIDAD.**

En la sección anterior se estableció el precedente de que los esfuerzos en la investigación sobre el razonamiento están dirigidos a desarrollar una teoría de la *competencia* humana de razonamiento en el sentido ya explicitado.

Es importante ahora ubicar otro precedente importante. Si bien es cierto que la investigación experimental sobre el razonamiento está encaminada al desarrollo de una teoría descriptiva capaz de dar cuenta de la manera en que los seres humanos efectivamente razonamos, también es cierto que esta tarea descriptiva ha sido realizada -de hecho- apelando, de un modo u otro, a ciertos principios normativos, es decir, principios que hacen alusión a las maneras “correctas” de razonar. Esto es así porque cuando un investigador declara, por ejemplo, que las respuestas de un grupo de personas a X problema de razonamiento son incorrectas o deficientes y busca la manera de dar cuenta de esta “falla” mediante algún factor externo a la competencia, está ya inevitablemente asumiendo ciertos principios normativos. De la misma manera lo hace así cuando juzga que las respuestas de los sujetos son correctas. Si nos queda claro que el estudio descriptivo del razonamiento no ha procedido completamente al margen de consideraciones normativas, es importante entonces conocer cuáles han sido los principios normativos que han orientado la investigación experimental sobre el razonamiento así como la manera en que la racionalidad fue entendida en la época de los primeros experimentos psicológicos sobre el razonamiento.

Como se hará evidente más adelante, la visión sobre la racionalidad que estuvo detrás de la investigación sobre el razonamiento, a lo menos durante –aproximadamente- las tres

primeras décadas (de los 60,s a los 80's), es la que ahora ha venido a conocerse como *visión estándar de la racionalidad*.

Hay tres supuestos básicos que subyacen a lo que se conoce como *visión estándar de la racionalidad* (ver Stein, 1996):

- 1) Ser racional es razonar de acuerdo con principios consistentes con las reglas de la lógica clásica, teoría de la probabilidad, teoría de la elección racional, etc., y tales reglas constituyen los principios normativos de razonamiento aplicables a todo ser racional.
- 2) La competencia humana de razonamiento está constituida por un sistema de reglas que, si no son las reglas de la lógica o teoría de la probabilidad propiamente dichas, sí deben ser consistentes con ellas.
- 3) Las reglas que constituyen la competencia humana de razonamiento son reglas de carácter formal, es decir, independientes de contenido. Esto significa que tales reglas nos permiten hacer razonamientos en cualquier ámbito temático, pues lo único relevante en la secuencia de nuestros razonamientos es su forma lógica, no su contenido.

Sin embargo, la *visión estándar de la racionalidad* debe enfrentar el enorme desafío que la investigación experimental en psicología del razonamiento le plantea: por un lado, una gran cantidad de experimentos parecen mostrar que el razonamiento humano se desvía sistemáticamente de los principios normativos de razonamiento que se desprenden de la lógica y teoría de la probabilidad, contraviniendo así el segundo supuesto básico de la visión estándar; por otro lado, los resultados experimentales sugieren que el contenido de un problema no es irrelevante para la representación que el sujeto se hace tanto del problema como de su solución, contraviniendo así el tercer supuesto básico de la visión estándar.

Estos resultados han llevado a muchos a cuestionar la racionalidad humana, toda vez que, según ellos, los experimentos han mostrado fallas sistemáticas en el razonamiento de los sujetos, mismas que no pueden ser consideradas como meras fallas de ejecución, sino que más bien parecen ser atribuibles a la competencia de razonamiento de los sujetos.

Otros consideran que las fallas en cuestión sí pueden ser explicadas como meras fallas de ejecución externas a la competencia; intentan salvar así la tesis de que el ser humano es racional, claro está, desde la perspectiva estándar de la racionalidad.

Aunque es un hecho que los resultados de los experimentos son polémicos, será mejor que juzguemos por nosotros mismos y obtengamos nuestras propias conclusiones. Vayamos pues a la revisión de los experimentos.

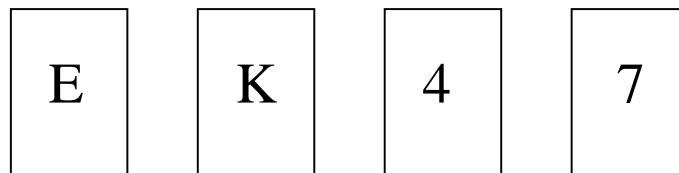


## 2.2 LA TAREA DE SELECCIÓN DE WASON.

Uno de los experimentos pioneros en el estudio del razonamiento deductivo fue realizado por Peter Wason (1966). Este experimento muestra las dificultades que los sujetos humanos enfrentan para hacer inferencias que involucran el manejo de una regla condicional de la forma “si P entonces Q”.

El experimento procedía como sigue:

Se presentaban a los sujetos 4 cartas desplegadas sobre una mesa, cada una de las cuales tenía una letra de un lado y un número del otro, aunque los sujetos sabían esto, ellos sólo podían ver el lado de las cartas que estaba hacia arriba. Sobre el frente de la primera carta aparecía una vocal (“E”), sobre el frente de la segunda aparecía una consonante (“K”), sobre el frente de la tercera aparecía un número par (“4”), y sobre el frente de la cuarta carta aparecía un número impar (“7”).



Despliegue de cartas utilizadas por Wason, 1966.

Se les presentaba también a los sujetos la siguiente regla condicional: **“Si hay una vocal en un lado de la carta entonces hay un número par sobre el otro lado”** y se les pedía seleccionar aquéllas cartas que era necesario voltear para descubrir si el experimentador mentía al afirmar la regla condicional anterior.

De acuerdo con el análisis lógico y con la tabla de verdad del condicional material, la selección correcta en este caso son las cartas “E” y “7”. Veamos porqué.

P	Q	$P \rightarrow Q$
V	V	V
V	F	F
F	V	V
F	F	V

Tabla de verdad del condicional material.

Si se simboliza la proposición “Hay una vocal de un lado” como P y la proposición “Hay un número par del otro lado” como Q, entonces las cartas presentadas pueden simbolizarse

como sigue: "E" (P), "K" ( $\neg$ P), "4" (Q), "7" ( $\neg$ Q).

Puesto que la única instancia falsificadora para un condicional es el caso en que el antecedente es verdadero y el consecuente es falso, las únicas cartas relevantes para la falsificación de la regla son las cartas "E" (P) y "7"( $\neg$ Q).

No obstante, menos del 10% de los sujetos a los que se presenta dicha tarea -la que ha venido a conocerse como tarea de selección- hacen la selección correcta, es decir, seleccionan las cartas que corresponden a P y  $\neg$ Q. La selección mayoritaria (aproximadamente un 59%) es "E"(P) y "4"(Q). Otras posibles respuestas registradas en relación a la tarea de selección, según datos aportados por Wason & Johnson-Laird (1972), son:

P:	42%
P, Q y $\neg$ Q:	9%
Otras:	13%

Por mi parte, en una aplicación de la tarea de selección a un grupo de 43 estudiantes de la Facultad de Psicología de la UNAM, de primer semestre, obtuve los siguientes resultados:

P y Q:	34%
P:	26%
$\neg$ P:	7%
$\neg$ P y $\neg$ Q:	7%
Q y $\neg$ Q:	7%
Todas las cartas:	5%
Ninguna carta:	5%
Q:	5%
$\neg$ Q:	2%
$\neg$ P y Q:	2%

Puesto que la selección mayoritaria es P y Q puede considerarse que, en la tarea de selección los sujetos cometen, fundamentalmente, dos tipos de errores: 1) por un lado, se selecciona la carta "4"(Q) la cual no es relevante para la falsación de la regla dado que una "K" del otro lado no es inconsistente con la regla; 2) por otro lado no se selecciona "7"( $\neg$ Q) la cual si es una carta relevante, pues una "E" del otro lado sería suficiente para falsificar la regla.

Como ya se había mencionado, desde la perspectiva de la visión estándar de la racionalidad, es posible asumir dos posturas ante los resultados obtenidos en experimentos de razonamiento tales como la tarea de selección. Por un lado puede concluirse que el ser

humano es irracional, toda vez que sus desviaciones del buen razonamiento no pueden atribuirse a factores de ejecución. Por otro lado, puede argumentarse que el mal razonamiento en las pruebas experimentales sí puede atribuirse a factores ajenos a la competencia de razonamiento de los sujetos y tratar de explicar estas *fallas* en términos de factores de ejecución. De probarse esto último, sería posible salvar la tesis de la racionalidad humana, o dicho en términos más precisos, la tesis de que los seres humanos somos racionales en el sentido *estándar* de la racionalidad. Lamentablemente, como veremos en los experimentos siguientes, los errores cometidos por los sujetos en la tarea de selección parecen ser sistemáticos y por ende, atribuibles únicamente a la competencia de razonamiento de los sujetos.

### 2.3 INVESTIGACIONES SUBSECUENTES SOBRE LA TAREA DE SELECCIÓN.

Tras el impacto que causara la mala ejecución de los sujetos en la tarea de selección, muchas manipulaciones experimentales se han realizado para tratar de entender la naturaleza de los errores y para producir alguna facilitación de la tarea, si fuera posible.

En el intento de aislar e identificar los factores que pudieran dar cuenta de la pobre ejecución de los sujetos en la tarea de selección, se han manipulado e investigado variables como las siguientes:<sup>4</sup>

#### A. Errores de procedimiento.

En primer lugar se trató de manipular lo que se consideraba podrían ser posibles errores de procedimiento. Por ejemplo, se pensó que la forma de presentación de la regla podría sugerir una relación temporal o causal entre antecedente y consecuente, lo cual implicaría que los sujetos se formaban una representación de la regla con un sentido mucho más amplio que aquel que corresponde al del *condicional material*. Recuérdese que, en el contexto de la lógica, el condicional material no sugiere ningún tipo de conexión entre antecedente y consecuente que vaya más allá de afirmar que no puede ocurrir que el antecedente sea verdadero y el consecuente sea falso. Es así que el sentido del condicional material es mucho más restringido que el sentido de otros muchos tipos de condicional que es posible encontrar en el discurso ordinario. Debe pues quedar claro que el significado del “si...entonces” de las reglas condicionales usadas en los experimentos iniciales con la tarea de selección refieren al “si...entonces” lógico y por lo tanto, su sentido es ese del condicional material.

Para eliminar la posibilidad de que los sujetos pudieran estar haciendo otras interpretaciones de la regla, los experimentadores modificaron la forma de presentación de ésta: “cada carta la cual tiene un triángulo rojo sobre un lado tiene un círculo azul sobre el otro”, esto elimina la

posibilidad de una interpretación donde los términos de la regla estén relacionados secuencial o causalmente. En este caso, la frecuencia del error de seleccionar Q se redujo, pero no hubo ningún efecto sobre el error de omitir  $\neg Q$ . (Wason, 1969).

Se creyó también posible que los sujetos estuvieran confundiendo ambos lados de las cartas, pensando en “el otro lado de la carta” como el lado que está hacia abajo. Para evitar esto se presentó toda la información de un sólo lado de la carta, tapando la información que en los previos experimentos se encontraba del otro lado. (Wason & Johnson-Laird, 1970; Goodwin & Wason, 1972). La ejecución tampoco mejoró en este caso. Así, las fallas de razonamiento no parecían atribuibles a estos errores de procedimiento.

#### B. Familiaridad con la regla.

Un posible factor que hacía difícil la ejecución de la tarea podría ser la falta de familiaridad de los sujetos con la regla. Así, antes de realizar la tarea de selección, se permitió a los sujetos tener una experiencia previa tanto con la regla como con un conjunto de cartas (24 en total) que instanciaban ejemplos de los cuatro tipos de cartas (P,  $\neg P$ , Q,  $\neg Q$ ). Esta experiencia previa se introdujo de dos formas diferentes, cada una de las cuales se aplicó a dos grupos experimentales independientes. A los sujetos en el llamado “grupo de evaluación” se les permitió voltear las cartas y se les pidió evaluar si la regla condicional dada era verdadera o falsa en relación a cada una de ellas. A los sujetos en el llamado “grupo de construcción” no se les permitió voltear las cartas, sólo se les pidió que imaginaran y enunciaran qué valor en el otro lado de la carta haría la regla condicional verdadera, o qué valor la haría falsa. Después de esta tarea previa, la cual permitió a los sujetos familiarizarse con la estructura lógica del problema, los sujetos realizaron la tarea de selección en relación con la misma regla con la que ya se habían familiarizado previamente. El resultado fue que no hubo una facilitación significativa de la tarea en relación a experimentos anteriores. (Wason y Shapiro, 1971).

De haberse producido alguna facilitación de la tarea debida a la experiencia previa introducida en este experimento, hubiera sido posible dar una explicación de las fallas de los sujetos registradas en los experimentos anteriores, en términos de algún tipo de factor de ejecución. En este caso en particular posiblemente podría argüirse que la complejidad de la estructura lógica de la tarea requería de parte de los sujetos un poco más de tiempo y manejo de la regla a fin de comprender su estructura plenamente, y que una vez logrado esto, la mejora observada en la ejecución de los sujetos era evidencia de que su competencia de razonamiento en relación a reglas de tipo condicional era acorde con los principios normativos de razonamiento. No obstante no ocurrió así.

### C. Limitaciones de memoria.

Otra posible explicación de los resultados era que la tarea producía una sobrecarga cognoscitiva, es decir, que el análisis combinatorial requerido en la realización de la tarea demandaba mucha más capacidad de memoria e imaginación por parte de los sujetos de la que ellos podían exhibir. De ser esto lo que estaba ocurriendo, podría apelarse a una teoría psicológica plausible sobre las limitaciones de la memoria operativa humana<sup>5</sup>, para dar cuenta de la deficiente ejecución de los sujetos en la tarea de selección. De tal manera se evitaría atribuir su mala ejecución a fallas en su competencia de razonamiento.

Con el fin de explorar la posibilidad de una sobrecarga cognoscitiva como un factor asociado a la mala ejecución de los sujetos, Goodwin y Wason (1972) dieron a los sujetos un conjunto de cartas plenamente reveladas que podían ser examinadas mientras se realizaba la tarea de selección con un conjunto equivalente de cartas. Las cartas plenamente reveladas podían subsanar las limitaciones de memoria o imaginación de los sujetos. Sin embargo, esto no produjo ninguna facilitación significativa de la tarea. Este resultado permitió descartar la tesis de que una sobrecarga de la memoria impidiera a los sujetos realizar el análisis combinatorial necesario, sugiriendo, además, que la necesidad de tal análisis ni siquiera surgía. Al parecer, los sujetos simplemente basaban sus juicios en lo que ellos podían ver, consideraban así relevantes sólo las cartas mencionadas en la regla. Este sesgo o tendencia en el razonamiento de los sujetos es analizado en el siguiente apartado.

### D. Énfasis en la tarea de verificación.

Una posible explicación del error de omitir  $\neg Q$  era que los sujetos se centraban en la tarea de verificar la regla, desatendiendo totalmente a su falsificación. En consecuencia se realizaron experimentos, que haciendo cambios en las instrucciones, trataban de hacer énfasis en la tarea de falsificación (Hughes, 1966; Wason, 1968). Así, en vez de pedir a los sujetos seleccionar las cartas que es necesario voltear a fin de conocer si la regla es verdadera o falsa<sup>6</sup>; se les pedía seleccionar las cartas que es necesario voltear a fin de saber si la regla es falsa, o si la regla ha sido rota, o si la regla ha sido violada, o si lo expresado por el experimentador es una mentira, etc.. Este cambio en las instrucciones no tuvo efectos significativos.

Otra manipulación experimental dirigida a enfatizar la falsación fue realizada por Wason (1968). El experimento procedió como sigue: primero los sujetos realizaron la tarea de selección; después, en una segunda etapa, se pidió a los sujetos que dijeran qué valores asociados con P, Q,  $\neg P$ ,  $\neg Q$ , harían la afirmación condicional falsa (a este procedimiento se le llamó "proyección de falsedad"). Tras esto, los sujetos revisaron la tarea de selección que previamente habían realizado, es decir, se les permitió cambiar sus respuestas si así lo

deseaban. Una vez más no hubo una facilitación significativa. Ante este resultado, Wason propuso que el error de omitir  $\neg Q$  podía deberse a una tendencia aprendida de los sujetos a esperar una relación de verdad o correspondencia entre oraciones y estados de cosas, centrándose así en las cartas que son mencionadas en la oración condicional.

Una propuesta análoga a la de Wason para explicar el error de omitir  $\neg Q$ , fue posteriormente sostenida por Jonathan Evans (1972). Él utilizó un conjunto de figuras geométricas coloreadas suministrando todas las posibles combinaciones de casos: V V, V F, F V, F F; (V=verdadero, F=falso). Y reglas de la forma:

- 1) Si **hay** un triángulo rojo sobre la izquierda entonces **hay** un cuadrado verde sobre la derecha.
- 2) Si **no hay** un triángulo rojo sobre la izquierda entonces **hay** un cuadrado verde sobre la derecha.
- 3) Si **hay** un triángulo rojo sobre la izquierda entonces **no hay** un cuadrado verde sobre la derecha.
- 4) Si **no hay** un triángulo rojo sobre la izquierda entonces **no hay** un cuadrado verde sobre la derecha.

A cada sujeto se le presentaron los cuatro tipos de reglas. Para cada regla, el sujeto realizó tanto una tarea de verificación como una tarea de falsificación. La primera consistió en construir todas aquellas posibles combinaciones que hacían la regla verdadera. La segunda consistió en construir todas aquellas posibles combinaciones que hacían la regla falsa. Cabe aclarar que cada carta representaba una sola figura geométrica, por lo tanto, para construir una combinación era necesario seleccionar un par de cartas. Así, si por ejemplo la regla era: "Si **hay** un triángulo rojo sobre la izquierda entonces **no hay** un cuadrado verde sobre la derecha" (regla 3), el sujeto podía seleccionar una carta con un triángulo rojo, ponerla a la izquierda, y luego colocar un círculo azul a la derecha a fin de construir una combinación que hiciera la regla verdadera.

Evans observó que la ejecución de los sujetos era afectada por la presencia y ausencia de componentes negativos en las reglas. Cuando, por ejemplo, en la tarea de falsificación, se presentaba a los sujetos una regla con consecuente negativo (regla 3), la mayoría de los sujetos seleccionó correctamente una carta con un triángulo rojo a la izquierda y una carta con un cuadrado verde a la derecha, esto es, seleccionó una combinación V F.

Sin embargo, cuando el componente negativo era el antecedente (regla 2) la ejecución de los sujetos fue muy diferente. Sólo 7 de 24 sujetos escogió una combinación V F cuando se les pidió falsar la regla, mientras que 15 sujetos escogieron una combinación F V, al colocar

un triángulo rojo a la izquierda y un cuadrado verde a la derecha.

Evans atribuyó este resultado a una preferencia de los sujetos para seleccionar las cartas de hecho nombradas en las reglas y a una dificultad para apreciar la relevancia de casos que no igualan ninguna de las instancias nombradas. Por esta razón denominó a este error peculiar “matching bias” que aquí traduciremos como “sesgo de apareamiento”.

Evans además observó que este sesgo afectaba la ejecución de los sujetos solamente en el caso de la tarea de falsación. En la tarea de verificación los sujetos mostraron preferencia por las instancias V V para los cuatro tipos de reglas, incluyendo aquella de la forma “Si  $\neg$ P entonces  $\neg$ Q” la cual requiere elegir cartas no mencionadas en la regla.

Pero, ¿cómo se explica que este *sesgo de apareamiento* afecte sólo a la tarea de falsación? Una posible explicación, para Evans, es apelar a otro tipo de sesgo en el razonamiento que podría estar afectando la ejecución de los sujetos: “confirmation bias” (*sesgo confirmatorio*). Muchos psicólogos han sostenido que los seres humanos tienen una tendencia a buscar información que es consistente con sus creencias, teorías o hipótesis, y a no considerar información inconsistente con estas creencias teorías o hipótesis. Este fenómeno ha sido observado en numerosos experimentos de laboratorio en los que los sujetos fallan en descubrir una regla general precisamente porque sus estrategias están dirigidas a buscar evidencia que confirme -más bien que evidencia que refute- sus hipótesis.

Evidencia que apoya la hipótesis de la existencia de un *sesgo confirmatorio* en los humanos ha sido provista por una interesante tarea de razonamiento originalmente ideada por Wason (1960): “el problema 2,4,6”. La tarea consistía en lo siguiente: se le decía al sujeto que la secuencia de tres números enteros 2, 4, 6 se conformaba a una regla que el experimentador tenía en mente. El sujeto tenía que descubrir cual era esa regla generando sus propias secuencias de tres (triples), para cada una de las cuales el experimentador decía al sujeto si éstas se conformaban a la regla o no. Los sujetos tenían un registro de los triples propuestos, de las respuestas del experimentador y de sus hipótesis tentativas. Además, se pidió a los sujetos no enunciar su hipótesis acerca de cual era la regla que el experimentador tenía en mente, sino hasta que ellos se sintieran seguros de que era correcta. En caso de que ellos enunciaran la regla incorrecta se les informaba que tal era el caso y se les pedía seguir probando.

Esta tarea suele sugerir a los sujetos la regla: “ascendente con iguales intervalos”, sin embargo, la regla del experimentador es: “cualquier secuencia ascendente”.

La estrategia de los sujetos para probar su hipótesis consiste generalmente en buscar evidencia positiva, por ejemplo: “10 12 14” , “10 20 30 “ , “100 500 900 “. Cuando ellos preguntan al experimentador, éste responde en cada caso que los triples se conforman a la regla, lo cual hace sentir a los sujetos muy seguros de que su hipótesis es correcta y

sumamente sorprendidos cuando se les informa que ésa no es la regla correcta. La dificultad para encontrar la regla precisa parece aquí ser muy clara: los sujetos fallan en probar con casos negativos que no se conforman a su hipótesis, casos tales como: “10, 15, 17 “. La prueba de tales casos negativos conduciría a la falsificación de la hipótesis del sujeto y aproximaría a éste al descubrimiento de la regla más general que el experimentador tiene en mente.

Una visión generalizada es que este *sesgo confirmatorio* refleja una especie de tendencia motivacional; ya sea por vanidad o mantenimiento de la propia estructura de creencias, los seres humanos intentan activamente verificar sus hipótesis. No obstante, Jonathan Evans (1989, Cap. 3) rechaza esta visión y propone que, más que una tendencia motivacional, es una falla cognoscitiva peculiar la que subyace al fenómeno de este *sesgo confirmatorio*. Esta falla consiste en la tendencia de los sujetos a pensar en términos de información positiva más que en términos de información negativa. Este sesgo o tendencia humana es denominada por Evans: *positivity bias* (*sesgo de positividad*). Él considera que la existencia de tal sesgo es sustentada por numerosos experimentos mostrando las grandes dificultades que enfrentan los sujetos para comprender y procesar negaciones lingüísticas y lógicas.

Como ejemplo de estas dificultades baste mencionar los resultados de un estudio en particular, dirigido a examinar los procesos mentales subyacentes a una tarea de verificación. Clark y Chase (1972), presentaban a los sujetos oraciones como las siguientes: “El *más* está arriba de la *estrella*”, “El *más* no está arriba de la *estrella*”; simultáneamente con un despliegue visual que podía ser  $\begin{pmatrix} + \\ * \end{pmatrix}$  o  $\begin{pmatrix} * \\ + \end{pmatrix}$ . La tarea del sujeto era decidir tan rápido como fuera posible si la oración era verdadera o falsa en relación al despliegue visual. Entre otras cosas, los resultados de este experimento muestran que el tiempo de reacción (tiempo que tarda el sujeto en generar la respuesta) aumenta cuando la respuesta emitida es “falso” en contraposición a “verdadero”. De igual manera, el tiempo de reacción aumenta cuando la oración presentada al sujeto es negativa (“el más no esta arriba de la estrella”), más bien que afirmativa. Este resultado nos da elementos para asumir que el procesamiento de información negativa es algo más complejo que el procesamiento de información positiva, toda vez que requiere de más tiempo de procesamiento.

Por otro lado, hay otro estudio cuyos resultados apoyan la tesis de Evans en relación a la existencia de un sesgo de positividad. El experimento, realizado por Tweney, et al. (1980), intentaba una manipulación que ayudara a los sujetos a eliminar esa “tendencia a la confirmación” que se había observado en la solución del problema 2 4 6. A los sujetos se les dijo que el experimentador tenía dos reglas en mente, una de estas reglas generaba un tipo de triples llamados DAX, mientras que la otra generaba otro tipo de triples llamados MED. Se



les dijo también que la secuencia 2 4 6 era un ejemplo de un DAX. Los sujetos debían averiguar cuál era la regla DAX. Al igual que en la versión original del problema, los sujetos proponían sus secuencias de triples al experimentador, sólo que aquí, en vez de recibir como respuesta: “correcto” o “incorrecto”; ellos recibían como respuesta DAX o MED. El experimentador consideraba como DAX cualquier secuencia de triples que se ajustara a la regla: “cualquier secuencia ascendente” y como MED cualquiera otra secuencia que no fuera un DAX. La mejoría en la ejecución de los sujetos para este problema fue significativa. Ahora bien, este resultado no parece ser explicable en términos de un cambio de estrategia de los sujetos en el sentido de intentar desconfirmar más bien que confirmar sus hipótesis. La explicación, más bien parece ser que los sujetos intentan no solamente probar sus hipótesis acerca de la regla DAX sino que alternativamente también prueban sus hipótesis para la regla MED. En tanto que todo lo que no es un DAX es un MED y a la inversa, una prueba positiva de MED es también una prueba negativa de DAX, y esto último es lo que básicamente se requiere para eliminar la hipótesis falsa “ascendente con iguales intervalos”. De alguna manera, lo que parece suceder aquí es que los sujetos, al dar una etiqueta positiva (MED) a una hipótesis negativa (que en este caso sería: “no ascendente con iguales intervalos”), son capaces de conceder a ésta el mismo status que a cualquiera otra hipótesis positiva. Este resultado da apoyo a la tesis de Evans de que, más bien que los sujetos estén intentando confirmar sus hipótesis, el problema subyacente es su incapacidad de pensar en términos negativos.

En resumen, lo que Evans sostiene es que, tanto la tendencia de los sujetos a considerar relevante sólo la información mencionada en las reglas (*matching bias*), cómo su tendencia a confirmar/verificar sus hipótesis más que a refutarlas, pueden ser explicadas igualmente, por una falla subyacente al funcionamiento cognoscitivo que hace del procesamiento de información negativa una tarea algo más complicada para los seres humanos que el procesamiento de información positiva.

Debe notarse que Evans habla ya de una *falla subyacente* al funcionamiento cognoscitivo humano. Los experimentos han mostrado el carácter sistemático de esta *falla* y en tanto que ésta no ha sido explicada en términos de algún factor psicológico general y plausible que sea externo a la competencia de razonamiento, parece haber fuertes razones para suponer que ésta debe ser atribuida a la competencia de razonamiento humana.

#### E. Tabla de verdad defectiva.

Wason (1966, 1968) atribuyó las fallas en la tarea de selección a dos posibles causas. Se ha hablado ya de una de ellas: “los individuos tienden a buscar y a esperar una relación de correspondencia entre oraciones y estados de cosas”. Otra causa que Wason consideró es

que posiblemente la ejecución de los sujetos no estuviera constreñida por las reglas del cálculo proposicional, pues los sujetos implícitamente asumen que una oración condicional puede tener tres valores de verdad: verdadero, falso e irrelevante. De tal manera, el sentido de la regla condicional, según la interpretación de los sujetos, estaría dado por una tabla de verdad alternativa a la tabla de verdad tradicional. Dicha tabla sería la siguiente:

P	Q	$P \rightarrow Q$
V	V	V
V	F	F
F	V	I
F	F	I

Tabla de verdad defectiva propuesta por Wason (1968)

Fue así que Wason creyó posible que una *tabla de verdad defectiva* diera cuenta de la ejecución de los sujetos en la tarea de selección. Si esto era lo que estaba ocurriendo, entonces consecuentemente debía aceptarse que la *lógica mental* subyacente al razonamiento humano no sería aquella postulada por la lógica proposicional clásica. Por lo tanto, el razonamiento humano sería divergente de los principios normativos de razonamiento y tal desviación sería en principio atribuible a la competencia de razonamiento humana, y no a factores de ejecución. Pero veamos qué luz pueden arrojar los experimentos en relación a esta tesis de Wason.

Un experimento que arrojó resultados positivos a favor de la hipótesis de Wason fue realizado por Johnson-Laird y Tagart (1969). Ellos presentaron a los sujetos el siguiente enunciado: "Si hay una letra A en el lado izquierdo de la tarjeta entonces hay un número 3 en el lado derecho".

Después se dio a los sujetos un paquete de tarjetas con varias combinaciones de símbolos y se les pidió clasificar cada una de ellas en las tres categorías siguientes:

- 1) Tarjetas descritas de forma verdadera por el enunciado.
- 2) Tarjetas descritas de forma falsa por el enunciado.
- 3) Tarjetas para las cuales el enunciado es irrelevante.

19 de los 24 sujetos examinados clasificaron las tarjetas de acuerdo con la tabla de verdad defectiva propuesta por Wason.

Sin embargo, el mismo Wason (1968) ha aportado datos experimentales que ponen en entredicho su hipótesis. Él realizó un diseño experimental dirigido a probar si una cierta

terapia llamada “proyección de falsedad” (de la cual ya se ha hablado anteriormente) podía producir una facilitación en la tarea de selección.

El experimento fue realizado con 34 sujetos, 18 asignados al grupo experimental y 16 asignados al grupo control. El procedimiento se describe en el siguiente cuadro:

1era. Etapa	2ª. Etapa	3era. Etapa	4ª. Etapa	
Tarea de selección (determinar si la regla es verdadera o falsa)	Proyección de falsedad: decir qué símbolo del otro lado de cada carta: $p$ , $\neg p$ , $q$ , $\neg q$ , haría la oración condicional falsa	Revisión de la tarea de selección inicial: los sujetos podían corregir las respuestas previamente dadas en la tarea de selección si así lo deseaban.	Tarea de evaluación: cada carta es volteada y la regla es evaluada cómo verdadera o falsa en relación a ésta.	<b>Grupo Experimental</b>
Tarea de selección (determinar si la regla es verdadera o falsa)	Sólo se les pidió a los sujetos pensar en sus selecciones iniciales con el fin de revisarlas.	Revisión de la tarea de selección inicial: los sujetos podían corregir las respuestas previamente dadas en la tarea de selección si así lo deseaban.	Tarea de evaluación: cada carta es volteada y la regla es evaluada como verdadera o falsa en relación a ésta.	<b>Grupo Control.</b>

Wason, 1968. Experimento 1.

La oración condicional utilizada fue la siguiente: “Si hay una D de un lado de la carta entonces hay un 3 del otro lado”.

Cómo se mencionó anteriormente, esta manipulación no tuvo un efecto significativo en la comisión de los errores que caracterizan la ejecución en la tarea de selección, pero arrojó datos interesantes sobre la hipótesis de Wason que aquí nos ocupa.

Si se revisan los resultados conjuntos de los grupos experimental y control en la tarea de evaluación (etapa 4), parece que la hipótesis de Wason se confirma sólo para los casos de las cartas PQ,  $P\neg Q$  y  $\neg P\neg Q$ , pues para el caso de la carta  $\neg PQ$ , 22 sujetos consideraron esta instancia como falsando la regla y sólo 10 la consideraron irrelevante. Sin embargo, llama la atención que en la tarea de selección revisada (etapa 3) sólo 4 de 34 sujetos seleccionó la carta  $\neg P$ .

Esto muestra una clara inconsistencia en la ejecución de los sujetos. Por un lado, ellos consideran una instancia como falsando la regla y por otro lado fallan en seleccionar ésta como relevante para determinar la falsedad de la regla.

Una inconsistencia similar puede encontrarse en el caso de la carta  $P\neg Q$ .

En el grupo experimental, 13 de 18 sujetos fallaron en seleccionar  $\neg Q$  inicialmente (etapa 1). De esos 13, 8 proyectaron P sobre  $\neg Q$ , (es decir, consideraron que P del otro lado de la carta  $\neg Q$  haría la oración condicional falsa. (etapa 2)). De esos 8, sólo 3 incluyeron la carta  $\neg Q$  en su selección revisada (etapa 3). Y finalmente, considerados en conjunto, tanto el grupo experimental como el grupo control, se observó que 32 de 34 sujetos evaluaron (etapa 4) la instancia  $P \rightarrow Q$  como falsando la regla.

Aquí, al igual que en el caso anterior, los sujetos consideran una instancia  $P \rightarrow Q$  como falsando la regla y por el otro lado fallan en seleccionar la carta  $\neg Q$ .

Estas inconsistencias hacen evidente que, para ciertas maneras de plantear el problema lógico que subyace a la tarea de selección, la interpretación de los sujetos en relación a los valores de verdad de oraciones condicionales, es congruente con una tabla de verdad defectiva. Sin embargo, a la luz de otros planteamientos del problema los sujetos generan interpretaciones alternativas a la tabla de verdad defectiva propuesta por Wason.

Dadas estas inconsistencias, la tesis general de que una tabla de verdad defectiva subyace al razonamiento de los sujetos en relación a oraciones condicionales, no puede sustentarse.

#### F. Manejo de autocontradicción.

Acabamos de mencionar que algunas inconsistencias o contradicciones han sido observadas en la ejecución de los sujetos, de ahí que ciertas técnicas que han intentado producir una facilitación en la tarea de selección hayan recurrido a hacer evidentes para los sujetos sus propias contradicciones. Esto les puede producir una sensación de que algo anda mal y les impulsa a buscar una solución que, se espera, les permita entender y corregir su error. La estrategia consiste en exponer a los sujetos a inconsistencias entre sus evaluaciones independientes de cartas específicas (como falsando o verificando la regla) y su selección inicial.

Las inconsistencias o contradicciones son de dos clases:

Contradicción hipotética.- ocurre cuando un sujeto reconoce que un valor de P del otro lado de la carta mostrando un valor de  $\neg Q$  falsaría la regla, y aún así no selecciona  $\neg Q$ .

La inconsistencia aquí se ubica meramente en el pensamiento.

Contradicción concreta.- Ocurre cuando el sujeto voltea la carta P encontrando un valor Q del otro lado, verificando así la regla. Luego voltea la carta  $\neg Q$ , la cual consideraba irrelevante, encontrando un valor de P sobre el otro lado y se da cuenta de que esto falsa la regla. La inconsistencia aquí, provoca un choque obvio y explícito.

El recurso de la técnica de provocar un choque en el sujeto enfrentándolo a sus propias contradicciones fue utilizado por Wason (1969). Él expuso a sus sujetos a contradicciones seriadas a lo largo de la tarea empezando con contradicciones hipotéticas y finalizando con

contradicciones concretas. Este tratamiento hizo posible que 26 de 36 sujetos incorporaran  $\neg Q$  en sus elecciones al final del experimento.

Una mejora mucho menos significativa fue lograda por Wason y Johnson Laird (1970). Ellos expusieron a los sujetos directamente a la contradicción concreta. Sólo 9 de 34 sujetos cambió a la solución correcta aún a pesar de haber reconocido previamente que P y Q verificaban la regla y que P y  $\neg Q$  la falsaban. Los autores sugieren que la exposición directa a la contradicción concreta pudo haber sido traumática.

Aunque en ambos de los experimentos anteriores, la confrontación de los sujetos con sus propias inconsistencias permitió que un número considerable de sujetos incorporara  $\neg Q$  en su elección final, es remarcable el hecho de que otro grupo importante de sujetos no logró a final de cuentas hacerlo. En el caso de estos últimos sujetos, nos enfrentamos nuevamente a una estrategia que falla, en el intento de producir una mejoría en la ejecución de los sujetos en la tarea de selección.

Los resultados de los experimentos hasta aquí descritos, sugieren fuertemente que las fallas de los sujetos en tareas de razonamiento y particularmente en la tarea de selección, no son atribuibles a simples fallas de procedimiento, a deficiencias de atención, a limitaciones de memoria, o a otros posibles factores de ejecución. Los experimentos parecen demostrar la existencia de “fallas” cognoscitivas consistentes y más profundas. En términos de la distinción competencia/ejecución, podemos concluir que dichas “fallas” no pueden ser atribuibles a factores de ejecución, sino a la naturaleza misma de la competencia de razonamiento de los sujetos.

Una vez alcanzada esta conclusión será interesante analizar ahora una manipulación experimental que sí ha podido producir, en algunos casos, una facilitación significativa de la tarea de selección. En tanto que este resultado ha generado nuevas concepciones sobre la naturaleza de la competencia humana de razonamiento y en tanto que ha sido el punto de partida de toda una serie de investigaciones subsecuentes sobre la tarea de selección, se le dedica aquí una sección aparte. La manipulación a la que me refiero y sus resultados observados han venido a conocerse como *efecto de material temático*.

### **3. EFECTO DE MATERIAL TEMÁTICO.**

Como recién se ha dicho, hubo una manipulación experimental que pareció haber tenido éxito en la producción de una facilitación significativa de la tarea de selección. Esta manipulación consistió en utilizar contenidos concretos (temáticos o realistas) y reglas que involucraran una relación natural entre el antecedente y el consecuente. Es decir, reglas que

hicieran alusión a temas sobre los cuales la gente suele ocuparse en la vida cotidiana y cuya relación fuera además capaz de generar un sentido para las personas, en el marco de su discurso ordinario. Esto contrasta profundamente con los contenidos abstractos y relaciones arbitrarias generalmente utilizados en los experimentos relacionados con la tarea de selección, los cuales constituyen sin duda, tipos de problemas que la gente no suele enfrentar en su vida corriente. Sin embargo, este resultado exitoso ha también ido de la mano con el reporte de fallas en el intento de producir un efecto de facilitación temática. Estas inconsistencias muestran al *efecto de material temático* como un efecto inestable, y es en gran medida esta inestabilidad lo que ha inyectado el aliciente para tratar de dilucidar la naturaleza del efecto de material temático a fin de entender justamente bajo qué condiciones el uso de reglas con contenidos temáticos será capaz de producir una facilitación y porqué. A continuación revisaremos una serie de experimentos dirigidos a la investigación del efecto de material temático. Estos experimentos involucran básicamente dos tipos de tareas de razonamiento: la ya conocida tarea de selección y, además, una tarea de razonamiento también ideada por Wason, aunque menos divulgada e investigada, conocida como el *problema THOG*. Se revisarán los experimentos que involucran cada uno de estos tipos de tarea por separado.

### 3.1 LA TAREA DE SELECCIÓN.

Uno de los primeros estudios reportando un efecto de facilitación temática para la tarea de selección fue realizado por Wason y Shapiro (1971). Ellos evaluaron la ejecución de los sujetos en la tarea de selección utilizando la siguiente regla:

*“Cada vez que voy a Manchester viajo por tren”*

la cual, se dijo a los sujetos, hacía alusión a algunos viajes realizados por el experimentador. Las cartas utilizadas fueron: “Manchester” (P), “Leeds” ( $\neg$ P), “tren” (Q), “carro” ( $\neg$ Q). Se pidió a los sujetos que seleccionaran aquéllas cartas y sólo aquéllas que era necesario voltear para decidir si lo afirmado por el experimentador era verdadero o falso. En este caso 10 de 16 (62.5%) sujetos hicieron la selección correcta en oposición a 2 de 16 (12.5%) en el grupo control para el cual se uso material abstracto.

Facilitación significativa fue también observada en la tarea de selección temática ideada por Johnson-Laird, Legrenzi y Legrenzi, (1972). Aquí se pidió a los sujetos imaginar que eran trabajadores de una oficina postal. Ellos tenían que descubrir si la siguiente regla había sido violada:

*“Si una carta está sellada entonces ésta tiene una estampilla de 5d sobre ésta”.*

El material consistió en cuatro sobres: la parte de atrás de un sobre sellado (P), la parte de atrás de un sobre no sellado ( $\neg$ P), el frente de un sobre con una dirección y una estampilla

de 5d (Q), el frente de un sobre con una dirección y una estampilla de 4d (-Q). Todos los sujetos del experimento fueron probados tanto para la condición de material temático como para la condición de material abstracto, en un orden contrabalanceado. Esto quiere decir que todos los sujetos realizaron dos tareas de selección: la típica versión con contenido abstracto (letras y números) y una versión con el contenido temático arriba mencionado. En relación al orden, “contrabalanceado”, se refiere al hecho de que la mitad del grupo realizó primero la versión estándar de la tarea de selección y en seguida la versión temática, mientras que la otra mitad realizó primero la versión temática y en segundo orden la versión estándar de la tarea de selección. Esta previsión permite controlar cualquier efecto que el orden pudiera tener en la ejecución de los sujetos. De tal manera, las diferencias (en caso de observarse) en las ejecuciones de los sujetos para ambas versiones de la tarea, pueden confiablemente ser atribuidas a la manipulación temática, descartando posibles efectos del orden.

En este experimento, 21 de 24 (87.5%) sujetos hicieron la selección correcta en la condición de material temático en oposición a sólo 2 de 24 (8.3%) en la condición de material abstracto. Además de la mejora sustantiva en la ejecución, otro resultado notable en este experimento fue que no hubo transferencia alguna entre las condiciones temática y abstracta, es decir, la realización previa y correcta de la tarea de selección temática no produjo mejoría en la tarea de selección abstracta.

Adicionalmente, Bracewell y Hidi (1974), realizaron un experimento similar al de Wason y Shapiro. Usando pueblos y modos de transporte obtuvieron similares resultados. La única diferencia fue que Bracewell y Hidi examinaron las contribuciones de dos factores: relación natural vs. relación arbitraria (entre antecedente y consecuente) y contenido concreto vs. contenido abstracto.

En el caso del problema con contenido concreto y relación natural ellos observaron un 75% de selecciones correctas. En el caso del problema con contenido abstracto y relación arbitraria ellos observaron un 8.33% de selecciones correctas. No obstante, Griggs y Cox (1982) y Manktelow (1979) consideran que este efecto de facilitación pudo deberse, en parte, al hecho de que Bracewell y Hidi dijeron a sus sujetos que la regla condicional no era reversible (es decir, que la conversa del condicional no se sostenía), esto parece haber reducido el error de seleccionar la carta Q.

Otro estudio que sustenta la existencia de un efecto de material temático es el de Van Duyne (1974). Él utilizó material extraído de un supuesto registro de estudiantes. Este material relacionaba materias académicas con universidades.

Las reglas utilizadas fueron cuatro:

- 1) *Cada estudiante que estudia Física está en Oxford.*
- 2) *Si un estudiante estudia Filosofía entonces él está en Cambridge.*

- 3) *O bien un estudiante no estudia Francés, o está en Londres.*
- 4) *No es el caso que un estudiante estudie Psicología y no esté en Glasgow.*

Las cartas para cada una de las cuatro reglas fueron las siguientes:

- 1) *Física (P), Español ( $\neg P$ ), Oxford (Q), Cambridge ( $\neg Q$ ).*
- 2) *Filosofía (P), Física ( $\neg P$ ), Cambridge (Q), Oxford ( $\neg Q$ ).*
- 3) *Francés (P), Leyes ( $\neg P$ ), Londres (Q), Bristol ( $\neg Q$ ).*
- 4) *Psicología (P), Química ( $\neg P$ ), Glasgow (Q), Edinburgo ( $\neg Q$ ).*

Se pidió a los sujetos seleccionar sólo las cartas necesarias para saber si se viola la regla o no. Se observó una diferencia significativa en relación al material abstracto solamente para las dos primeras reglas. Para la regla universal (1) se registraron 50% selecciones correctas y para la regla condicional (2) se registraron 58% selecciones correctas.

Los experimentos hasta aquí descritos dan sustento a la idea de que el uso de contenidos concretos y de relaciones naturales entre el antecedente y el consecuente de la regla condicional, producen una facilitación significativa en la tarea de selección. Veamos ahora algunos de los experimentos para los que, el intento de producir una facilitación temática, ha sido desafortunado.

Manktelow y Evans (1979) hicieron una replicación casi exacta del experimento de Wason y Shapiro, pero no encontraron ninguna facilitación. Ellos probaron además con una regla diferente: *"If I eat haddock, then I drink gin"*. ("haddock" es un tipo de pez, "gin" es un tipo de bebida alcohólica). De igual modo, no encontraron efecto de facilitación alguno.

Posteriores fallas para replicar los resultados de Wason y Shapiro, utilizando el mismo contenido temático (pueblos y modos de transporte), fueron también reportadas por Brown, et. al. (1980) y Griggs & Cox (1982). Además, Yachanin y Tweney (1982), utilizando locaciones en un campo universitario y modos de transporte, obtuvieron también resultados negativos.

Por otro lado, Griggs y Cox (1982), replicaron el experimento de Johnson-Laird, Legrenzi y Legrenzi, sin observar ninguna facilitación. Mientras que Yachanin y Tweney (1982) fallaron en replicar los resultados de Van Duyne.

Los resultados arrojados por estos últimos experimentos muestran un efecto de material temático inconsistente. Pareciera que el uso de contenidos temáticos no es una condición suficiente para producir una facilitación. Surge entonces la pregunta de porqué algunos tipos de contenidos temáticos son capaces de producir una facilitación, mientras que otros no.



Una de las primeras respuestas a esta cuestión fue sugerida por Griggs y Cox (1982). Veamos cuál es su propuesta.

Si bien Griggs y Cox no observaron facilitación temática alguna en la replicación de los experimentos de Wason y Shapiro (1971) y Johnson-Laird et. al. (1972), ellos sí encontraron facilitación con otro tipo de material temático:

*“Si una persona está bebiendo cerveza, entonces la persona debe ser mayor de 19 años”* (19 años es la edad mínima legal para beber en Florida, dónde el experimento fue realizado). En este estudio se pidió a los sujetos imaginar que ellos eran oficiales de policía responsables de que la mencionada regla fuera cumplida. Las cartas fueron las siguientes: Bebiendo cerveza (P), Bebiendo Coca (¬P), 22 años de edad (Q), 16 años de edad (¬Q). Se informó a los sujetos que las cartas hacían alusión a cuatro personas sentadas a una mesa y se les pidió seleccionar aquéllas que era necesario voltear para saber si alguien estaba violando la regla. En este problema, *la selección correcta fue realizada por 74% de los sujetos.*

La explicación que Griggs y Cox dieron de estos resultados fue que el material temático es capaz de producir una facilitación, cuando el contexto en el que la tarea es presentada permite al sujeto *evocar experiencias pasadas que involucran el contenido del problema, la relación expresada y posibles contraejemplos a la regla.* Esta explicación ha sido nombrada por los autores como *“memory-cueing explanation”*.

Griggs y Cox consideran que *memory-cueing* permite dar cuenta de muchos de los resultados obtenidos en experimentos anteriores. Por ejemplo, la falla en replicar los resultados de Johnson-Laird et. al. con el problema postal pudo deberse a que los sujetos que participaron en el experimento de Griggs y Cox (sujetos norteamericanos) no estaban familiarizados con la relación entre sobres sellados y el costo del timbre postal; cosa que sí ocurría en el caso de los sujetos que participaron en el experimento de Johnson-Laird et. al. (sujetos británicos).

Esta hipótesis es apoyada por los resultados de Golding (1981). La regla postal utilizada por Johnson-Laird et. al. obedece a una regulación que estuvo vigente en Gran Bretaña hasta 1968. En conformidad con este dato y con *“memory-cueing explanation”*, Golding observó -utilizando una versión actualizada del problema postal y sujetos británicos- una diferencia significativa entre un grupo de sujetos menores de 45 años y un grupo de sujetos mayores de 45 años. El grupo de mayor edad logró 59% de selecciones correctas, mientras que el grupo de menor edad alcanzó sólo un 9% de selecciones correctas. Ella consideró que este resultado podía deberse al diferente grado de familiaridad que ambos grupos tenían con la vieja regla postal.

*“Memory-cueing explanation”* también puede dar cuenta de los resultados de Van Duyne y

de la falla en replicarlos de Yachanin y Tweney.

La facilitación observada por Van Duyne para las reglas:

1) *Cada estudiante que estudia Física está en Oxford.*

2) *Si un estudiante estudia Filosofía entonces él está en Cambridge*

puede explicarse del siguiente modo: las cartas que constituían la selección correcta para ambas reglas: "Física" y "Cambridge" para la regla 1, y "Filosofía" y "Oxford" para la regla 2, constituían además reconocidos contraejemplos a las reglas, pues Cambridge tiene reputación por su excelencia en Física mientras que Oxford tiene reputación por su excelencia en Filosofía. Este hecho pudo haber marcado la diferencia con el experimento de Yachanin y Tweney, pues las selecciones correctas en los problemas que ellos plantearon no constituían reconocidos contraejemplos a las reglas.

Otra versión temática de la tarea de selección que produjo facilitación significativa fue ideada por Roy D'Andrade (descrito en Griggs, 1983). Aquí se pedía a los sujetos imaginar que ellos eran gerentes en una tienda Sears y que estaban encargados de checar los recibos de venta para establecer si la regla:

*"Si una compra excede de 30 dólares entonces el recibo debe ser aprobado por el gerente de departamento",*

estaba siendo seguida. Se presentó a los sujetos cuatro recibos: uno para una compra de 45 dólares (P), otro para una compra de 15 ( $\neg$ P), uno firmado (Q) y uno no firmado ( $\neg$ Q).

Para esta tarea se observó un 70% de selecciones correctas.

Pero, ¿cómo aplicar *"memory-cueing explanation"* en un caso como este, si parece muy poco probable que los sujetos hayan tenido alguna experiencia previa como gerentes de Sears? Griggs (1983), sugiere que *"memory-cueing"* puede dar cuenta de estos resultados, sólo que debe agregarse otro elemento a la explicación: los sujetos no necesariamente deben haber tenido experiencia con el contenido y relaciones específicas de la tarea, aunque sí es necesario que los sujetos hayan tenido experiencia con contenidos y relaciones *"análogas"*. Así, *"memory-cueing"* más un *"razonamiento por analogía"* pueden dar cuenta de la facilitación lograda en versiones como la del gerente de Sears.

Sustento para esta última hipótesis es provisto por otro experimento realizado por Cox y Griggs (1982). Ellos utilizaron una regla análoga a aquella que involucra la edad para beber, pero en lugar de bebida, la regla hacía alusión al color de la ropa que la persona estaba usando. La regla era la siguiente:

*"Si una persona está usando azul, entonces la persona debe ser mayor de 19 años".*

Para esta tarea se pidió también a los sujetos imaginar que eran policías encargados de determinar si ciertas personas, sentadas a una mesa, estaban violando la regla. Las cuatro cartas fueron: usando azul (P), usando verde ( $\neg$ P), 22 años de edad (Q), 16 años de edad

(¬Q).

En este estudio, cuando la regla relacionando edad con el color de la ropa, fue la primera versión de la tarea de selección realizada por los sujetos, se observó un 25% de selecciones correctas. Cuando dicha regla se presentó como la segunda versión de la tarea de selección realizada por los sujetos -después de haberse enfrentado a una primera versión con la regla relacionando edad y tipo de bebida- se observó un 75% de selecciones correctas. En este experimento, la regla manipulada no era familiar a los sujetos; no obstante, es posible que ellos hayan podido vincular esta regla con múltiples y reconocidas restricciones sociales que tienen que ver con la edad de las personas, por ejemplo la edad adecuada para casarse, la edad para votar, la edad para tener responsabilidad penal, etc. Además, apelar a un razonamiento por analogía, parece plausible aquí, pues éste permite dar cuenta de la transferencia observada cuando la tarea fue precedida por la realización previa de la tarea con la regla relacionando edad y tipo de bebida.

Hasta aquí hemos visto cómo la utilización de ciertos contenidos temáticos es capaz de producir una enorme mejoría en la ejecución de los sujetos en la tarea de selección, así como una explicación tentativa de las razones de dicho resultado. Veamos ahora cuál es el panorama para el efecto de material temático en relación a un tipo de tarea diferente.

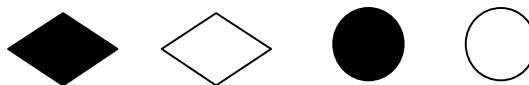
### 3.2 EL PROBLEMA THOG.

El problema THOG es un problema que involucra el manejo de la disyunción exclusiva y que, al igual que la tarea de selección, es especialmente difícil para las personas.

La versión estándar del problema es como sigue (Wason & Brooks, 1979):

*“Frente a ti están cuatro diseños: Diamante Negro, Diamante Blanco, Círculo Negro y Círculo Blanco. Tú debes asumir que yo [el experimentador] he escrito, en algún lado, uno de los colores (negro o blanco) y una de las formas (diamante o círculo). Ahora lee la siguiente regla cuidadosamente:*

*Si, y solamente si, alguno de los diseños incluye ya sea el color que yo he escrito o la forma que yo he escrito, pero no ambos, entonces éste es llamado un THOG. Yo te diré que el Diamante Negro es un THOG.*



*Cada uno de los diseños puede ahora ser clasificado dentro de una de las siguientes categorías:*

- a) Definitivamente es un THOG*
- b) Insuficiente información para decidir*
- c) Definitivamente no es un THOG “*

La respuesta correcta a este problema es que el diamante blanco y el círculo negro definitivamente no son THOGS mientras que el círculo blanco sí es un THOG.

La explicación es la siguiente: Sabemos que el diamante negro es un THOG, y sabemos que un THOG tiene la propiedad de poseer el color o la forma que el experimentador ha escrito, pero no ambos. En consecuencia, podemos deducir que: 1) el experimentador no ha escrito Diamante Negro y, 2) que el experimentador ha escrito una de dos posibles combinaciones: Diamante Blanco o Círculo Negro. Ahora bien, supongamos que **Diamante Blanco** es la combinación que el experimentador ha escrito, entonces: el Diamante Blanco no puede ser un THOG, pues éste posee tanto el color como la forma que el experimentador ha escrito. El Círculo Negro tampoco es un THOG, pues éste no posee ni el color ni la forma escrita por el experimentador. En cambio, el Círculo Blanco sí es un THOG, pues posee sólo una de las propiedades escritas (blanco), pero no ambas.

Supongamos ahora que **Círculo Negro** es la combinación que el experimentador ha escrito, entonces: Diamante Blanco no puede ser un THOG, pues éste no posee ninguna de las propiedades círculo o negro. Círculo Negro tampoco puede ser un THOG, pues éste posee ambas propiedades. Círculo Blanco, en cambio, sí es un THOG, pues éste posee una de las propiedades escritas (círculo), pero no ambas. Así, cualesquiera que sea la combinación que el experimentador haya escrito, se da el caso que el Diamante Blanco y el Círculo Negro no son THOGS, mientras que el Círculo Blanco sí lo es.

En la aplicación de este problema, Wason y Brooks encontraron que sólo aproximadamente 20% de los sujetos dio la solución correcta. La respuesta comúnmente dada fue que el Diamante Blanco y el Círculo Negro eran THOGS mientras que el Círculo Blanco no era un THOG, precisamente la respuesta invertida.

En el intento de entender las razones de la complejidad del problema THOG los psicólogos se han dado a la tarea de establecer y analizar cada uno de los procesos cognoscitivos involucrados en la solución correcta del problema. Se considera que estos procesos son los siguientes: 1) comprensión de la regla, 2) formación de hipótesis, 3) prueba de hipótesis, 4) análisis combinatorial.

En primer lugar, a fin de descartar la falta de competencia de los sujetos para comprender y usar la disyunción exclusiva, como una posible fuente de complejidad del problema, Wason y Brooks (1979), pidieron a sus sujetos escribir ellos mismos uno de los colores y una de las formas y luego entonces clasificar los diseños en relación a lo que ellos escribieron. Ellos encontraron que los sujetos no tenían ninguna dificultad para hacer esta tarea, lo cual indica que los sujetos comprenden y usan bien la disyunción exclusiva. Si es que no hay ningún problema en la comprensión de la regla, el siguiente paso ahora es indagar si los sujetos son capaces de generar las hipótesis correctas acerca de lo que el experimentador escribió. Los mismos Wason y Brooks (1979), en un segundo experimento, mostraron que la mayoría de sus sujetos eran capaces de generar las hipótesis correctas. Una vez que se tienen las hipótesis, el paso siguiente es probarlas sistemáticamente en relación a cada uno de los diseños. Como ya se ha mencionado, cuando los sujetos escribían ellos mismos un color y una forma, eran perfectamente capaces de probar sistemáticamente cada uno de los diseños en relación a la regla, sin embargo, puesto que en el problema THOG hay dos hipótesis en juego y no sólo una, es necesario que los sujetos realicen un análisis combinatorial adecuado que les permita checar cada hipótesis contra cada diseño y caer en la cuenta de que dos de los diseños no son THOGS, mientras que uno sí lo es, para cualquiera de las dos hipótesis consideradas.

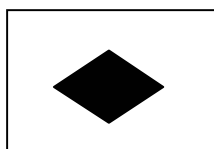
Realizar esta última parte de la tarea resulta algo más complicado para los sujetos, posiblemente porque el probar simultáneamente dos hipótesis es una tarea algo más compleja que probar una sola. También es factible suponer que la complejidad del problema THOG no se encuentra precisamente en alguno de los procesos cognoscitivos involucrados en la solución, sino más bien surge de la demanda de realizar todos estos procesos al mismo tiempo, lo cual puede producir una sobrecarga cognoscitiva. De ser así, podría argüirse que las personas son competentes en el razonamiento disyuntivo. El problema estaría simplemente en que la tarea demanda el procesamiento de una cantidad de información que desborda la capacidad de la memoria operativa humana. De tal manera las fallas de los sujetos serían atribuibles a factores de ejecución. No obstante, la explicación aquí presentada no es la única explicación que se ha dado de la complejidad del problema THOG y hasta ahora no se sabe bien a bien el porqué este problema es difícil para las personas.

En el intento de dilucidar el porqué de la complejidad del problema THOG, los investigadores del razonamiento han también explorado la posibilidad de facilitación mediante el uso de versiones temáticas o realistas. Los resultados de la manipulación temática en el caso del problema THOG han mostrado, al igual que con las manipulaciones temáticas de la tarea de selección, un efecto inconsistente. Así, mientras que algunas versiones temáticas no

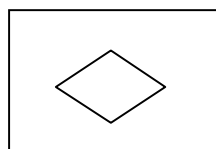
producen facilitación alguna, otras producen una facilitación significativa. Veamos una de las versiones temáticas que produjo resultados positivos. Newstead, Griggs, y Warner (1982), presentaron a los sujetos un problema gastronómico. El experimentador dice a los sujetos que él comería una comida si ésta tuviera sólo uno de dos alimentos que un amigo suyo había escrito, en x lugar [un sólido, ya sea carne o helado, y una salsa, ya sea *gravy* (tipo de salsa especial para la carne) o chocolate]. Luego se presentan cuatro comidas y se señala que el experimentador comería carne con *gravy*. Para este problema la solución correcta es que el experimentador también comería helado con chocolate, pero que no comería carne con chocolate o helado con *gravy*. Esta respuesta fue dada por la mayor parte de los sujetos, lo cual habla de una facilitación significativa de la tarea, sin embargo, esta facilitación ha sido atribuida a la concordancia de la solución correcta con las experiencias alimenticias previas de los sujetos. En consecuencia, es muy probable que los sujetos ni siquiera hayan tenido que considerar qué era lo que el amigo del experimentador había escrito y mucho menos realizar análisis combinatorial alguno para probar las hipótesis. Es por esto que los resultados observados para esta versión han sido entendidos más como una pseudofacilitación, que como una facilitación genuina de la tarea.

Por el contrario, en la siguiente versión ideada por Girotto y Legrenzi (1989) es posible hallar una facilitación genuina. Cabe mencionar que, aunque en este problema los contenidos de la regla siguen siendo abstractos, se introduce un contexto realista.

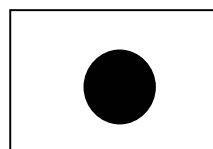
*Cinco amigos se encuentran todas las noches en el club. Una noche, Charles decide jugar un juego: “yo he traído un paquete de cartas, éste contiene solamente cuatro tipos de cartas [a los sujetos se les muestra en este momento los cuatro diseños usados en la versión original del problema THOG]. Yo tomaré del paquete una para mí y no se las mostraré a ustedes. Ahora yo repartiré a cada uno de ustedes una carta, y pagaré una comida a cada persona que tenga una carta incluyendo ya sea el color de mi carta, o la forma de mi carta, pero no ambos”. Las siguientes son las cartas de los amigos de Charles:*



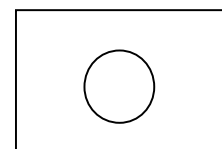
*Rob*



*Tim*



*Paul*



*John*

*“Sin mostrarles a ustedes mi carta, yo puedo decirles que yo debo una comida a Rob. ¿Qué carta piensan ustedes que yo podría tener? ¿Piensan que yo tengo que pagar comida a alguien más? Si es así ¿a quién?”*

Con esta versión, 89% de los sujetos obtuvieron la solución correcta.

Los resultados experimentales hasta aquí descritos registran tanto facilitación como no facilitación como consecuencia de la manipulación temática, en dos tipos de tareas de razonamiento: la tarea de selección y el problema THOG. Esto muestra, como ya se había anticipado, que el llamado efecto de material temático es un efecto inestable. Podemos concluir que el uso de material temático *per se* no es una condición suficiente para producir una facilitación en tareas de razonamiento y no sabemos si sea una condición necesaria. Es por esto que algunos autores han cuestionado la existencia de un efecto de facilitación debido al uso de material temático. Tal es la posición de Richard Griggs (1983). Él argumenta que cuando una facilitación temática puede ser observada, ésta es generalmente atribuible a la evocación de memorias específicas (*memory-cueing*), anexada, en algunos casos, a un razonamiento por analogía. Esto último implica que, en realidad, no hay ninguna facilitación del razonamiento lógico, sino más bien, 1) o una eliminación del mismo o 2) una sustitución de este por estrategias que permiten una reducción del esfuerzo cognoscitivo (*cognitive short circuiting*).

Analicemos en detalle estas dos posibilidades. La posibilidad uno destaca el hecho de que lo que puede estar ocurriendo cuando una facilitación se logra, es que la tarea de razonamiento es reemplazada, a final de cuentas, por una tarea de memoria. El punto sobre el que Griggs llama la atención aquí es importante, pues nos advierte de la existencia de casos donde efectivamente, la familiaridad de los sujetos con los problemas y situaciones planteadas, hace que las respuestas correctas sean fácilmente accesibles para los sujetos sin la necesidad de que ningún proceso de razonamiento esté teniendo lugar. Y por supuesto, si nuestro interés está en el proceso de razonamiento humano, es importante poder distinguir cuándo estamos tratando con una genuina tarea de razonamiento, y cuándo con una tarea de memoria. Un buen ejemplo de esta situación en que una pretendida tarea de razonamiento resulta más bien en una tarea de memoria, puede ser el caso del problema gastronómico ideado por Newstead et. al. (1982) arriba presentado. En un caso como este podemos estar perfectamente de acuerdo con Griggs en que no es posible hablar de un efecto de facilitación temática.

La posibilidad dos destaca otro hecho que puede estar ocurriendo cuando una facilitación se logra: puede ser el caso que la tarea temática planteada no se reduzca a una simple tarea de memoria, y que la solución de la misma sólo sea posible mediante la realización de algún tipo de proceso de razonamiento. Sin embargo, sostiene Griggs, cuando esto es lo que ocurre, el tipo de razonamiento involucrado no es en modo alguno el razonamiento formal requerido para la solución de la versión abstracta de la tarea de selección, lo cual implica

que, más que de una facilitación temática, hemos de hablar de un cambio en la naturaleza original de la tarea de selección o, en otras palabras, de un cambio de tarea. La evidencia en la que Griggs se apoya para afirmar lo anterior, es que no ha podido observarse ningún efecto de transferencia hacia los problemas de razonamiento abstractos una vez que se han solucionado correctamente versiones temáticas de esos mismos problemas; transferencia que según Griggs debería observarse si efectivamente se tratará de la misma tarea lógica. En relación a este segundo argumento esgrimido por Griggs para rechazar la existencia de un efecto de facilitación temática, es posible objetar lo siguiente: en realidad las versiones temáticas de la tarea de selección no cambian la naturaleza original de la tarea de selección abstracta, pues desde el punto de vista de su estructura lógica, las versiones temáticas son idénticas a la versión abstracta. Si, como señala Griggs, no es posible observar una transferencia a los problemas abstractos una vez solucionados correctamente problemas temáticos, este hecho no se explica por diferencias en la estructura lógica de los problemas temático y abstracto, sino más bien por diferencias que sólo pueden ser explicadas en términos de la psicología de los individuos a los que se han planteado dichos problemas. Por lo tanto, me parece que el argumento de Griggs no nos da realmente razones para rechazar la existencia de un efecto de material temático.

#### **4. CONCLUSIÓN.**

En el presente capítulo se ha hecho una rápida inspección de los experimentos más relevantes que han orientado las discusiones en psicología del razonamiento durante aproximadamente dos décadas, a partir de que Peter Wason realizara su experimento original con la tarea de selección.

Hemos visto que ciertas versiones abstractas de tareas de razonamiento son particularmente difíciles para los sujetos y aún para sujetos muy inteligentes. Hemos visto también que diferentes manipulaciones experimentales: algunas modificaciones de procedimiento, cambios en las instrucciones, familiarización previa de los sujetos con la regla, subsanar posibles limitaciones de memoria, etc., se han mostrado ineficaces para producir una facilitación de la tarea. Estas fallas sistemáticas sugieren fuertemente que las dificultades enfrentadas son atribuibles a factores que tienen que ver con la competencia humana de razonamiento, toda vez que no hay evidencia para dar una explicación plausible en términos de factores de ejecución. Por lo tanto, hay razones para pensar que los seres humanos razonamos de maneras que se desvían de los principios normativos de razonamiento que la visión estándar postula.



Por otra parte, se ha presentado una serie de estudios experimentales que nos muestran la existencia de una afectación en la ejecución de los sujetos como una consecuencia de los contenidos utilizados en las distintas tareas de razonamiento. Afectación que, según Griggs y Cox, puede ser explicada en términos de la evocación de ciertas memorias específicas (*memory-cueing*). En el siguiente capítulo veremos propuestas alternativas a la de Griggs y Cox y se analizará más en detalle el papel del contenido en el razonamiento humano.

## NOTAS.

---

<sup>4</sup> Debo aclarar que la siguiente presentación de experimentos en torno al razonamiento está organizada de acuerdo con las variables que se manipulan e investigan y no sigue en modo alguno un estricto orden cronológico.

<sup>5</sup> El término *memoria operativa* (también conocida como *memoria de corto plazo*) se refiere al dispositivo de memoria que sostiene la pequeña cantidad de información que está siendo atendida por el sujeto en un momento determinado y que puede rápidamente perderse si no es transferida a un dispositivo de memoria de mayor duración y capacidad al que los psicólogos cognoscitivos identifican como *memoria de largo plazo*.

<sup>6</sup> Aquí cabe señalar que aunque el experimento original de Wason pedía a los sujetos descubrir si el experimentador mentía al afirmar la oración condicional, gran parte de los experimentos posteriores con la tarea de selección pedían a los sujetos descubrir si la regla era verdadera o falsa.

## CAPITULO DOS

### TEORÍAS DEL RAZONAMIENTO.

Como se vio en el capítulo precedente, Griggs y Cox (1982) propusieron que el factor clave responsable del efecto de facilitación observada en algunos de los experimentos que manipulan contenidos temáticos, era la evocación de memorias específicas (*memory-cueing*). Sin embargo, posteriores investigaciones han dado lugar al desarrollo de nuevas teorías, las cuales han propuesto factores diferentes como los estrictamente relevantes.

En el presente capítulo se presenta una parte importante de dichas investigaciones, así como las propuestas teóricas generadas. Específicamente hablando, se presentan y contrastan aquí tres propuestas teóricas: 1) las *teorías de disponibilidad (TD)* dentro de las cuales queda incluida la explicación de evocación de memorias específicas de la cual ya tenemos antecedente; 2) la *teoría de esquemas de razonamiento pragmático (TERP)* de Patricia Cheng y Keith Holyoak, y 3) la *teoría del contrato social (TCS)* desarrollada fundamentalmente por Leda Cosmides.

Por otro lado, en la última parte de este capítulo abordaremos algunas cuestiones relativas a la investigación sobre el razonamiento no deductivo.

#### 1. TEORIAS DE DISPONIBILIDAD (TD)

Hay varias propuestas teóricas comprendidas dentro de lo que ha venido a conocerse como *teorías de disponibilidad*. Sin embargo, es común a todas ellas el explicar la facilitación temática observada en tareas de razonamiento en términos de la cantidad de experiencias que los sujetos han tenido previamente con los contenidos de las reglas utilizadas. Todas ellas sostienen, básicamente, que las experiencias recurrentes con ciertos tipos de contenidos crean *ligas asociacionales* entre ciertos tipos de elementos temáticos, mismas que tienden a hacerse mucho más fuertes en la medida en que la persona está mucho más expuesta a los contenidos relevantes. En el caso de la tarea de selección temática, las teorías de disponibilidad sostienen que cuando alguna facilitación se observa, ésta se debe invariablemente a la familiaridad de la regla, lo cual implica que los sujetos ya han experimentado previamente y con la frecuencia necesaria la co-ocurrencia de P&Q (la regla probada), así como la co-ocurrencia de P&¬Q (un contraejemplo de la regla) lo cual ha

creado una fuerte liga asociacional entre los elementos de la regla, de manera que, cuando se trata de falsar la regla,  $P \& \neg Q$  llegan a estar mucho más disponibles en la mente del sujeto.

No obstante, veremos más adelante que aunque un factor como el aquí propuesto permita dar cuenta del acierto o el error en ciertos tipos de problemas, la *familiaridad per se* no puede dar cuenta cabal de la ejecución mostrada por los sujetos en la tarea de selección.

## **2. TEORÍA DE ESQUEMAS DE RAZONAMIENTO PRAGMÁTICO (TERP)**

Por su parte, Patricia Cheng y Keith Holyoak (1985), proponen que la gente no razona utilizando representaciones de memorias específicas, como sostienen las teorías de disponibilidad, sino más bien que la gente razona a partir de *esquemas de razonamiento pragmático*. Un esquema de razonamiento pragmático consiste de un conjunto de reglas relativamente abstractas que, aunque son dependientes de contenido y por tanto propias de un dominio temático específico, tienen el suficiente grado de generalidad como para poder ser aplicadas a un rango más o menos amplio de situaciones de razonamiento. Dicho en otros términos, las reglas que constituyen un determinado esquema darán cuenta del razonamiento de las personas solamente cuando los contenidos del problema acerca del cual se razona se ubiquen dentro de un cierto ámbito temático capaz de elicitar ese esquema. Si las cuestiones acerca de las cuales versa el razonamiento pertenecen a un ámbito temático distinto, entonces seguramente un esquema de razonamiento alternativo deberá elicitar. En este sentido, un esquema de razonamiento es específico de un dominio, no obstante, su especificidad no llega al punto de requerir una regla específica para cada tipo particular de contenido. Es por esto que se dice que las reglas dentro de un esquema de razonamiento son relativamente abstractas y relativamente generales.

Por otra parte, un esquema de razonamiento tiene un carácter pragmático. Si bien es el caso que, desde un punto de vista lógico, varias inferencias pueden ser válidas, no es el caso que todas esas inferencias puedan ser útiles. Cheng y Holyoak ilustran este punto con un claro ejemplo: es simple ver que inferir que: *“si dos partículas no se repelen entre sí, entonces ellas no tienen la misma carga eléctrica”*, a partir de: *“si dos partículas tienen la misma carga eléctrica, entonces ellas se repelen entre sí”*, es obtener una conclusión útil. Por el contrario, inferir que: *“si no es el caso que yo deba tomar una aspirina, entonces yo no tengo dolor de cabeza”*, a partir de: *“si yo tengo dolor de cabeza, entonces yo debo tomar una aspirina”*, parece a toda vista inútil. Para la teoría de esquemas, las inferencias que importan son aquéllas que puedan ser útiles en la solución de problemas que de hecho nos demandan

una solución en la vida cotidiana. Por lo tanto, el proceso de razonamiento debe estar guiado por metas pragmáticas.

Por último, TERP sostiene que las reglas que conforman un esquema de razonamiento son *inducidas* a partir de la experiencia cotidiana con ciertas clases de problemas. Esto implica que los esquemas de razonamiento pragmático son adquiridos en el curso del desarrollo cognoscitivo tras la experiencia recurrente con cierto tipo de problemas. Puesto que este proceso de inducción -por el cual los esquemas de razonamiento pragmático son producidos- opera sobre cualesquiera tipos de problemas que la gente pueda encarar en el transcurso de la vida, dicho proceso es independiente de contenido. Debe, sin embargo quedar claro que las reglas comprendidas en el esquema son, en sí mismas, dependientes de contenido.

Esta es básicamente la teoría de esquemas. Veamos ahora qué puede ésta decirnos en relación a la tarea de selección y a la facilitación observada en el caso de algunas de sus versiones temáticas.

Cheng y Holyoak consideran que la facilitación temática observada en la tarea de selección puede ser explicada en términos de un esquema de razonamiento particular, a saber, en términos de un *esquema de permiso*. Este esquema consiste de cuatro reglas que nos permiten razonar acerca de regulaciones del tipo en el que realizar una determinada acción requiere de la satisfacción de una determinada precondition. Estas reglas son las siguientes:

- 1) *Si la acción va a ser realizada, entonces la precondition debe ser satisfecha.*
- 2) *Si la acción no va a ser realizada, entonces la precondition no necesita ser satisfecha.*
- 3) *Si la precondition es satisfecha, entonces la acción puede ser realizada.*
- 4) *Si la precondition no es satisfecha, entonces la acción no debe ser realizada.*

La forma en que el esquema de permiso puede producir una facilitación en la tarea de selección, según Cheng y Holyoak proponen, es la siguiente: en la tarea de selección temática un sujeto es enfrentado con una oración condicional que tiene la forma de la regla 1, esto permite evocar el esquema de permiso, lo cual significa que el conjunto total de reglas que componen a este esquema llegará a estar disponible al sujeto. Por un lado, la regla 3 permite bloquear la selección de la carta Q, pues esta regla nos dice que *si la precondition es satisfecha, entonces la acción puede, (aunque no necesariamente debe), ser realizada*. Así, el disponer de esta regla permite al sujeto darse cuenta de que la carta Q ("la precondition es satisfecha"), sea que tenga P ("la acción se realiza") o sea que tenga  $\neg P$

("la acción no se realiza") del otro lado, es perfectamente compatible con la regla acerca de la cual se razona (alguna instancia particular de la regla 1). Por otra parte, la regla 4 permite reducir significativamente el error de omitir  $\neg Q$ , pues, en tanto que asume la forma de la contrapositiva (es decir, si  $\neg Q$  entonces  $\neg P$ ) y está directamente disponible, exime al sujeto de la necesidad de inferir la contrapositiva de la regla probada, la cual es una inferencia necesaria a fin de percibir la relevancia de la carta  $\neg Q$ .

Es así como el *esquema de permiso* puede producir una facilitación temática.

Cheng y Holyoak (1985) realizaron tres experimentos encaminados a sustentar su teoría y a refutar la explicación, según la cual, es la evocación de memorias específicas, el factor responsable de la facilitación temática.

En lo que sigue, se hace una descripción rápida de sus dos primeros experimentos.

### Experimento 1

Cheng y Holyoak proponen que la deficiente ejecución en la tarea de selección es atribuible a la falla en evocar un esquema facilitador. Ellos consideran que en la medida en que dicha evocación pueda ser facilitada, la ejecución en la tarea de selección será sustancialmente mejorada. Consideran además que la evocación, en este caso, de un esquema de permiso, puede ser facilitada por dos factores:

- 1) la provisión de una razón o propósito social para la regla y,
- 2) la experiencia previa o familiaridad con la regla.

El objetivo pues de este experimento es probar si es el caso que los dos factores señalados pueden facilitar la evocación de un esquema de permiso.

Se utilizaron dos problemas temáticos para cada uno de los cuales se construyó una versión con propósito social y una versión sin propósito social. Estos problemas fueron los siguientes:

- Problema Postal.

Sin propósito social.- Se pide al sujeto imaginar que es un trabajador postal en un país extraño en el cual la regulación postal dispone que: *"Si una carta está sellada, entonces ésta debe llevar una estampilla de 20 centavos"*. El sujeto debe checar que la regulación sea seguida.

Con propósito social.- Es un problema igual al anterior, excepto que se agrega información señalando que, la razón de esta regulación es incrementar los beneficios del correo personal el cual es generalmente sellado. Si las cartas selladas se entienden como personales, éstas deben entonces pagar más.

- Problema del Cólera.

Sin propósito social.- Se pide al sujeto imaginar que es un oficial de migración en el aeropuerto de Manila. Su trabajo es checar, entre otros documentos, la forma H. Un lado de esta forma dice si el pasajero está entrando al país o está sólo en tránsito. El otro lado lista nombres de enfermedades tropicales. El sujeto debe asegurar que se cumpla la siguiente regla: *“Si la forma de un pasajero dice “entrar” de un lado, entonces el otro lado debe incluir “cólera”*”.

Con propósito social.- El problema es igual al anterior excepto que en lugar de decir que la forma H lista enfermedades tropicales, se dice que la forma lista inoculaciones que el pasajero ha recibido en los pasados seis meses y que con esto se pretende asegurar que los pasajeros que entran al país están protegidos contra estas enfermedades.

La otra variable manipulada fue la familiaridad de los sujetos con la regla. Para evaluar la contribución de este factor participaron sujetos en Michigan (88 estudiantes de la Universidad de Michigan), los cuales no tenían familiaridad con la regla postal, y sujetos en Hong Kong (82 estudiantes de la Universidad de Hong Kong), los cuales si tenían familiaridad con dicha regla, pues una regla similar estuvo vigente ahí hasta seis meses antes del experimento.

El procedimiento fue el siguiente: a cada sujeto, tanto en Michigan como en Hong Kong, se le dieron dos problemas: o bien la versión del problema postal con un propósito social y la versión del problema del cólera sin un propósito social o, la versión del problema postal sin un propósito social y la versión del problema del cólera con un propósito social. El problema postal siempre se presentó en primer lugar. Se dio a los sujetos el tiempo que fuera necesario y se les permitió hacer correcciones.

En el cuadro de la siguiente página se presenta esquemáticamente el diseño de este experimento.

De acuerdo con las predicciones de la teoría de esquemas, es de esperarse una ejecución mejor para el caso en que la regla tiene un propósito social que para el caso en el que no lo tiene. Así mismo se espera que, para el caso del problema postal sin un propósito social, la ejecución de los sujetos de Hong Kong, los cuales están más familiarizados con la regla postal, sea mejor que la de los sujetos de Michigan. Para el caso del problema postal con un propósito social, se espera una ejecución similarmente buena para los sujetos de ambas ciudades.

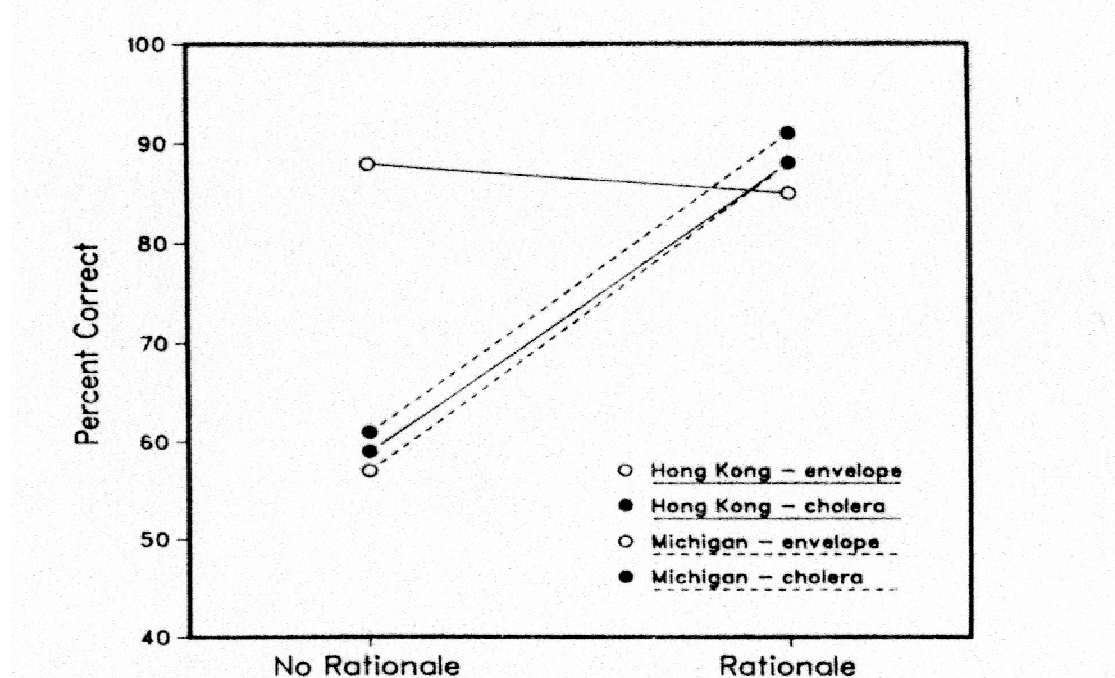
	Hong Kong University N = 82 Familiaridad de los sujetos con la regla postal.		Michigan University N = 88 No familiaridad de los sujetos con la regla postal	
	N = 41	N = 41	N = 44	N = 44
Problema Postal.	Con propósito social	Sin propósito social	Con propósito social	Sin propósito social
Problema del Cólera	Sin propósito social	Con propósito social	Sin propósito social	Con propósito social

Cheng & Holyoak 1985. Experimento 1

(N=número de sujetos)

En contraste, las teorías de disponibilidad predicen una ejecución similar para las dos versiones del problema del cólera, independientemente de si la regla tiene o no tiene un propósito social. Y para el caso del problema postal, la teoría de disponibilidad esperaría una ejecución mejor por parte de los sujetos de Hong Kong, sin importar si la regla tiene o no, un propósito social.

Los resultados de este experimento, mostrados en la gráfica, son congruentes con las



Resultados del experimento 1 presentados por Cheng & Holyoak, 1985, p. 401

predicciones de la teoría de esquemas, y no sustentan, en cambio, las predicciones de la teoría de disponibilidad.

### Experimento 2

Pudo haber sido el caso que, en el experimento 1, la provisión de un propósito social a la regla pudiera haber evocado ciertas memorias en los sujetos. De ser así, los resultados serían congruentes no sólo con la teoría de esquemas sino también con las teorías de disponibilidad, a las cuales se pretende excluir como una explicación factible de la facilitación observada en las reglas con contenidos temáticos. Por lo tanto, Cheng y Holyoak diseñaron un segundo experimento con el fin de excluir esta posibilidad.

A fin de establecer que el factor clave para la facilitación temática es la elicitación de un esquema de permiso y no la presencia de contenidos concretos (familiares), los autores contrastaron la ejecución de los sujetos en dos tipos de problemas: 1) un problema que describe una situación de permiso en términos abstractos y 2) un problema arbitrario (se refiere a la relación arbitraria entre el antecedente y el consecuente de la regla probada) en términos concretos. El problema de permiso/abstracto pedía a los sujetos imaginar que ellos eran una autoridad checando si la gente obedece o no ciertas regulaciones. Estas regulaciones tienen la forma general:

*“Si uno va a realizar la acción A, entonces uno debe primero satisfacer la precondition P”.*

Por otra parte, el problema arbitrario/concreto pedía a los sujetos establecer si la siguiente regla estaba siendo seguida:

*“Si una carta tiene una A en un lado, entonces ésta debe tener un 4 del otro lado”.*<sup>7</sup>

El procedimiento fue el siguiente: cada sujeto resolvió los dos problemas mencionados. El orden de estos fue contrabalanceado. Aquí, en contraste con el experimento anterior, no se permitió a los sujetos hacer correcciones sobre sus respuestas anteriores.

Lo que este experimento pretende mostrar es que un problema de permiso es capaz de producir una facilitación aún cuando éste no refiere a contenido concreto alguno que pueda evocar memorias específicas en los sujetos. Por otra parte se espera mostrar que para un problema con contenidos concretos, pero que no describe ninguna situación de permiso, no se producirá facilitación alguna. Esto dará soporte a la tesis de que es la evocación de un esquema de permiso, más bien que el carácter concreto y familiar del contenido, el factor relevante y responsable de la facilitación temática observada en la tarea de selección.

Aquí, nuevamente, Cheng y Holyoak obtienen resultados que se adhieren a las predicciones de la teoría de esquemas, pues 55% de los sujetos (N=44) solucionaron el problema de permiso/abstracto correctamente en contraste con un 30% que solucionó el problema arbitrario/concreto correctamente. El alto porcentaje de respuestas correctas en el problema



arbitrario/concreto en relación a la ejecución observada en experimentos anteriores para problemas de esta clase (problema *abstracto* en la literatura precedente), sugiere la ocurrencia de una transferencia positiva del problema de permiso/abstracto al problema arbitrario/concreto.

Estos son dos de los experimentos relevantes realizados por Cheng y Holyoak para sustentar su teoría y para refutar a las teorías de disponibilidad. Sin embargo, estos experimentos son susceptibles de varias críticas relativas a su diseño que ponen en entredicho su aceptabilidad como una pieza de evidencia a favor de la teoría de esquemas y que permita la exclusión de la “evocación de experiencias pasadas” como una posible explicación de sus resultados.

Algunas críticas importantes son señaladas por Leda Cosmides (1989): en el caso del primer experimento el problema es el siguiente: a fin de probar que dar un propósito social a la regla facilita la elicitación de un esquema de permiso, los sujetos son enfrentados con una regla a la que se da un propósito social y una regla a la que no. Pero para dar a la regla un propósito social debió agregarse a ésta mucha más información contextual que la dada a la regla que carecía de un propósito social. Se observó que para las reglas que tenían un propósito social (es decir, mayor información contextual) el porcentaje de respuestas P&-Q fue mayor que para las reglas que no lo tenían, lo cual es consistente con las predicciones de la teoría de esquemas. Sin embargo, este resultado también es consistente con las predicciones de las teorías de disponibilidad por dos razones: en primer lugar, las reglas usadas (aún las que no tenían un propósito social) no eran totalmente no familiares para los sujetos y, en segundo lugar, la información contextual agregada pudo haber provocado la evocación de memorias específicas, lo cual permitiría una explicación de la facilitación observada en las reglas con un propósito social, acorde con las teorías de disponibilidad.

En el segundo experimento Cheng y Holyoak buscaron evitar el error del experimento anterior. Tratando de eliminar información contextual que pudiera sugerir experiencias previas a los sujetos, ellos compararon una regla de permiso sin contenido concreto con una regla que relacionaba letras y números (tarea de selección abstracta estándar). Sin embargo, el problema de permiso no pudo dejar de referir a una autoridad checando regulaciones, lo cual es de cualquier modo, información contextual capaz de evocar experiencias específicas, mientras que la otra regla sí carecía efectivamente de contexto. La regla de permiso pudo así haber sugerido a los sujetos experiencias previas en las que ellos tuvieron que satisfacer una precondition a fin de obtener un permiso para hacer algo. En consecuencia, el resultado obtenido, a saber, una mejor ejecución para la regla de permiso, puede ser plenamente explicado por las teorías de disponibilidad.

Es así que los experimentos de Cheng y Holyoak, sostiene atinadamente Cosmides, no pueden ser usados para refutar a las teorías de disponibilidad.

Por mi parte, quisiera agregar dos observaciones críticas relativas al experimento 1: por un lado, si la teoría de esquemas asume que la familiaridad de los sujetos con la regla es capaz de facilitar la evocación de un esquema de permiso y consecuentemente se espera una buena ejecución por parte de los sujetos cuando la regla presentada es familiar, entonces sus predicciones en relación a los efectos de la manipulación de esta variable no difieren en absoluto de las predicciones de las teorías de disponibilidad. Asumiendo esto, la única variable verdaderamente relevante en la contrastación de ambas teorías es el que la regla tenga o carezca de un “propósito social”. Es por esto que no veo el sentido o pertinencia de la comparación de la ejecución de los sujetos para el caso de reglas familiares y no familiares (problema postal), esta manipulación en el primer experimento me parece que está de más. Por otro lado, considero que no es claro el porqué o el cómo la familiaridad de la regla o el tener un propósito social son factores capaces de producir la elicitación de un esquema de permiso. Creo que no es claro el tipo de vínculo entre estos factores y el esquema de permiso. Tal vez un pequeño párrafo por parte de los autores en la clarificación de este punto hubiera sido pertinente. Y lo considero pertinente porque si no queda clara la naturaleza de este vínculo, entonces no veo cómo puede afirmarse que la mejor ejecución de los sujetos para el caso de la regla con un propósito social (o para el caso de la regla familiar) constituye evidencia a favor de que el contar con un propósito social (o el ser familiar) es un factor que favorece la elicitación de un esquema de permiso, pues lo que nosotros podemos observar es simplemente la mejor ejecución por parte de los sujetos, en modo alguno hemos observado la elicitación de un esquema de permiso.

### **3. TEORÍA DEL CONTRATO SOCIAL (TCS)**

La teoría del contrato social ha sido originalmente desarrollada por Leda Cosmides y John Tooby. Esta teoría se caracteriza fundamentalmente por asumir un enfoque evolutivo de la psicología. En conformidad con dicho enfoque, la TCS sostiene que, tanto la estructura cerebral como los mecanismos cognoscitivos que subyacen a su funcionamiento son el resultado de la evolución por selección natural, es decir, son adaptaciones. Estas adaptaciones -en tanto que se han mostrado eficientes en la solución de problemas específicos, que como especie, los seres humanos hemos tenido que enfrentar- han sido modeladas y seleccionadas en el curso de nuestro desarrollo evolutivo. Es así que, desde esta perspectiva, la biología evolutiva viene a ser una herramienta esencial para la investigación en psicología cognoscitiva, pues ésta nos brinda información acerca de los

problemas adaptativos que la mente humana ha debido ser capaz de solucionar y da sugerencias sobre las características de diseño requeridas por cualquier mecanismo capaz de solucionar dichos problemas. Esto significa reducir significativamente el rango de mecanismos posibles que pueden dar cuenta de las capacidades cognoscitivas humanas, lo cual simplifica enormemente la tarea del psicólogo cognoscitivo, pues si él conoce de antemano los problemas adaptativos que los mecanismos cognoscitivos han debido solucionar, su tarea se reduce a investigar sólo aquellos mecanismos que están bien diseñados para la solución de dichos problemas. Esto le permitirá finalmente descubrir la naturaleza de los algoritmos que efectivamente comprende la mente humana.

TCS sostiene además, que los mecanismos cognoscitivos, en tanto que son adaptaciones eficientes, deben ser sistemas altamente especializados, es decir, eficientes en la solución de problemas en dominios específicos. Ahora bien, uno de estos dominios específicos, según TCS, es el *intercambio social*. El intercambio social, entendido como *la cooperación entre dos o más individuos para beneficio mutuo*, ha evolucionado en los seres humanos permitiéndoles incrementar sus posibilidades de supervivencia a través del intercambio de múltiples bienes. La biología evolutiva nos sugiere restricciones sobre cómo los humanos deben procesar información relativa a intercambios sociales. En otras palabras, nos brinda información sobre las características que debe tener un mecanismo cognoscitivo que funcione bien (adaptativamente) en la regulación de intercambios sociales.

Partiendo de esta información, Cosmides & Tooby (1989) y Cosmides (1989), proponen que cualquiera que sea el mecanismo cognoscitivo implicado en la conducta de intercambio social éste debe incluir las siguientes características de diseño:

- Debe contener algoritmos que operen sobre representaciones costo-beneficio. Esto es necesario para que el individuo pueda ser capaz de establecer los costos y beneficios que conlleva la realización de varias posibles acciones a fin de elegir aquel curso de acción que le resulte más benéfico.
- Debe incluir procedimientos de inferencia que permitan al individuo la detección de posibles tramposos en las relaciones de intercambio. Así podrá evitar intercambios sociales desventajosos en los que el costo pueda exceder al beneficio.

Evidencia para la existencia de un mecanismo cognoscitivo encargado de regular el intercambio social ha sido aportada por los experimentos de Cosmides (1989). Estos experimentos están basados en la ya conocida tarea de selección. Cosmides sostiene que los resultados de sus experimentos permiten sustentar la tesis de que existe un mecanismo cognoscitivo especializado para la solución de problemas de intercambio social y, además, que este mecanismo puede dar cuenta de la facilitación temática observada en la tarea de selección, pues un efecto de facilitación consistente puede ser observado cuando la regla

presentada a los sujetos puede ser interpretada como un intercambio social, es decir, una en la que se relacionan términos que pueden claramente ser identificados como costos y beneficios<sup>8</sup>.

Según TCS, dos tipos de reglas subyacen en nuestra capacidad para razonar acerca de intercambios sociales, cuando de la tarea de selección se trata:

Regla 1. Contrato Social Estándar.- “Si tú tomas el beneficio, entonces tú pagas el costo”.

Regla 2. Contrato Social Invertido.- “Si tú pagas el costo, entonces tú tomas el beneficio”.

De acuerdo con estas reglas, las cartas mostradas a los sujetos pertenecen a las categorías siguientes:

	Beneficio aceptado.	Beneficio no aceptado.	Costo pagado.	Costo no pagado.
Regla 1	P	¬P	Q	¬Q
Regla 2	Q	¬Q	P	¬P

Cosmides argumenta que, independientemente de la categoría lógica, un procedimiento de detección de tramposos debe llevar a las personas a escoger las cartas “costo no pagado” y “beneficio aceptado”, y a ignorar las cartas “costo pagado” y “beneficio no aceptado”, pues las primeras representan potenciales tramposos mientras que las segundas no representan gente que pueda haber hecho trampa o engaño alguno. En consecuencia, las respuestas correctas según la TCS son:  $P \& \neg Q$  para la regla 1 y  $\neg P \& Q$  para la regla 2. Podemos observar que las predicciones de TCS, para el caso de la regla 2, difieren de la que sería considerada una respuesta correcta desde el punto de vista de la lógica proposicional. Esta divergencia es interesante, pues la comparación de la ejecución de los sujetos para el caso de las reglas 1 y 2 permite establecer si el razonamiento está guiado por un mecanismo de detección de tramposos, como predice TCS, o simplemente por la estructura lógica de la tarea, como lo establece la visión estándar de la racionalidad.

Vayamos pues a los experimentos. Cosmides intenta construir pruebas críticas que permitan sustentar TCS al mismo tiempo que se refuta a las teorías alternativas: TD y TERP. Con este fin Cosmides (1989) diseña nueve experimentos: los primeros cuatro están dirigidos a contrastar la TCS con las teorías de disponibilidad, los restantes cinco están dirigidos a contrastar la TCS con la teoría de esquemas propuesta por Cheng y Holyoak.

Aquí, por razones de brevedad, se describirán sólo los experimentos más relevantes.

### 3.1 CONTRASTES: TEORIA DEL CONTRATO SOCIAL VS. TEORIAS DE LA DISPONIBILIDAD.

Para la TCS el mayor determinante de las respuestas en la tarea de selección, es el hecho de que el contenido de los problemas presentados pueda ser interpretado como un contrato social o no. Para las teorías de disponibilidad, el mayor determinante de las respuestas es el hecho de que la regla sea familiar o no. En consecuencia, TCS predice una buena ejecución en el caso de las reglas de contrato social, sean estas familiares o no. Las teorías de disponibilidad predicen, en cambio, una buena ejecución para las reglas familiares, independientemente de que éstas sean contratos sociales o no.

Para probar estas predicciones se realizó el siguiente experimento:

#### Experimento 1. (experimento 1 de Cosmides)

Participaron 24 sujetos. A cada sujeto se le pidió realizar cuatro tareas de selección que involucraban:

- 1) Un problema de contrato social estándar no familiar (CSE-noF).
- 2) Un problema descriptivo no familiar (D-noF).
- 3) Un problema descriptivo familiar (D-F).
- 4) Un problema abstracto (ABS).

(Ver versiones completas de estos problemas y de algunos de los planteados en los siguientes experimentos en el Apéndice A).

Para el caso de los dos problemas no familiares, las reglas utilizadas fueron:

*“Si un hombre come raíz de cassava entonces él debe tener un tatuaje en la cara”* y

*“Si tú comes carne de “duiker” entonces tú has encontrado una cáscara de huevo de avestruz”.*

Asimismo, para cada uno de los dos problemas no familiares se construyeron dos versiones: una versión “raíz de cassava” y una versión “carne de duiker”. A cada sujeto se le presentó, ya sea la versión “raíz de cassava” del problema CSE-noF y la versión “carne de *duiker*” del problema D-noF, o la versión “carne de *duiker*” del problema CSE-noF y la versión “raíz de cassava” del problema D-noF. Este manejo de dos versiones temáticas tuvo la finalidad de evitar algún tipo de transferencia entre los dos problemas no-familiares.

Para el caso del problema descriptivo familiar la regla utilizada fue:

*“Si una persona va a Boston, entonces toma el metro”.*

Y para el problema abstracto la regla utilizada fue:

*“Si una persona es clasificada D, entonces sus documentos deben ser marcados código 3”.*

El orden de los cuatro problemas fue contrabalanceado. Los sujetos debían resolver los cuatro problemas en el orden dado. No se les permitió volver a problemas anteriores y no hubo límite de tiempo.

Los resultados de este experimento parecen confirmar las predicciones de TCS, pues el porcentaje de respuestas correctas P&-Q más alto observado fue para el problema CSE-noF con un 75%, en contraste con el 46% y el 21% para el problema D-F y para el problema D-noF, respectivamente.

PROBLEMAS	RESPUESTAS CORRECTAS (P&-Q)
Contrato social estándar no familiar CSE-noF	75%
Descriptivo no familiar D-noF	21%
Descriptivo familiar D-F	46%
Abstracto ABS	25%

Cosmides, 1989. Resultados del experimento 1.

Se observa así que un problema de contrato social, aunque no familiar, es capaz de elicitar un porcentaje más alto de respuestas correctas que un problema familiar que no tiene la estructura de un contrato social. Este es un resultado que la teoría de disponibilidad no podría explicar. Los resultados podrían, sin embargo, ser explicados también por una hipótesis diferente a TCS. Podría argüirse que los problemas de contrato social pueden facilitar el razonamiento lógico. Esto querría decir que no habría necesidad de apelar a reglas de dominio específico relativas a contratos sociales para dar cuenta del buen razonamiento de los sujetos en problemas de contrato social, la explicación sería simplemente que, por alguna razón, los problemas de contrato social son capaces de elicitar en los sujetos el uso de los principios lógicos adecuados. Para descartar esta alternativa Cosmides realizó el siguiente experimento:

Experimento 2. (experimento 3 de Cosmides)

Participaron 24 sujetos. El procedimiento y los problemas utilizados fueron los mismos del experimento 1, sólo que, para los problemas no familiares las reglas fueron invertidas. Es decir, las reglas utilizadas tenían la estructura de la regla 2 mencionada arriba: “Si tú pagas el costo, entonces tú tomas el beneficio”.

Por lo tanto, las categorías de los problemas presentados fueron:

- 1) Contrato social invertido no-familiar (CSI-noF).
- 2) Descriptivo no-familiar (regla invertida), (D-noF).
- 3) Descriptivo familiar (D-F).
- 4) Abstracto (ABS).

Como ya se había mencionado, la respuesta correcta para un contrato social invertido es  $\neg P \& Q$ , en tanto que estas son las cartas que nos permitirían identificar un posible tramposo. Como vemos, dicha respuesta difiere de la respuesta lógicamente correcta  $P \& \neg Q$ . Por lo tanto, si la respuesta de los sujetos es acorde con lo predicho por TCS, esto sería indicativo de que su razonamiento está guiado por un algoritmo para la detección de tramposos, más bien que por algoritmos lógicos. Tesis, ésta última que parece ser confirmada por los experimentos de Cosmides. El porcentaje de respuestas  $P \& \neg Q$  fue muy bajo en comparación al porcentaje de respuestas  $\neg P \& Q$  para el problema CSI-noF: 4% vs. 67%.

Estos resultados permiten descartar la tesis de que el problema de contrato social pudiera producir una facilitación del razonamiento lógico.

También se observó un porcentaje más alto de respuestas de contrato social ( $\neg P \& Q$ ) al problema CSI-noF, que respuestas falsificantes ( $P \& \neg Q$ ) al problema D-F: 67% vs. 50%. Esto muestra que el tamaño del efecto de contrato social es mayor que el efecto de la familiaridad, dándose así soporte a la tesis sostenida por TCS de que la estructura de contrato social de una regla es el mayor determinante de las respuestas de los sujetos a problemas que involucran intercambio social aún cuando estos problemas no sean familiares.

PROBLEMAS	RESPUESTAS $P \& \neg Q$	RESPUESTAS $\neg P \& Q$
Contrato social invertido no familiar	4%	67%
Descriptivo no familiar (regla invertida)	12%	4%
Descriptivo familiar	50%	0%
Abstracto	12%	0%

Cosmides, 1989. Resultados del experimento 2.

Otros resultados interesantes de los experimentos 1 y 2 son los siguientes: El porcentaje de respuestas correctas para el problema D-F fue superior al porcentaje de respuestas correctas para el problema ABS. (Exp.1 46% vs. 25%; Exp.2: 50% vs. 12%).

Se observó también que la ejecución de los sujetos fue mucho mejor en el problema D-F en comparación con el problema D-noF. (Exp.1: 46% vs. 21%; Exp.2: 50% vs. 12%).

Cosmides reconoce que la disponibilidad pudo haber tenido un efecto sobre los problemas descriptivos familiares, sin embargo, ella considera que el efecto de contrato social es mucho mayor que el efecto de disponibilidad, ya que más respuestas de contrato social son elicitadas por contratos sociales no-familiares, que respuestas falsificantes  $P \& \neg Q$  por problemas descriptivos familiares. Por lo tanto, Cosmides concluye que algoritmos de contrato social son el mayor determinante de la ejecución de los sujetos en la tarea de selección.

### 3.2 CONTRASTES: TEORIA DEL CONTRATO SOCIAL VS. ESQUEMAS DE RAZONAMIENTO PRAGMÁTICO.

La TCS y la teoría de esquemas coinciden en un punto muy importante: ambas consideran que las reglas de inferencia que subyacen al razonamiento humano son específicas del dominio pertinente para el problema acerca del cual se razona. Por lo tanto, ambas asumen que las reglas de razonamiento pueden ser diferentes para diferentes dominios. Difieren sin embargo en lo siguiente: primero, en los algoritmos propuestos. Para TCS los algoritmos de contrato social representan al mundo en términos de costos y beneficios; mientras que para la teoría de esquemas, un esquema de permiso representa al mundo en términos de “acciones a ser realizadas” y “precondiciones a ser satisfechas”. Estas dos maneras de representarse un problema de razonamiento son, en esencia, representaciones acerca de intercambios sociales. La diferencia entre ambas reside en su grado de generalidad. La representación en términos de “acciones a ser realizadas” y “precondiciones a ser satisfechas” es mucho más general que la representación en términos de costos y beneficios. Esto es así porque todo beneficio aceptado puede ser visto como una acción realizada, y todo costo pagado puede ser visto como una precondición satisfecha. Aunque lo contrario no se sostiene, no toda acción realizada es necesariamente un beneficio aceptado, ni toda precondición satisfecha implica necesariamente un costo pagado. En consecuencia, toda regla de contrato social es también una regla de permiso, pero no toda regla de permiso es una de contrato social. Una regla de permiso es un contrato social sólo cuando los sujetos pueden interpretar “la acción a ser realizada” como un beneficio y “la precondición a ser satisfecha” como un costo.

Una segunda diferencia entre TCS y TERP se encuentra en el origen de los algoritmos propuestos. Hablaremos ampliamente de este punto más adelante. Permítaseme previamente presentar los experimentos de Cosmides para refutar TERP y sustentar TCS.

Conforme con esta primera divergencia justo señalada, la TCS considera que una regla de permiso puede producir un efecto de facilitación temática sólo si ésta es, además, un



contrato social. La teoría de esquemas, por el contrario, considera que una regla de permiso puede elicitar una facilitación en la tarea de selección aunque ésta no sea un contrato social. Para decidir qué clase de representación es la psicológicamente relevante, Cosmides realiza un experimento en el que se contrasta la ejecución de los sujetos en dos tipos de tareas: una de ellas implica un problema de permiso que carece de la estructura costo-beneficio del contrato social, la otra implica un problema de permiso que es, además, un contrato social. Si la TCS es correcta, entonces la regla de permiso que es también un contrato social elicitará muchas más respuestas correctas que la regla de permiso sin la estructura costo-beneficio del contrato social. Si la teoría de esquemas es correcta, entonces ambas reglas elicitarán un porcentaje igualmente alto de respuestas correctas.

### Experimento 3. (experimento 5 de Cosmides)

Participaron 40 sujetos. A cada sujeto se le presentaron dos tareas de selección: un problema de permiso con la estructura de un contrato social estándar (P-CSE) y un problema de permiso sin la estructura de un contrato social (P-noCS).

Las reglas usadas fueron las siguientes:

- 1) *Si un estudiante es asignado a la preparatoria Grover, entonces ese estudiante debe vivir en Grover City.*
- 2) *Si un estudiante es asignado a la preparatoria Milton, entonces ese estudiante debe vivir en el poblado de Milton.*

Para cada una de estas reglas había una versión P-CSE y una versión P-noCS. A cada sujeto se le presentó, ya sea la versión “preparatoria Grover” del problema P-CSE y la versión “preparatoria Milton” del problema P-noCS, o la versión “preparatoria Grover” del problema P-noCS y la versión “preparatoria Milton” del problema P-CSE. Al igual que en experimentos anteriores la razón para este manejo de dos versiones es evitar transferencias entre los dos problemas presentados a los sujetos.

Veáse la versión completa de estos problemas en el apéndice A.

Los resultados de este experimento se ajustan a lo esperado por TCS. El problema P-CSE elicató más respuestas P&¬Q que el problema P-noCS: 75% vs. 30%.

### Experimento 4. (experimento 8 de Cosmides)

En otro experimento, Cosmides exploró la ejecución de los sujetos (N=40) para el caso de reglas invertidas. Como ya se ha dicho, TCS espera un alto porcentaje de respuestas ¬P&Q para reglas que tienen la estructura de un contrato social invertido. Del mismo modo, TERP espera un alto porcentaje de respuestas ¬P&Q para reglas de permiso invertidas (i. e. *Si la precondition se satisface entonces la acción puede ser realizada*). Sin embargo, TCS

predice una alto índice de respuestas  $\neg P\&Q$  para la regla de permiso con la estructura de un contrato social invertido (P-CSI), no así para la regla de permiso que carece de la estructura de un contrato social (P-noCS); mientras que TERP espera un porcentaje igualmente alto de esta misma respuesta para ambos problemas (P-CSI y P-noCS).

El procedimiento y las reglas en este experimento fueron idénticos a los del experimento anterior, excepto porque las reglas fueron invertidas:

- 1) *Si un estudiante vive en Grover City, entonces ese estudiante será asignado (..is to be assigned to..) a la preparatoria Grover.*
- 2) *Si un estudiante vive en el pueblo de Milton, entonces ese estudiante será asignado a la preparatoria Milton.*

Los resultados de este experimento nuevamente se ajustan a las predicciones de TCS.

El problema P-CSI elicó un 65% de respuestas correctas  $\neg P\&Q$  mientras que el problema P-noCS no elicó ninguna respuesta  $\neg P\&Q$ .

#### **4. ALGUNAS CONSIDERACIONES ADICIONALES SOBRE TD, TERP y TCS.**

En la presente sección detallaré algunos puntos adicionales, tanto de acuerdo, como de contraste, entre las teorías del razonamiento arriba presentadas.

En primer lugar, tanto TERP como TCS se distinguen de las teorías de disponibilidad (TD) por una característica fundamental: mientras que TD caracterizan a la competencia de razonamiento como un mecanismo cognoscitivo independiente de contenido -pues sólo la familiaridad, sin importar el dominio de los contenidos, es condición necesaria para producir un efecto de facilitación temática- TERP y TCS caracterizan a la competencia de razonamiento como un mecanismo dependiente de contenido, es decir, asumen que las reglas o principios que son constitutivos de la competencia humana de razonamiento hacen alusión a los contenidos de dominios específicos de razonamiento.

No obstante este importante punto de acuerdo entre TERP y TCS, estas teorías difieren en un punto sustancial, a saber, en el origen propuesto de las reglas que constituyen la competencia de razonamiento humano. Para dejar bien claro este punto de controversia permítaseme aquí citar directamente a Cosmides:

*“Cheng et. al. creen que sus esquemas se originan en la inducción, que sus reglas de inferencia son el producto de “experiencia” estructurada únicamente por mecanismos de procesamiento de información innatos que son de dominio general. En contraste, la teoría de contrato social propone que las reglas de inferencia de los algoritmos de contrato social son*

*ellas mismas innatas, o si no el producto de "experiencia" estructurada por algoritmos innatos que son de dominio específico". (Cosmides, 1988, p. 235).*

Dicho en otras palabras, TERP propone que el ser humano dispone innatamente de un mecanismo de dominio general para procesar información de muy diversa índole, sin embargo tras la experiencia y el aprendizaje éste mecanismo general va desarrollando especializaciones, hasta que es posible hablar de diversos mecanismos cognoscitivos que dan cuenta del razonamiento humano en diferentes dominios específicos (entre ellos el dominio de aquellos problemas que involucran reglas de permiso). Por su parte TCS propone que el ser humano dispone innatamente de un mecanismo de dominio específico, producto de nuestra evolución biológica, que da cuenta del razonamiento humano en el caso de problemas que involucran intercambio social. En resumidas cuentas, mientras que TERP propone que los *esquemas de razonamiento pragmático* son inducidos desde la experiencia, esto es adquiridos; TCS propone que los *algoritmos de contrato social* son un producto de la evolución biológica y por ende innatos.

Este último es el punto en donde TERP y TCS genuinamente entran en conflicto, pues si consideramos el otro punto de divergencia entre estas dos teorías, a saber, la proposición de dos distintos tipos de representación de los problemas planteados: permiso vs. contrato social, posiblemente la teoría de esquemas no tuviera ninguna dificultad en aceptar, dados los resultados de Cosmides, que el nivel genuinamente relevante de representación del problema en las tareas de selección donde una facilitación ha sido lograda, es el de contrato social. Considero que TERP podría perfectamente incorporar dentro de su propuesta teórica la tesis de Cosmides acerca de la existencia de un mecanismo cognoscitivo de dominio específico que hace a los seres humanos competentes en la detección de posibles tramposos. En cambio, en lo referente al carácter innato o adquirido/aprendido de los algoritmos propuestos, TERP y TCS son irreconciliables.

Ahora bien, ¿cómo podríamos dirimir esta controversia? Hemos de reconocer que resolver esta cuestión es una tarea harto compleja, fundamentalmente porque no puede haber una separación tajante entre *regla innata* y *regla adquirida/aprendida*. La respuesta más ampliamente aceptada al cuestionamiento sobre el carácter innato o adquirido/aprendido de ciertas conductas o capacidades cognoscitivas en el individuo, ha sido que ellas son el resultado de la interacción de ambos factores: genética y ambiente, aunque se reconoce que el peso de cada factor puede variar. Si se acepta esto y se reconoce el hecho de que la expresión genética requiere de sutiles y complejas interacciones de los genes entre sí y con el ambiente, entonces se concederá que resulta más apropiado, como sostiene Kupfermann (1991), hablar de una gradación continua desde respuestas estereotipadas que son casi

independientes de la historia del individuo, a respuestas que son altamente sensitivas a factores ambientales. A la luz de esta concepción es posible aceptar que conductas y capacidades cognoscitivas altamente estereotipadas puedan llegar a ser modificadas por el ambiente, también es posible aceptar que conductas o capacidades cognoscitivas altamente plásticas estén influenciadas por factores genéticos.

Las consideraciones precedentes, así como las establecidas al inicio del capítulo anterior, sugieren que no es posible ya entender “innato” como “enteramente codificado en los genes” ni “adquirido/aprendido” como “enteramente libre de control genético”. Esta es una concepción demasiado simplista.

Teniendo esto en mente podemos volver a la controversia de TERP y TCS sobre el carácter innato o adquirido/aprendido de los mecanismos cognoscitivos de dominio específico postulados por cada una de estas teorías. Debe quedar claro que la discusión no puede situarse en el punto de establecer si la señalada competencia de razonamiento está enteramente genéticamente determinada o si ésta es enteramente adquirida. Si aceptamos que ambos factores, tanto genéticos como ambientales (aquí debe considerarse no solamente el ambiente físico sino también y fundamentalmente el ambiente social y cultural) deben estar interactuando en la producción de los mecanismos cognoscitivos involucrados en el razonamiento humano, entonces la discusión debe situarse en el punto de establecer el peso o la importancia relativa de cada factor en el resultado final. Desde este planteamiento, la diferencia entre TERP y TCS sería la siguiente: mientras que la primera atribuye más importancia o peso al factor ambiente la segunda atribuye mayor importancia a factores genéticos.

Este replanteamiento del problema, aunque clarifica un punto importante, no simplifica mucho la tarea de dirimir la controversia que aquí nos ocupa, pues dada la complejidad de las interacciones entre los dos factores (genético y ambiental) y el incipiente conocimiento que se tiene del papel que juega uno y otro en la conformación de la estructura cerebral, al igual que el incipiente entendimiento de la bases neurofisiológicas de procesos mentales superiores; resulta muy difícil poder determinar para un resultado específico, ya sea una conducta o una capacidad cognoscitiva, si ésta es “innata” o “adquirida/aprendida”.

Reconociendo la dificultad del problema, creo que debemos ser cautelosos con cualquiera postura que decidamos asumir.

Una vez conscientes de la dificultad de superar esta controversia entre TERP y TCS, es ahora pertinente conocer los argumentos esgrimidos por Leda Cosmides a favor de su posición. Ella argumenta, en contra de TERP en los siguientes términos:

*“La conexión hecha por Cheng y sus colegas entre sus estructuralmente bien especificados esquemas de permiso,.....y el hipotético proceso inductivo de dominio general que se supone los ha producido es, al momento, simplemente una aseveración sin contenido. Es posible probar la estructura de sus esquemas, pero ellos no han publicado teoría o datos demostrando que estos esquemas pudieran haber surgido a través de un proceso inductivo de propósito general.” (Cosmides 1989, p. 255).*

Además Cosmides argumenta que TERP debe aún especificar el conjunto de algoritmos a que refiere el término “inducción” pues en la actualidad dentro de la psicología dicho término refiere a un proceso cognoscitivo aún no especificado.

Sin duda Cosmides tiene razón en estos señalamientos. Sin embargo, la misma situación prevalece en el caso de TCS en el sentido de que ninguno de los experimentos realizados por Cosmides aportan elementos para sustentar la tesis de que los algoritmos de contrato social son el resultado de la evolución por selección natural y por tanto innatos.

Por otra parte hemos de recordar la evidencia señalada por Elman et. al. presentada en el capítulo precedente, a favor de la plasticidad cerebral y de la implausibilidad de que las funciones mentales superiores estén pre-especificadas desde las primeras etapas del desarrollo.

## **5. RAZONAMIENTO INDUCTIVO**

Toda la investigación y teorías previamente presentadas versan sobre lo que podemos llamar razonamiento deductivo. Esto es, aquél tipo de razonamiento que puede ser instanciado mediante un argumento de tipo deductivo, el cual se define por el hecho de que aquello que llamamos su conclusión se sigue *necesariamente* de aquello que llamamos sus premisas. Expresado en otros términos, un argumento deductivo es aquel en el que la verdad de las premisas garantiza la verdad de la conclusión de manera tal que, si sus premisas son verdaderas, la verdad de su conclusión se sigue *necesariamente*.

Para este tipo de razonamiento el estándar normativo ha sido invariablemente la lógica proposicional clásica, aunque como hemos visto, la investigación psicológica ha mostrado fallas sistemáticas de los seres humanos para ajustarse a dicho estándar. La pregunta pertinente que podemos hacernos ahora es ¿cuál es la competencia humana para resolver problemas que implican un razonamiento no deductivo? En lo que sigue echaremos un vistazo rápido a parte de la investigación que los psicólogos han realizado en torno al razonamiento no deductivo.

El razonamiento no deductivo, al que en términos generales podemos englobar bajo el rubro de *razonamiento inductivo*, se caracteriza porque, para el caso del argumento que nos permite instanciar dicho razonamiento –al cual llamamos argumento inductivo- la verdad de las premisas no garantiza la verdad de la conclusión, lo cual quiere decir que de la verdad de las premisas no se sigue *necesariamente* la verdad de la conclusión. Sólo podemos aseverar que de la verdad de las premisas se sigue la verdad de la conclusión con un cierto *grado de probabilidad*. De ahí que una inferencia inductiva involucre siempre un elemento de incertidumbre, pues la información dada en la conclusión va más allá de la información dada en las premisas. Los seres humanos razonamos inductivamente siempre que adoptamos creencias y/o elegimos cursos de acción sobre la base de información incompleta o de información relativa a la probabilidad de eventos inciertos. Por ejemplo, cuando nosotros inferimos que mañana será un día lluvioso y nublado sobre la base del estado del tiempo en los últimos días y consecuentemente decidimos cancelar el día de campo que habíamos planeado, nosotros estamos razonando inductivamente.

#### 5.1 LA FALACIA DE LA CONJUNCION.

Veremos enseguida que el panorama en la investigación sobre el razonamiento inductivo es similar al de la investigación en el ámbito del razonamiento deductivo en el sentido de que se han encontrado desviaciones sistemáticas de la ejecución de los sujetos de los principios normativos para este tipo de razonamiento, que en este caso han sido los principios establecidos por la teoría estándar de la probabilidad, así como por la teoría de la elección racional desarrollada originalmente en el ámbito de la investigación económica.

Una de las violaciones sistemáticas a un principio de la teoría de la probabilidad que ha sido ampliamente investigada es la violación de la *“regla de la conjunción”*. Esta regla deriva a su vez de la llamada *“regla de extensión”* la cual sostiene que:

Si la extensión de A incluye la extensión de B, entonces la probabilidad de A es mayor o igual que la probabilidad de B:  $(A \supset B)$  entonces  $P(A) \geq P(B)$

Puesto que el conjunto de eventos incluidos en la conjunción A&B están incluidos en el conjunto de eventos B, es decir  $B \supset A\&B$ , se sigue por la regla de extensión lo que se conoce como *“regla de la conjunción”*, a saber, que una conjunción no puede ser más probable que cualquiera de sus constituyentes:  $P(A\&B) \leq P(B)$  ;  $P(A\&B) \leq P(A)$

Una tarea experimental que ha llamado la atención por la violación sistemática de la regla de la conjunción por parte de las personas a las que se ha aplicado, fue diseñada por Amos Tversky y Daniel Kahneman (1983)<sup>9</sup>. La tarea consiste en lo siguiente: se presenta a los

sujetos un breve esquema sobre la personalidad de un individuo ficticio seguido de una lista de ocupaciones. Después se pide a los sujetos ordenar las ocupaciones en esta lista de acuerdo con su grado de probabilidad. La tarea es la siguiente:

*Linda tiene 31 años, es soltera, abierta y muy brillante. Ella se licenció en filosofía. Como estudiante ella estaba profundamente interesada en temas de discriminación y justicia social, y también participó en protestas antinucleares.*

*Linda es una maestra en escuela primaria.*

*Linda trabaja en una librería y toma clases de yoga.*

*Linda es activa en el movimiento feminista. (F)*

*Linda es una trabajadora social psiquiátrica.*

*Linda es miembro de la liga de mujeres votantes*

*Linda es una cajera de banco. (B)*

*Linda es una vendedora de seguros.*

*Linda es una cajera de banco y está activa en el movimiento feminista. (B&F)*

Lo que Tversky y Kahneman observaron es que la mayoría de sus sujetos (88%, N=173) ordenó la conjunción B&F como más probable que B, violando así la regla de la conjunción. Cabe destacar que esta tarea fue presentada a tres distintos grupos de respondientes: un grupo de estudiantes *estadísticamente ingenuos*, es decir sin conocimiento alguno de probabilidad o de estadística; un grupo de estudiantes *informados*, es decir con conocimientos básicos de probabilidad y estadística; y un grupo *sofisticado* el cual consistió de estudiantes de doctorado que habían tomado cursos avanzados de probabilidad y estadística. Sin embargo, no se observó ninguna diferencia significativa en las respuestas de los sujetos para estos tres grupos.

Tras este resultado, Tversky y Kahneman (1983) idearon una serie de manipulaciones experimentales intentando producir alguna mejoría en la ejecución de los sujetos, lo cual permitiría identificar si esta “falla” -la que ha venido a conocerse como “falacia de la conjunción”- podía ser el producto de un mero factor de ejecución o si era más bien atribuible a la competencia de razonamiento de los sujetos. Podemos adelantar que, al igual que en el estudio del razonamiento deductivo, el carácter sistemático de la violación de la regla de la conjunción sugiere que se trata más bien de una cuestión de competencia. A continuación se describen algunos de estos experimentos.

Una de las primeras manipulaciones experimentales intentaba facilitar la detección de la relación de inclusión del evento B&F en B. Con este fin Tversky y Kahneman presentaron a

un nuevo grupo de sujetos la descripción original de Linda seguida solamente de las dos oraciones relevantes:

*Linda es una cajera de banco (B)*

*Linda es una cajera de banco y está activa en el movimiento feminista. (B&F)*

Los sujetos debían establecer cual de las dos alternativas era la más probable. El resultado fue que 85% (N=142) de los respondientes considero a B&F como más probable que B.

Otra prueba experimental en la que se intentó clarificar la relación de inclusión de los dos eventos relevantes, consistió en presentar a los sujetos la misma descripción de Linda seguida de la oración B&F y la oración B\*, en lugar de B:

*Linda es una cajera de banco si o no ella está activa en el movimiento feminista. (B\*)*

En seguida los sujetos debían estimar la probabilidad de cada uno de los eventos utilizando una escala de 9 puntos: 1 para designar a un evento extremadamente improbable hasta el 9 para designar a un evento altamente probable. Aunque aquí la naturaleza inclusiva del evento B\* (el cual es extensionalmente equivalente a B) parece mucho más evidente, 57% (N=75) de los sujetos participantes violaron la regla de la conjunción al asignar una probabilidad más alta a B&F.

Con el fin de descubrir si las personas son capaces de reconocer la validez de la regla de la conjunción, Tversky y Kahneman realizaron otro experimento. Ellos presentaron a los sujetos la descripción de Linda seguida de las dos oraciones relevantes B&F y B, en seguida se les pidió señalar *cual de los dos argumentos siguientes les parecía más convincente*:

*Argumento 1. Es más probable que Linda sea una cajera de banco a que sea una cajera de banco feminista, porque cada cajera de banco feminista es una cajera de banco, pero algunas mujeres cajeras de banco no son feministas, y Linda podría ser alguna de ellas.*

*Argumento 2. Es más probable que Linda sea una cajera de banco feminista a que sea una cajera de banco, porque ella se asemeja a una feminista activa más de lo que ella se asemeja a una cajera de banco.*

Para este problema, 65% (N=58) de los sujetos escogió el argumento de semejanza inválido (argumento 2), como más convincente.



Hasta aquí, la comisión de la falacia de la conjunción parece persistente. No obstante los resultados de los experimentos son más complejos, pues aquí al igual que en la investigación del razonamiento deductivo, se han ideado tareas y se han hecho ciertas manipulaciones experimentales que se han mostrado capaces de producir una mejoría en la evitación de la falacia de la conjunción. Por ejemplo, Tversky y Kahneman reportan el logro de una mejoría –aunque pequeña- cuando ellos pidieron a los sujetos decidir a cual de los dos eventos apostarían si a B o a B&F: *Si tú pudieras ganar diez dólares por apostar a un evento, ¿a cuál de los siguientes le apostarías?* Se uso la misma descripción de Linda y las dos oraciones relevantes de los experimentos anteriores. El porcentaje de violaciones de la regla de la conjunción para esta tarea fue del 56% (N=60), en contraste con el 88% de violaciones en la versión original citada más arriba.

Los mismos Tversky y Kahneman también reportan que en una aplicación posterior del problema de Linda -se uso la descripción de Linda seguida de las dos oraciones relevantes y se pidió a los sujetos estimar la probabilidad de cada oración en una escala de 9 puntos: 1 (extremadamente improbable), 9 (extremadamente probable)- a un grupo de estudiantes graduados en ciencias sociales, los cuales habían acreditado varios cursos de estadística, sólo 36% (N=64) de los sujetos cometieron la falacia de la conjunción. Lo cual parece indicar que la *sofisticación estadística* si tuvo un efecto aquí.<sup>10</sup>

Otra manipulación que tuvo un efecto favorable fue presentar a los sujetos una versión diferente del problema de Linda en la que la presentación esquemática de su personalidad fue omitida, a los sujetos solamente se les dijo que Linda era una mujer de 31 años. Para esta tarea, Tversky y Kahneman reportan que casi todos los sujetos (ellos no establecen un resultado en términos de porcentaje) respetaron la regla de la conjunción al señalar la oración B&F como menos probable que sus constituyentes.

Tras los resultados obtenidos, Tversky y Kahneman (1983) concluyen que, mientras las leyes de la probabilidad derivan de consideraciones extensionales, los juicios intuitivos de probabilidad no. Esto quiere decir que mientras una medida de probabilidad está determinada por la *frecuencia relativa* así como por las relaciones de *inclusión de clase* de un determinado evento, nuestros juicios intuitivos de probabilidad no están así determinados por esos factores, pues cuando las personas hacen estimaciones de la probabilidad de eventos inciertos ellas no siempre hacen listas exhaustivas de eventos posibles para ponderar la posibilidad de ocurrencia de un cierto evento. En lugar de eso, Tversky y Kahneman consideran, la gente *comúnmente usa un número limitado de heurísticos*<sup>11</sup> *tales como representatividad y disponibilidad* (p. 294), los cuales no se conforman a la lógica extensional de la teoría de la probabilidad.

Por un lado, el **heurístico de disponibilidad** opera cuando la gente hace un juicio de probabilidad sobre la base de la información que le es inmediatamente disponible –como es el caso de la información susceptible de una más fácil recuperación desde la memoria-. Por ejemplo, Kahneman y Tversky (1973) mostraron que la gente considera como más frecuentes a las palabras que inician con la letra “k” (dentro del idioma inglés) en relación a las palabras que tienen una “k” en la tercera posición. Aunque esta respuesta no es correcta, Kahneman y Tversky consideran que la razón de esta respuesta no es que la gente conozca más palabras que inician con “k” que palabras con “k” en la tercera posición, más bien lo que ocurre es que la recuperación desde la memoria de la primera clase de palabras es mucho más fácil que la recuperación de palabras de la segunda clase.

Por otro lado, el **heurístico de representatividad** opera cuando el juicio de probabilidad está basado en *“una estimación del grado de correspondencia entre una muestra y una población, una instancia y una categoría, un acto y un actor o, más generalmente, entre un resultado y un modelo”* (Tversky & Kahneman, 1983; p.295). Así, por ejemplo, un individuo puede ser representativo de un determinado grupo social en tanto que sus características son similares a las características del miembro estereotípico de ese grupo. O una acción X puede ser representativa de un determinado individuo en tanto que ese individuo es reconocido como teniendo la tendencia a actuar de manera X.

Tversky y Kahneman señalan que la *representatividad* tiende a covariar con la frecuencia, de modo que los eventos más frecuentes tienden a ser más representativos que los eventos menos frecuentes. Sin embargo, hay casos donde esta covariación no se da, de modo que un evento puede ser representativo aunque infrecuente. Un caso como este queda ejemplificado con la siguiente evidencia empírica (Tversky & Kahneman, 1983): 65% (N=105) de los sujetos en un grupo experimental consideraron como más representativo de una actriz de Hollywood “ser divorciada más de 4 veces” que “votar por el partido demócrata”; mientras que en otro grupo 83% (N=102) de los sujetos consideró que entre las actrices de Hollywood hay más “mujeres que votan por el partido demócrata” que “mujeres divorciadas más de 4 veces”. Vemos pues que aunque entre las actrices de Hollywood la característica de “votar por el partido demócrata” es considerada como más frecuente, es al mismo tiempo considerada como una característica menos representativa de una actriz de Hollywood que la característica de ser divorciada más de 4 veces.

Este, entre otros resultados, es presentado por Tversky y Kahneman como evidencia de que la *representatividad* no es extensional.

El heurístico de representatividad es, según Tversky y Kahneman, la estrategia psicológica que opera cuando las personas son enfrentadas a problemas como el de Linda. En este problema uno puede darse cuenta de que la descripción de Linda es representativa de una

feminista activa y no representativa de una cajera de banco. Puesto que el agregar la cualidad de ser feminista a una cajera de banco incrementa el grado de semejanza con el modelo de Linda, los sujetos tienden a juzgar a B&F como más probable que su menos representativo constituyente B.

Es así que los sujetos, más que plantearse la tarea como un cálculo de probabilidades, dirigen su atención al grado de representatividad (correspondencia) de cada una de las ocupaciones respecto de un modelo, que en este caso es la personalidad de Linda. Ahora bien, en los casos donde una mejor ejecución ha podido observarse, algunos factores deben ser responsables de la facilitación del razonamiento extensional. Entre estos factores Tversky y Kahneman consideran:

- 1) la sofisticación estadística.
- 2) la anulación de la relación entre el modelo de Linda y el evento altamente representativo (F), la cual se da cuando la descripción de Linda es omitida
- 3) la presentación de la tarea en escenarios en los que se combina el azar y la apuesta.

Aunque estos factores puedan de alguna manera promover el razonamiento extensional, su contribución debe todavía ser investigada, pues como hemos visto en el caso de la sofisticación estadística, así como en el caso del uso de escenarios de apuesta (lo cual se verá más adelante), los resultados no son consistentes.

La evidencia a favor de la utilización por parte de los sujetos de un heurístico de representatividad no se limita a la prueba de la regla de la conjunción mediante tareas como la de Linda. Tversky y Kahneman investigaron los juicios intuitivos de probabilidad en otros ámbitos donde la incertidumbre de los eventos relevantes, más que ser atribuible a una información incompleta, era atribuible a un factor azaroso. Veamos pues una de estas tareas experimentales, la cual resulta además interesante por una hipótesis adicional que es sometida a prueba.

Tversky y Kahneman intentaron descartar la hipótesis de que los sujetos pudieran estar entendiendo el término “probabilidad” de una manera alternativa. Tal vez los sujetos tenían una concepción distinta de la probabilidad que, aunque no estándar, podría considerarse como legítima. La *representatividad* podría en este caso ser vista como una concepción legítima, aunque no estándar, de la probabilidad. De ser esto lo que estaba ocurriendo, la falacia de la conjunción podría ser atribuible a un mal entendido de la palabra *probabilidad*. Para probar esta hipótesis, los autores idearon un problema de decisión en el que había un incentivo para elegir el evento más probable, aunque la palabra *probabilidad* nunca se menciona. El problema fue el siguiente:

Considere un dado regular de 6 lados con 4 caras verdes y dos caras rojas. El dado será lanzado 20 veces y la secuencia de verdes y rojos será registrada. A ti se te pide seleccionar una secuencia, de un conjunto de tres, y tú ganarás 25 dólares si la secuencia que tú escoges aparece en los lanzamientos sucesivos del dado. Por favor checa la secuencia de verdes y rojos a la cual tú prefieres apostar.

1. RRRR
2. RRRR
3. RRRRR

Este problema fue presentado a 260 estudiantes. A 125 de los 260 sujetos se les ofrecieron pagos reales. Los sujetos restantes del grupo (135) jugaron la apuesta de un modo hipotético. De cualquier manera había un incentivo para elegir la secuencia más probable.

En este problema la secuencia 1 es más probable que la secuencia 2, puesto que como podrá notarse la secuencia 1 queda incluida dentro de la secuencia 2 de modo tal que si la secuencia 2 ocurre, forzosamente ocurre la secuencia 1; pero no al revés, la secuencia 1 puede ocurrir sin que ocurra la secuencia 2. No obstante, 65% de los sujetos en el grupo de pago real apostaron a la secuencia 2, mientras que así lo hicieron 62% de los sujetos que apostaron de modo hipotético.

Para Tversky y Kahneman la explicación de este resultado es nuevamente que las respuestas de los sujetos están guiadas por la operación del *heurístico de representatividad*. Aunque las tres secuencias presentadas son poco representativas del dado en cuestión puesto que exhiben pocas apariciones de verde que es el color más probable, la secuencia 2 parece mejor a los sujetos que la secuencia 1 desde el momento en que presenta una proporción más alta de verdes:  $1/3$ , en comparación con la proporción de verdes para la secuencia 1, la cual es de  $1/5$ , resultando esto en un mayor grado de representatividad de la secuencia 2.

Estos resultados permiten además descartar que la comisión de la falacia de la conjunción pueda ser explicada como un malentendido de la palabra *probabilidad*, pues aunque ésta no fue mencionada el planteamiento del problema alentaba la estimación de las probabilidades, éstas si en el sentido estándar. No obstante, los sujetos se guiaron por el heurístico de representatividad.

En este experimento podemos también observar que la presentación de la tarea en un escenario de apuesta tuvo poco efecto en la promoción de un razonamiento extensional, lo cual es inconsistente con la tesis de Tversky y Kahneman según la cual, la combinación de un escenario de *azar* y una tarea de apuesta promueven el razonamiento extensional *al enfatizar las condiciones bajo las cuales se pagará una apuesta*. (p. 304).

Tras este resultado son de llamar la atención los resultados del siguiente experimento.

A un grupo de 88 estudiantes se les presentó el mismo problema anterior, con la diferencia de que la secuencia 3 fue omitida y de que a ellos no se les pidió elegir a que secuencia le apostarían. En lugar de esto, ellos tenían que *indicar cual de los dos argumentos siguientes, si es que alguno, les parecía correcto*:

*Argumento 1*

*La primera secuencia (RVRRR) es más probable que la segunda (VRVRRR) porque la segunda secuencia es la misma que la primera con una V adicional al principio. De aquí que, cada vez que la segunda secuencia ocurre, la primera secuencia debe también ocurrir. Consecuentemente, tú puedes ganar sobre la primera y perder sobre la segunda, pero tú nunca puedes ganar sobre la segunda y perder sobre la primera.*

*Argumento 2*

*La segunda secuencia (VRVRRR) es más probable que la primera (RVRRR) porque las proporciones de R y V en la segunda secuencia se aproximan más que aquellas de la primera secuencia a las proporciones esperadas de R y V para un dado con cuatro caras verdes y dos rojas.*

Sorprendentemente en este caso 76% de los sujetos escogió el argumento extensional válido (argumento 1).

Tras los resultados de sus investigaciones Tversky y Kahneman han concluido que, en lo que a juicios de probabilidad refiere, los seres humanos no seguimos las reglas de la teoría estándar de la probabilidad; sino que más bien nos valemos de un cierto número de heurísticos, los cuales algunas veces nos permiten emitir juicios razonables mientras que otras veces nos conducen a errores sistemáticos en el razonamiento.

Ellos también destacan el hecho de que, mientras en los estudios sobre el razonamiento deductivo se ha observado que la gente frecuentemente falla en entender y aplicar principios lógicos abstractos, no obstante que parecen aplicarlos apropiadamente para el caso de problemas concretos en situaciones familiares; en el caso de sus estudios sobre la falacia de la conjunción la gente parece exhibir el patrón opuesto: ellos tienden a aceptar la regla de la conjunción en su forma abstracta, pero la desafían en problemas concretos (como es el caso del problema de Linda) donde la regla de la conjunción parece conflictuar con una impresión intuitiva. (Tversky y Kahneman, 1983; p. 311)

## 5.2 LA REPRESENTACION DE LA INFORMACION PROBABILISTICA.

Para ampliar un poco más nuestro panorama en relación a la investigación experimental sobre la falacia de la conjunción, será interesante considerar una investigación adicional que desafía la tesis de que la llamada *falacia de la conjunción* constituye verdaderamente un error en el razonamiento probabilístico.

Gerd Gigerenzer (1993), considera que la manera en que un problema de razonamiento probabilístico es planteado -esto es, las palabras utilizadas, la forma en que se dan las instrucciones y el contexto en el que el problema es presentado- da lugar a cierta representación de la información contenida en el problema. Dicha información, según Gigerenzer, puede representarse en la mente del sujeto a lo menos de dos maneras diferentes: 1) como probabilidad de un evento singular o grados de creencia en un evento singular, expresado en otros términos; o 2) como frecuencia relativa.

Toda vez que para la teoría matemática de la probabilidad, *probabilidad* se entiende como frecuencia relativa -es decir, como la frecuencia con que ocurre un evento en relación a una clase de referencia dada-, a la luz de la teoría matemática de la probabilidad tiene sentido hablar, por ejemplo, de la probabilidad de que Linda sea una cajera de banco dado que un porcentaje alto de mujeres en la comunidad donde Linda vive son cajeras de banco; pero no tiene sentido hablar simplemente de la probabilidad de que Linda sea una cajera de banco así sin más, sin haber definido una clase de referencia.

En tanto que la teoría de la probabilidad no se aplica a la probabilidad de eventos singulares, Gigerenzer argumenta, las leyes de la probabilidad no tienen aplicación alguna en el caso del problema de Linda -el cuál pide a los sujetos determinar la probabilidad de un evento singular-. Por lo tanto la llamada falacia de la conjunción en el caso de Linda, no puede implicar la violación por parte de los sujetos experimentales de ningún principio derivado de la teoría de la probabilidad. Desde luego, lo anterior puede sostenerse desde la perspectiva de una interpretación frecuentista de la probabilidad.

Por otra parte, Gigerenzer sostiene que los seres humanos, así como muchas otras especies animales, somos altamente sensibles a la frecuencia con que ocurren eventos en nuestro medio ambiente. Él especula que si los humanos tenemos alguna capacidad para el razonamiento estadístico, ésta tuvo que haber sido el resultado de la evolución por selección natural; y si nos preguntamos acerca de cuál es el tipo de representación de la información a que tendría que estar ajustado el algoritmo subyacente a dicha capacidad, muy posiblemente -afirma Gigerenzer- no sería a la probabilidad de eventos singulares, ni a probabilidades en términos de porcentajes (como -nos dice Gigerenzer- suele ser el caso en los experimentos típicos sobre el razonamiento humano), pues el desarrollo de tales conceptos requirió miles de años de alfabetización y de conocimiento de los números. Para él, en un mundo no

alfabetizado, la representación de la información tendría que darse en términos de frecuencias de eventos. Por ejemplo, 3 de 20, en lugar de 15% o de  $p=0.15$

Sobre la base de las consideraciones precedentes, Gigerenzer sustenta la hipótesis de que la falacia de la conjunción puede ser eliminada si el planteamiento del problema es capaz de evocar en el sujeto una representación de la información en términos de frecuencia.

A favor de lo anterior, Gigerenzer nos refiere los resultados obtenidos por Klaus Fiedler (1988), el cual presentó a un grupo de sujetos la siguiente versión del problema de Linda:

*Linda tiene 31 años, es soltera, abierta y muy brillante. Ella se licenció en filosofía. Como estudiante ella estaba profundamente interesada en temas de discriminación y justicia social, y también participó en protestas antinucleares.*

*Hay 100 personas que se adaptan a la descripción de arriba. ¿Cuántas de ellas son:  
cajeras de banco?*

*cajeras de banco y activas en el movimiento feminista?*

Para este problema, el porcentaje de violaciones de la regla de la conjunción descendió de 91% en la versión donde la información se representó como probabilidad de un evento singular, a 22% en la versión donde la información se representó como frecuencia.

El punto importante que Gigerenzer quiere destacar con todo esto es que hay una mucho mejor explicación del fenómeno conocido como falacia de la conjunción que la apelación a un heurístico de representatividad. Según Gigerenzer, si la mente se vale de un heurístico de representatividad para solucionar el problema de Linda, entonces los cambios en la representación de la información llevados a cabo por Fiedler en la versión frecuentista justo presentada, no tienen porqué afectar la ejecución de los sujetos, pues el grado de similitud entre la descripción de Linda y las oraciones B&F y B permanece el mismo. No obstante, como hemos visto, la presentación del problema en un formato de frecuencia da lugar a una disminución significativa de la violación de la regla de la conjunción. Para Gigerenzer la mejor explicación de este resultado es que los sujetos son capaces de hacer una distinción conceptual que desde hace tiempo ya había sido señalada por los filósofos, a saber, la distinción entre la probabilidad de un evento singular y frecuencias relativas. Esta distinción, nos dice Gigerenzer, es una distinción de fundamental importancia a la hora de evaluar los datos experimentales sobre juicios de probabilidad, pues nos permite clarificar que la llamada falacia de la conjunción, así como otros supuestos errores en el razonamiento probabilístico tales como sobre-confianza o la falacia de las tasas base<sup>12</sup> no pueden ser apropiadamente considerados como violaciones de principios de la teoría de la probabilidad, consiguientemente, no pueden ser considerados como fallas en el razonamiento.

En conclusión, lo que Gigerenzer está argumentando es que los sujetos pueden concebir por lo menos dos sentidos distintos de probabilidad, y que la forma en que el problema se plantea, así como su contenido y contexto, dan los indicios clave para que el sujeto se represente la información del problema en un sentido o en otro, de los cuales, ambos son igualmente legítimos.

## **6. CONCLUSIÓN.**

En el presente capítulo se han presentado las principales teorías sobre el razonamiento humano desarrolladas en la psicología. Hemos visto que dos de ellas, TERP y TCS, se comprometen con la tesis de que los principios de razonamiento son principios de dominio específico. Concretamente, hemos visto que TERP postula la existencia de principios de razonamiento aplicables exclusivamente a “problemas de permiso”, principios que juntos constituyen lo que los proponentes de TERP llaman un “esquema de permiso”. Por su parte, TCS, postula la existencia de principios de razonamiento aplicables exclusivamente a “problemas de contrato social” los cuales involucran la detección de posibles tramposos. A favor de lo anterior los proponentes de estas teorías han realizado un sin número de experimentos, de los cuales, he presentado los más relevantes. Por un lado, los resultados de estos experimentos sugieren fuertemente que hay reglas, subyacentes a la competencia humana de razonamiento, que son de dominio específico. Por otro lado, dichos resultados nos dan elementos para dudar que haya un cálculo formal que pueda modelar apropiadamente las inferencias de los sujetos, simplemente porque, como se ha visto, el contenido parece crucial.

En este capítulo se ha presentado también un breve panorama de la investigación en el ámbito del razonamiento probabilístico. Por un lado, se han visto los resultados experimentales de Tversky y Kahneman, los cuales nos muestran desviaciones sistemáticas, por parte de los sujetos, de la llamada regla de la conjunción. La explicación que los autores dan a sus resultados es que los sujetos, más que razonar de acuerdo con la regla de la conjunción, razonan de acuerdo con un principio heurístico: el heurístico de representatividad. Si Tversky y Kahneman están en lo correcto, entonces, nos enfrentamos aquí, del mismo modo que en el caso del razonamiento deductivo, al hecho de que la ejecución de los sujetos se desvía sistemáticamente del principio normativo considerado como correcto y, al hecho de que esta desviación parece ser atribuible a la competencia de razonamiento de dichos sujetos.

Por otro lado, se ha visto la propuesta de Gigerenzer a favor de la tesis de que los sujetos son sensibles a diferencias en la manera de plantear la información probabilística: ya sea en



términos de la *probabilidad de un evento singular* o ya sea en términos de *frecuencia relativa*. Él propone, además, apoyado en evidencia experimental, que cuando en una tarea de razonamiento probabilístico se plantea la información en términos de frecuencia, la ejecución de los sujetos mejora significativamente, al punto de que la falacia de la conjunción puede desaparecer.

La propuesta de Gigerenzer, así como la evidencia experimental que él nos refiere, apuntan a la existencia de una regla, subyacente en la competencia de razonamiento de los sujetos, la cual es de dominio específico. Esto parece ser así, dado el hecho de que una facilitación en problemas como el de Linda, sólo puede ser observada cuando la información y el contexto propios del problema, sugieren una interpretación en términos de *frecuencia*.

## NOTAS.

---

<sup>7</sup> Aquí los autores señalan que se debe distinguir entre arbitrariedad de la relación y el carácter abstracto o concreto de las entidades involucradas en la regla. Para el caso de esta regla, ellos consideran que ésta especifica una relación arbitraria entre entidades concretas.

<sup>8</sup> Gigerenzer y Hug (1992) aclararon posteriormente que la estructura costo-beneficio de un contrato social no es una condición suficiente para producir una facilitación en la tarea de selección. Ellos argumentan que la condición verdaderamente crítica para lograr una facilitación, es el hecho de que el sujeto que realiza la tarea de selección pueda asumir la perspectiva de alguien que puede ser engañado. Esto último significa que la estructura de contrato social de una regla no necesariamente activará un algoritmo de detección de tramposos. De cualquier modo, nos dicen Gigerenzer y Hug, aunque Cosmides no haya hecho una distinción entre “contrato social” y “algoritmo para la detección de tramposos”, dicha distinción subyace a la teoría de contrato social.

<sup>9</sup> Cabe mencionar que los primeros estudios de la regla de la conjunción datan de 1974. En el artículo citado de 1983, Tversky y Kahneman replican y amplian sus estudios previos.

<sup>10</sup> Debe señalarse que los sujetos participantes en los experimentos previamente descritos (con excepción del primer experimento en el que participaron tres grupos de respondientes) eran estudiantes estadísticamente ingenuos.

<sup>11</sup> Por *heurístico* ha de entenderse un procedimiento computacional, no algorítmico, el cual es lo suficientemente rápido y frugal (en lo que refiere al uso de recursos computacionales) como para poder ser realizado en un tiempo lo suficientemente corto de modo que sea posible que la toma de decisiones y las consecuentes acciones puedan ocurrir en el tiempo real de una vida humana. Este

---

procedimiento computacional debe también ser lo suficientemente confiable como para llevarnos a generar las respuestas apropiadas que el ambiente demanda, la mayor parte del tiempo. Esto último implica que puede haber ocasiones en que la aplicación de ese mismo procedimiento no nos conduzca a las respuestas apropiadas.

<sup>12</sup> “Sobre-confianza” se refiere al fenómeno consistente en la tendencia de los sujetos a confiar de más en su pretendido conocimiento. Esto se refleja en el hecho de que, cuando ellos realizan una prueba de conocimientos generales, ellos suelen tener una estimación más alta de su propia ejecución en dicha prueba, de lo que de hecho es el caso.

El fenómeno conocido como “falacia de las tasas base” consiste en lo siguiente:

A un grupo de estudiantes de Medicina se le presenta el siguiente problema:

*“Si una prueba para detectar una enfermedad cuya prevalencia es 1/1000 tiene una tasa de falsos positivos del 5%, ¿cuál sería entonces la probabilidad de que una persona que obtuvo un resultado positivo tenga efectivamente la enfermedad, asumiendo que tú no sabes nada acerca de los signos o síntomas de la persona?”*

La respuesta de la mayoría de los sujetos a este problema es 0.95 (95%), no obstante esta respuesta se considera incorrecta. El error cometido por los sujetos deriva de apoyar su razonamiento únicamente en la información de que la tasa de falsos positivos es del 5% y de la casi nula consideración de la llamada tasa base (1/1000). Por eso se ha llamado a este supuesto sesgo en el razonamiento probabilístico, rechazo de las tasas base. (ver Gigerenzer 1991 y 1993).

## **CAPITULO TRES.**

### **ACERCA DE LA UNIVERSALIDAD DE LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO.**

Como bien señala Stephen P. Stich (1990), uno de los proyectos centrales y tradicionales de la epistemología ha sido el análisis de nuestros métodos de investigación a fin de dilucidar cuáles de dichos métodos son correctos y cuáles no lo son, en el sentido de que garanticen la obtención de conocimiento genuino. El producto final de una indagación epistemológica abocada a tal proyecto, será el establecimiento de un conjunto de principios normativos que nos permitan discernir cuándo los procedimientos que utilizamos en aras de generar conocimiento son epistémicamente aceptables. Puesto que el razonamiento es el proceso cognoscitivo fundamental mediante el cual los seres humanos generamos y revisamos nuestros sistemas de creencias, los epistemólogos han puesto especial interés en el establecimiento de principios normativos que nos permitan evaluar nuestros procesos de razonamiento.

Un supuesto ampliamente generalizado en relación a dichos principios es que, cualesquiera que éstos sean, ellos deben tener un carácter universal. Lo que esto quiere decir es que se presupone que los principios normativos de razonamiento deben ser aplicables en la evaluación de los procesos de razonamiento de cualquier ser humano “normal” en cualquier tiempo y lugar. La razón para asumir que los principios normativos de razonamiento deben ser universales es que la posición alternativa –el relativismo- resulta a primera vista inaceptable. Pues si los principios normativos que son apropiados para ciertas personas pueden no serlo para otras, y a la inversa, y cada cual se rige por sus propios principios, entonces parecería que cualquier principio normativo sería igualmente válido. El relativismo nos conduciría así a la posición de que en materia de razonamiento “todo vale”, lo cual a final de cuentas implicaría que no hay en realidad principios normativos.

El rechazo al relativismo es un pilar fundamental para la tesis de que los principios normativos de razonamiento son universales. No obstante, los filósofos que se adhieren a dicha tesis han aportado muchas otras razones, igualmente poderosas, a favor de que los principios normativos de razonamiento deben ser universales. Entre estos filósofos podemos contar a Donald Davidson.

El presente capítulo está dedicado a la presentación de los argumentos de Davidson a favor de la tesis de que debe existir un núcleo importante de racionalidad compartida por todos los

seres humanos. La postura de Davidson es que no pueden existir sujetos racionales que utilicen reglas de razonamiento que sean fundamentalmente distintas de las nuestras.

Una vez presentados los argumentos de Davidson, el objetivo será intentar minar su tesis. Se tratará de mostrar, sobre la base de argumentos meramente conceptuales, que sí es posible que haya sujetos racionales que utilicen reglas de razonamiento distintas de las nuestras y que, por tanto, los principios normativos de razonamiento no son necesariamente universales.

## **1. EL ARGUMENTO DE DAVIDSON.**

De los escritos de Davidson es posible extraer el siguiente argumento:

- a) Si hubiera sujetos que utilizaran principios normativos de razonamiento fundamentalmente distintos de los nuestros, entonces no podríamos interpretarlos.
- b) La idea de un sujeto que sea racional y al mismo tiempo no podamos interpretar es ininteligible.
- c) De (b) se sigue que todo sujeto racional es en principio interpretable (Inteligibilidad universal).
- d) Por lo tanto, de (a) y (c) se sigue que: No hay sujetos racionales que utilicen principios normativos de razonamiento fundamentalmente diferentes de los nuestros, lo cual significa que los principios normativos de razonamiento son universales.

En lo que sigue iremos desglosando y analizando los argumentos que es posible extraer desde los textos de Davidson para dar soporte, a su vez, a cada una de las premisas del argumento arriba presentado.<sup>13</sup>

Veamos pues los argumentos para sustentar (a).

### **1.1 ¿POR QUÉ RAZÓN SI HUBIERA SUJETOS QUE UTILIZARAN PRINCIPIOS NORMATIVOS DE RAZONAMIENTO FUNDAMENTALMENTE DISTINTOS DE LOS NUESTROS NO PODRIAMOS INTERPRETARLOS?**

Responder al cuestionamiento recién planteado nos remite forzosamente a la teoría de la interpretación que Davidson propone. Por tal razón dedicaré una parte importante de esta sección a la presentación de la teoría davidsoniana de la interpretación. Una vez que haya quedado suficientemente claro lo que Davidson tiene en mente cuando habla de “interpretación”, estaremos en condición de responder la pregunta que aquí nos ocupa.

Davidson considera que una teoría adecuada de la interpretación (i.e. del significado) debe cumplir tres requisitos fundamentales (“Semantics For Natural Languages”; en Davidson, 1984):

- 1) Dicha teoría debe, en principio, **dar el significado de cada oración** del lenguaje del sujeto que queremos interpretar.
- 2) Es un hecho que el número de oraciones significativas que se pueden generar en un lenguaje es potencialmente infinito; no obstante, una teoría de la interpretación **debe tener un carácter finito**, por lo tanto, dicha teoría debe dar cuenta de cómo, a partir de un conjunto de elementos finitos (esto es, el conjunto de palabras de que se compone una lengua), así como de la aplicación de ciertas reglas, es posible construir un número infinito de oraciones significativas. En otras palabras, una teoría de la interpretación **debe dar cuenta de la productividad del lenguaje**.
- 3) Puesto que la única manera en que una teoría puede lograr lo anterior es postulando que el conjunto de oraciones posibles puede formarse por la distinta composición, de acuerdo con ciertas reglas, de los elementos finitos del lenguaje, dicha teoría **debe ser composicional**. Así pues hay un tercer requisito que una teoría adecuada de la interpretación debe satisfacer: dicha teoría debe mostrar cómo el significado de cada oración depende del significado de las palabras de que se compone.

Davidson sostiene que una teoría de la verdad al estilo de Tarski puede cumplir los requisitos anteriores. Recordemos que la teoría de la verdad de Tarski consiste de un número finito de axiomas sobre la base de los cuales es posible generar (infinitamente), para cualquier oración ‘s’ en el lenguaje, una oración de la forma:

s es verdadera si y sólo si p

donde ‘s’ representa a la oración en el lenguaje objeto para la cual están siendo dadas condiciones de verdad, y ‘p’ representa una oración usada en el metalenguaje. Tarski denomina oraciones-V a estos bicondicionales.

La teoría tarskiana puede así dar cuenta de la productividad del lenguaje; además, en tanto que los axiomas contenidos en la teoría tarskiana de la verdad establecen cómo determinar cuáles son las oraciones-V de oraciones compuestas a partir de las oraciones-V de sus componentes, una teoría de la verdad para un lenguaje natural al estilo Tarski podrá mostrar cómo el significado de la oración depende de su estructura, cumpliendo así el requisito de composicionalidad. Queda ahora mostrar cómo el conocimiento de una teoría de la verdad para un cierto lenguaje L, nos permitirá generar interpretaciones o dar sentido a las oraciones emitidas por los hablantes de L.

En relación a esto último, Davidson nos dice que es un error asumir que todo lo que se requiere para dilucidar el significado de una cierta oración está contenido en el bicondicional expresado en la oración-V para esa oración. Más bien, lo que nos permite generar interpretaciones es el proceso por el cual sometemos a prueba a las oraciones-V. En otras palabras, lo que nos permite generar interpretaciones es el proceso por el cual determinamos que nuestra teoría de la verdad para el lenguaje que deseamos interpretar es formalmente correcta y empíricamente adecuada. Veamos pues en qué consiste dicho proceso de prueba. Sólo que antes será necesario establecer algunas cuestiones pertinentes.

En primer lugar debe quedar claro cuál es el tipo de evidencia sobre la que puede sustentarse una teoría de la interpretación. Davidson declara que la evidencia a favor de una teoría de la interpretación *debe ser evidencia plausiblemente disponible al intérprete*. Por un lado, parecería fácil suponer que tal evidencia puede consistir en una muestra apropiada de instancias de interpretaciones que se reconocen como correctas sobre la base de la captación de los significados del hablante. El problema con esto es que justamente lo que se espera de una teoría de la interpretación es que ésta le permita, a un intérprete que no tiene entendimiento alguno del lenguaje en cuestión, conocer el significado de las emisiones de un hablante. Es claro que para un intérprete que desconoce totalmente el lenguaje del interpretado –esto es, el intérprete radical- la evidencia de que ciertas interpretaciones son correctas -en tanto que asignan el sentido propio de las emisiones del hablante- no está disponible. Por otro lado, podríamos estar tentados a buscar evidencia a favor de una teoría de la interpretación apoyándonos en información acerca de los deseos, intenciones y creencias de los hablantes. A este respecto Davidson objeta que esta estrategia está totalmente desencaminada dada la conexión que hay entre creencia y significado: no es posible atribuir detalladamente creencias a un hablante sin conocer el significado de sus palabras; de igual manera no es posible conocer el significado de las palabras de un hablante sin atribuirle ciertas creencias, deseos y/o intenciones. En una palabra: la atribución de creencias y la asignación de significado forman parte de un mismo proyecto, ambos juegan roles entremezclados en la interpretación del habla. La siguiente cita de Davidson dejará más claro este punto:

*“We have still to say what evidence is available to an interpreter- evidence, we now see, that T-sentences are true. The evidence cannot consist in detailed descriptions of the speaker’s beliefs and intentions, since attributions of attitudes, at least where subtlety is required, demand a theory that must rest on much the same evidence as interpretation. The interdependence of belief and meaning is evident in this way: a speaker holds a sentence to be*

*true because of what the sentence (in his language) means, and because of what he believes. Knowing that he holds the sentence to be true, and knowing the meaning, we can infer his belief; given enough information about his beliefs, we could perhaps infer the meaning". (En "Radical Interpretation"; en Davidson 1984, pp.134-135)*

Dadas las consideraciones anteriores es posible concluir que la evidencia apropiada para sustentar una teoría de la interpretación no puede estar basada en ningún conocimiento previo por parte del intérprete, ni de las creencias ni de los significados de las oraciones de aquéllos a quienes se está tratando de interpretar, pues simplemente este tipo de evidencia no está disponible a un intérprete radical. En cambio, un tipo de evidencia que sí está disponible a un intérprete radical es la actitud de un hablante de sostener una oración como verdadera. En tanto que dicha actitud puede ser identificada por un intérprete que no sepa de antemano como interpretar el lenguaje en cuestión, Davidson argumenta que éste es el tipo de evidencia en la que debe sustentarse una teoría de la interpretación.

Tener en mente la conexión entre creencia, significado y verdad, nos permite definir con mayor claridad en qué consiste el trabajo de interpretación: en la medida en que el intérprete va construyendo una teoría de la verdad para el lenguaje de aquellos a quienes desea interpretar -apoyándose en observaciones de las oraciones a las que el hablante está dispuesto a asentir, así como de las condiciones en que lo hace- él va simultáneamente construyendo una teoría de la creencia y una teoría del significado. En resumen, el trabajo del intérprete es inferir las creencias y los significados del hablante al que se quiere interpretar a partir de lo que dicho hablante considera como verdadero.

Pero ¿cómo inferir la creencia y el significado a partir solamente del conocimiento de una teoría de la verdad? Un poco más arriba he apuntado que –según Davidson- lo que nos permite generar interpretaciones a partir de una teoría de la verdad es el proceso por el cual sometemos a prueba dicha teoría:

*"It is tempting, therefore, simply to say that a T-sentence 'gives the meaning' of a sentence. Not, of course, by naming or describing an entity that is a meaning, but simply by saying under what conditions an utterance of the sentence is true.*

*But on reflection it is clear that a T-sentence does not give the meaning of the sentence it concerns....*

*A move that might seem helpful is to claim that is not the T-sentence alone, but the canonical proof of a T-sentence, that permits us to interpret the alien sentence." (En "Radical Interpretation"; en Davidson 1984, p.138)*

Enfoquémonos pues en la siguiente cuestión: ¿Cómo sometemos a prueba una teoría de la verdad? Responder a esta pregunta es una tarea algo compleja, así que vayamos paso a paso. Podemos empezar diciendo que someter a prueba una teoría de la verdad implicará ir sometiendo a prueba a las oraciones-V generadas por la teoría a fin de establecer que éstas son correctas. Ahora bien, establecer que las oraciones-V generadas por nuestra teoría de la verdad son correctas requerirá establecer que dichas oraciones cumplen con ciertas restricciones formales y empíricas. De aquí que establecer que las oraciones-V son correctas requerirá dos tipos de prueba: una formal y una empírica.

Es importante aclarar que al hablar aquí de dos tipos de prueba no se pretende decir que éstas se realizan con independencia una de la otra, o que una se realiza primero y la otra después; se trata más bien de un proceso simultáneo en el que la evidencia se va ajustando a modo de poder satisfacer las restricciones formales y, a la inversa, las conexiones lógicas que se presuponen entre las oraciones del lenguaje se van ajustando a fin de adecuarse a la evidencia. Si en lo que sigue se presenta por separado lo que implica una prueba formal y una prueba empírica de las oraciones-V, es sólo con fines expositivos.

Empecemos pues por establecer en qué consiste la prueba formal de las oraciones-V. En *Semantics for Natural Languages* (Davidson, 1984) Davidson nos dice que la prueba del bicondicional expresado en una oración-V debe mostrar paso a paso cómo el valor de verdad de la oración 's' (una oración particular cualquiera emitida por el hablante al que se intenta interpretar) depende de su estructura. En tanto que nosotros –como intérpretes– hemos establecido ya las condiciones de verdad para una gran parte de las oraciones del lenguaje del hablante (de aquí en adelante "H") y no sólo para 's', entonces también hemos determinado ya cómo el valor de verdad de esas otras oraciones depende de su estructura. Esta descripción estructural de una buena parte de las oraciones del lenguaje de H nos permite, a la vez, darnos cuenta de las relaciones lógicas que unas oraciones mantienen con otras. De este modo, las pruebas conjuntas de las oraciones-V nos permiten construir un mapa general de la estructura lógica del lenguaje de H considerado como un todo. Dicho mapa estructural nos permitirá conocer el rol que cierta oración juega en la estructura total del lenguaje. Tendremos además elementos para asumir que la oración-V para 's' ha sido generada desde nuestra teoría de la verdad de una manera formalmente apropiada y que sus relaciones con otras oraciones-V también generadas por la teoría son consistentes.

No obstante, la prueba de las oraciones-V debe ir más allá de establecer que las restricciones formales se cumplen. Resta aún determinar si la teoría es empíricamente adecuada. Establecer que una oración-V (asumamos que se trata de la oración-V para 's') es empíricamente adecuada requiere determinar que dicha oración-V se adapta a la evidencia relativa a las condiciones en las cuales H asiente o disiente de la oración 's'. Pero la



evidencia a favor de nuestra oración-V para 's' no se limita sólo a la evidencia señalada. Nuestra oración-V debe además adaptarse a la evidencia que se tiene relativa a las condiciones en las cuales el hablante asiente o disiente a otras oraciones del lenguaje, las cuales mantienen conexiones lógicas con la oración 's' en cuestión. En conclusión, someter a prueba empírica una cierta oración-V particular implica confrontar a ésta con gran parte de, sino es que con toda, la evidencia disponible hasta ese momento para el intérprete.

Tras lo anteriormente descrito, hemos de darnos cuenta de que las restricciones formales y empíricas que debe satisfacer una teoría de la verdad para ser una teoría aceptable, tienen un carácter holista: la totalidad de las oraciones-V deben adecuarse a la totalidad de la evidencia y deben también constituir ellas mismas una estructura lógica coherente y consistente.

Volvamos ahora al cuestionamiento que originalmente motivó nuestro interés por la manera en que sometemos a prueba una teoría de la verdad. La pregunta era ¿cómo el conocimiento de una teoría de la verdad para un cierto lenguaje L, nos permitirá generar interpretaciones o dar sentido a las oraciones emitidas por los hablantes de L? Bueno, la respuesta de Davidson a esta cuestión es que una teoría de la verdad que cumpla las restricciones formales y empíricas arriba señaladas producirá de hecho interpretaciones aceptables de los hablantes de L. Nuevamente podremos preguntarnos porqué esto es así. La razón es básicamente que Davidson asume que el significado está dado en gran medida por el rol que una oración juega en la estructura total del lenguaje, y puesto que, como hemos visto, el procedimiento mediante el cual sometemos a prueba las oraciones-V de nuestra teoría de la verdad nos lleva a determinar precisamente esto, dicho procedimiento nos permitirá también inferir el significado de las emisiones del hablante (al mismo tiempo que es posible atribuirle creencias).

A fin de que todo lo anterior quede más claro, permítaseme citar nuevamente a Davidson:

*“A final suggestion along these lines is that we can interpret a particular sentence provided we know a correct theory of truth that deals with the language of the sentence. For then we know not only the T-sentence for the sentence to be interpreted, but we also ‘know’ the T-sentences for all other sentences; and of course, all the proofs. Then we would see the place of the sentence in the language as a whole, we would know the role of each significant part of the sentence, and we would know about the logical connections between this sentence and others”. (En “Radical Interpretation”; en Davidson 1984, pp.138-139)*

Hasta aquí se ha buscado establecer cómo es que una teoría de la verdad al estilo de Tarski puede fungir como una teoría de la interpretación, según Davidson propone. Sin embargo, la exposición de la teoría davidsoniana de la interpretación no concluye aquí, nos resta todavía presentar otros aspectos fundamentales de dicha teoría.

Se ha dicho ya que una teoría de la interpretación debe estar sustentada en evidencia empírica, se ha dicho también cuál es el tipo de evidencia en la que puede y debe sustentarse dicha teoría: a saber en la actitud de un hablante de asentir ante ciertas oraciones -dadas ciertas condiciones-, así como a disentir ante esas mismas oraciones cuando las pretendidas condiciones no se cumplen. Considérese el siguiente ejemplo:

Imaginemos que nosotros somos intérpretes radicales y que hemos observado en varias ocasiones que H suele emitir ciertos sonidos –pongamos por caso “#+\*&”- cada vez que él se encuentra frente a árboles de manzana. Observamos también que estos sonidos no se emiten frente a otro tipo de árboles y que de hecho H disiente de “#+\*&” cuando no hay frente a él árboles de manzana. Estas observaciones nos llevan a presuponer que “#+\*&” podría ser interpretado como “Ahí manzano”. Lo que sigue entonces es determinar si tal interpretación es correcta, para lo cual será necesario corroborar que la oración-V para “#+\*&” cumple con las restricciones formales y empíricas que una teoría aceptable de la interpretación debe satisfacer. Imaginemos que efectivamente este es el caso. Asumimos entonces que “#+\*&” puede confiablemente entenderse como diciendo “Ahí manzano”. Sin embargo, no obstante la adecuación formal y empírica de nuestra interpretación, siempre será posible dar una interpretación alternativa que se adecue igualmente bien con la evidencia disponible. Por ejemplo, “#+\*&” tal vez podría entenderse como “Ahí árbol hermoso”, o “Ahí fruta sabrosa”, lo cual implica que siempre habrá una indeterminación en nuestras interpretaciones y consecuentemente que siempre será posible generar más de una teoría interpretativa satisfactoria para las emisiones de un hablante. Si esto es así, cabe entonces preguntarse cómo elegir entre todas aquéllas teorías interpretativas que pueden ser igualmente aceptables. La respuesta de Davidson a esta cuestión es que, ante la indeterminación de la interpretación, la única opción para el intérprete que busca elegir entre interpretaciones posibles es apegarse a un principio general, a saber, al *Principio de Caridad*, el cual se enuncia en seguida:

***Principio de Caridad.***- el intérprete siempre debe elegir aquéllas interpretaciones que atribuyan a aquél al que se interpreta creencias verdaderas y consistentes, tanto como esto sea posible.

Dado lo establecido en dicho principio, en aquéllas situaciones en que nuestra interpretación nos lleve a asumir que H sostiene creencias falsas o absurdas, lo consecuente será más bien rechazar nuestra interpretación como incorrecta.

Apegarse al principio de caridad no es una opción para el intérprete en el sentido de que éste pueda decidir ser o no ser caritativo en la interpretación. Davidson nos dice que la adhesión al principio de caridad es forzosa si es que realmente queremos entender a otros.

*“Charity is forced on us; whether we like it or not, if we want to understand others, we must count them right in most matters”. (En “On the very idea of a conceptual scheme”; en Davidson, 1984, p. 197)*

La formulación del principio de caridad levanta dos cuestiones: la primera tiene que ver con la justificación de dicho principio, esto es, con las razones por las cuales nosotros hemos de asumir que aquellos a quienes tratamos de entender tienen creencias verdaderas y consistentes la mayor parte del tiempo ¿acaso las personas no pueden tener fallas masivas en el razonamiento y sustentar una gran cantidad de creencias inconsistentes, o en todo caso falsas? La segunda cuestión tiene que ver con lo siguiente: si el principio de caridad nos exhorta a considerar a los sujetos a los que se interpreta como estando –generalmente– en lo correcto, podemos entonces preguntarnos: en lo correcto ¿según cuáles estándares de racionalidad?

Veamos lo que Davidson nos dice en relación a estas interrogantes. Consideremos pues el primer cuestionamiento: ¿por qué debemos asumir que los sujetos a quienes se busca interpretar sustentan generalmente creencias verdaderas y consistentes so pena de que la interpretación no sea posible? Davidson sostiene que la razón de que esto sea así deriva de una conexión conceptual existente entre creencia (y demás actitudes proposicionales) y racionalidad. Él sostiene que las creencias, intenciones, deseos, etc. pueden ser identificados dadas sus relaciones lógicas con otras actitudes proposicionales y dadas sus relaciones causales con los objetos y los eventos en el mundo. Pensemos por ejemplo en una persona, digamos María, la cual asevera que “el agua hirviendo quema”. Dada esta aseveración, nosotros no dudamos en atribuirle la creencia de que el agua hirviendo quema. No obstante María también asevera que “las quemaduras son dolorosas y deben evitarse” y que “bañarse en agua hirviendo es placentero”; María además intenta bañarse en el agua hirviendo. Cómo podemos darnos cuenta, las creencias que hemos atribuido a María y su acción son inconsistentes. Esto obviamente nos llevará a poner en duda el hecho de que María genuinamente crea que el agua hirviendo quema. En realidad, en tales circunstancias ya no sabríamos que creencia o creencias atribuir a María. Este es pues el punto que

Davidson quiere establecer: si la creencia de que “el agua hirviendo quema” no mantiene las relaciones lógicas apropiadas con otras creencias, así como con las acciones de María, entonces dicha creencia no puede contar genuinamente como la creencia de que el agua hirviendo quema.

La moraleja de esta historia sobre María es que si una persona sustentara un sistema de creencias masivamente irracional, entonces simplemente no podríamos atribuirle creencias (ni otras actitudes proposicionales). Pero a alguien a quien no pueden atribuirse creencias – nos dice Davidson- tampoco puede atribuírsele pensamiento, por lo tanto ese alguien no podría considerarse, a final de cuentas, como un agente cognoscitivo.

De lo anterior se desprende que ser racional es una condición necesaria de tener creencias. Esta es pues la conexión que Davidson señala entre creencia y racionalidad, la cual se establece con claridad en la siguiente cita:

*“...these obvious logical relations amongst beliefs, desires and intentions; between beliefs and the world, make beliefs the beliefs they are; therefore they cannot in general lose these relations and remain the same beliefs. Such relations are constitutive of the propositional attitudes.” (Davidson, 1985, pp. 351-352)*

Con lo anterior queda ya resuelto nuestro primer cuestionamiento: la razón por la que hemos de asumir que los sujetos a los que deseamos interpretar sustentan creencias consistentes y verdaderas en la mayoría de los casos es que, de no ser así, sencillamente no podríamos considerarlos como agentes cognoscitivos y por ende no tendría caso el pretender interpretarlos. Si nosotros queremos interpretar a otros, es un requisito previo asumir que esos otros son agentes cognoscitivos, que tienen creencias, deseos, intenciones, etc.; y puesto que ser racional es una condición necesaria de tener creencias y otras actitudes proposicionales, entonces nosotros debemos asumir forzosamente que aquellos a quienes buscamos interpretar son racionales.

Vayamos ahora al segundo cuestionamiento al que ha dado lugar la formulación del principio de caridad. Estamos ya claros en el punto de que, cómo intérpretes, debemos elegir aquellas interpretaciones que hagan a nuestros interpretados sujetos racionales. Ahora queda dejar claro cuáles son los principios normativos de razonamiento según los cuales hemos de determinar la racionalidad de los sistemas de creencias de aquellos a quienes interpretamos. La posición de Davidson a este respecto es que los principios normativos de razonamiento que todos los seres humanos suscribimos –sepámoslo o no, nos guste o no- son los principios establecidos por la teoría de la elección racional, la lógica clásica y principios como

el de evidencia total en el caso del razonamiento inductivo. En relación a estos principios Davidson nos dice:

*“These are principles shared by all creatures that have propositional attitudes or act intentionally; and since I am (I hope) one of those creatures, I can put it in this way: all thinking creatures subscribe to ‘my’ basic standards or norms of rationality. This sounds sweeping, even authoritarian, but it comes to no more than this, that it is a condition of having thoughts, judgments, and intentions that the basic standards of rationality have application.” (Davidson, 1985, p. 351)*

Aquí Davidson hace alusión a la conexión entre racionalidad y creencia de la que justo se ha hablado un poco más arriba. Pero ahora podemos agregar que las conexiones entre creencias y demás estados intencionales que se consideran apropiadas, son aquéllas que son acordes con la teoría de la elección racional, la teoría de la probabilidad y, fundamentalmente, con la lógica estándar de primer orden. El compromiso de Davidson con los principios que se desprenden de la lógica estándar, como los principios normativos correctos, es perfectamente entendible si consideramos que su teoría del significado y de la interpretación está basada en la teoría tarskiana de la verdad, la cual subsume a la lógica estándar de primer orden.

En tanto que suscribir los principios normativos de razonamiento correctos es una condición necesaria de ser un agente racional, y en tanto que ser racional es una condición necesaria de ser un agente cognoscitivo, se sigue entonces que suscribir los principios normativos que emanan de la lógica y los cálculos ya mencionados es una condición necesaria de ser un agente cognoscitivo. De lo anterior se desprende que si un sujeto cuenta como un agente cognoscitivo y por ende puede ser interpretado, entonces dicho sujeto razona de acuerdo con nuestros propios principios normativos de razonamiento (en tanto que nosotros también somos agentes cognoscitivos). Es así que, como lo señala Davidson en la cita precedente, “toda criatura pensante suscribe mis estándares básicos o mis normas de razonamiento”.

Una consecuencia importante que también se desprende de todo lo anterior es la siguiente: si un sujeto no suscribe los principios normativos de razonamiento ya señalados, entonces simplemente dicho sujeto no cuenta como un agente cognoscitivo. De aquí la idea de que un sujeto que utilizará principios normativos de razonamiento distintos de los nuestros no podría ser interpretado.

## 1.2 ¿POR QUÉ RAZÓN LA IDEA DE UN SUJETO QUE SEA RACIONAL Y AL MISMO TIEMPO NO PODAMOS INTERPRETAR ES ININTELIGIBLE?

Las razones de Davidson para declarar que no es posible dar sentido y claridad a la idea de un sujeto racional al que no podamos interpretar, nos remiten a las razones que él esgrime para rechazar la idea de “esquema conceptual”.

Según los adherentes a la idea de que los seres humanos nos movemos en el marco de diferentes “esquemas conceptuales”, éstos últimos constituyen sistemas de categorías a partir de los cuales se organiza y da sentido a la experiencia. En tanto que los seres humanos pertenecientes a diferentes culturas o épocas pueden poseer esquemas conceptuales diferentes, sus visiones del mundo pueden diferir drásticamente al punto de que la comunicación entre ellos se vuelva sumamente difícil, sino es que imposible. La razón de estas fallas en la comunicación reside en el carácter no homologable de las categorías con las que el suscriptor de cada uno de los diferentes esquemas conceptuales organiza la experiencia. En otras palabras, las respectivas clasificaciones de los distintos objetos que pueblan el mundo no son equiparables, de modo que muchos de los conceptos del suscriptor de un esquema no se pueden expresar en el lenguaje del suscriptor de un esquema conceptual diferente, y a la inversa. El simpatizante de los esquemas conceptuales suele afirmar también que no sólo las categorías de objetos pueden diferir de un esquema a otro, sino que incluso los objetos mismos que llegan a postularse como existentes pueden variar dependiendo de los distintos esquemas conceptuales. De este modo la realidad misma, lo que cuenta como real, se constituye a partir de un esquema conceptual.

Como podemos darnos cuenta, la aceptación de una visión como esta implica la aceptación de un relativismo conceptual. Tesis, ésta última, a la que Davidson se opone rotundamente, pues según él nosotros no podemos hacer sentido de la idea de una diferencia conceptual masiva. La razón de esto es básicamente que la única manera de identificar diferencias conceptuales entre nosotros y cualesquiera otros sujetos es que haya un sistema común de conceptos sobre el cual representarnos estas diferencias. Pero si tal sistema común existe, entonces nuestros respectivos sistemas de conceptos no son genuinamente inequ岸parables ni radicalmente diferentes. Por lo tanto, cualquier intento por establecer diversidad conceptual se autorrefuta. Esta misma conclusión puede presentarse en los siguientes términos: “cualquier intento por establecer no intertraducibilidad se autorrefuta”.

Esta segunda formulación de la conclusión precedente, obedece a la propuesta de Davidson de identificar esquemas conceptuales con conjuntos de lenguajes intertraducibles. Partiendo de esta propuesta podemos asumir que lo que se quiere decir cuando se afirma que dos sujetos tienen distintos esquemas conceptuales es justamente que ellos hablan lenguajes que fallan en intertraducirse. En este punto, la estrategia de Davidson para rechazar la

posibilidad de divergencia conceptual radical será negar la posibilidad de fallas de traducción entre lenguajes posibles. Específicamente, Davidson analiza dos casos: el caso de falla total y el caso de falla parcial de traducción.

En relación a la posibilidad de una *falla total* de traducción, entendida ésta como el caso en el que ni siquiera una mínima parte de un lenguaje puede ser traducida en otro, Davidson nos dice:

*“...nothing, it may be said, could count as evidence that some form of activity could not be interpreted in our language that was not at the same time evidence that that form of activity was not speech behavior. If this were right, we probably ought to hold that a form of activity that cannot be interpreted as language in our language is not speech behavior.”*  
(En “On the very idea of a conceptual scheme”; en Davidson, 1984, pp. 185-186)

Como vemos, Davidson establece aquí la traducibilidad a un lenguaje familiar como un criterio de lingüisticidad (*languagehood*). Sus razones a favor de este criterio derivan de la estrecha conexión entre atribuir significado a las emisiones de un hablante y atribuirle creencias, deseos e intenciones. Como se recordará, estas actividades forman parte de un mismo proceso en el que significado y creencia se fijan simultáneamente, de modo que ninguna de estas actividades puede completarse antes que la otra sea también completada. Así las cosas, si nosotros no podemos hacer sentido de las emisiones de un hablante, dada la no intertraducibilidad de sus emisiones a nuestro lenguaje, entonces nosotros tampoco podemos atribuirle creencias, deseos, intenciones, etc. y si nosotros no podemos hacer esto último, entonces no tenemos ningún elemento para sostener que los sonidos que él emite constituyen en realidad oraciones de algún lenguaje. En resumen, si un sujeto emitiera sonidos que de ninguna manera pudieran traducirse a expresiones en nuestro lenguaje, entonces simplemente hemos de concluir que dichos sonidos no constituyen en realidad conducta de habla. De aquí que la idea de una falla total de traducción sea ininteligible.

En relación a una *falla parcial* de traducción, entendida esta como el caso en que un cierto conjunto de expresiones de un lenguaje pueden traducirse a otro, mientras que otro conjunto de expresiones no pueden traducirse, Davidson nos dice lo siguiente:

*“We can be clear about breakdowns in translation when they are local enough, for a background of generally successful translation provides what is needed to make the failures intelligible.”* (En “On the very idea of a conceptual scheme”; en Davidson, 1984, p. 192)

La tesis de Davidson es que nosotros podemos hacer sentido de la idea de que ciertas expresiones de un lenguaje no pueden traducirse a nuestro lenguaje, sólo cuando la mayor parte de las expresiones de ese lenguaje pueden traducirse al nuestro sin ningún problema. Dicho en otros términos, podemos hacer sentido de la idea de que otras personas asignan a sus emisiones lingüísticas significados muy diferentes de aquéllos que nosotros asignamos a nuestras propias emisiones, y por ende de la idea de que esas personas sustentan ciertas creencias diferentes de las nuestras, solamente en el trasfondo de un acuerdo masivo en lo que respecta a nuestros sistemas de creencias. La razón de que esto tenga que ser así se encuentra en el procedimiento mismo subyacente a toda interpretación, tal como Davidson entiende a esta última. Como se recordará, dicho procedimiento demanda la adhesión al principio de caridad como una condición necesaria para la interpretación. Hemos de recordar también que el principio de caridad nos exhorta a asumir que aquél a quien se interpreta suscribe creencias verdaderas y consistentes -de acuerdo con nuestros propios estándares- en la mayoría de los casos. Si se reflexiona un poco en la tarea de interpretar caritativamente a otros, podremos darnos cuenta que el ser caritativo implica asumir que esos otros a quienes buscamos interpretar son esencialmente como nosotros en lo que respecta a sus sistemas cognoscitivos. Es decir, sus creencias deben ser causadas por el mundo de manera similar a cómo nuestras creencias son causadas por el mundo; de igual manera, las interconexiones entre sus creencias, deseos e intenciones deben adherirse a los mismos principios lógicos que subyacen en las interconexiones entre nuestras creencias, deseos e intenciones.

De lo anterior se sigue que podemos interpretar a otras personas en tanto que asumimos un acuerdo masivo entre sus sistemas de creencias y el nuestro. Es, por ende, en el marco de este acuerdo que nosotros podemos hacer sentido del hecho de que hay algunas diferencias en nuestras creencias. En palabras de Davidson:

*“The point is rather that widespread agreement is the only possible background against which disputes and mistakes can be interpreted. Making sense of the utterances and behavior of others, even their most aberrant behavior, requires us to find a great deal of reason and truth in them.” (En “Belief and the Basis of Meaning”; en Davidson, 1984, p. 153)*

Es también en el marco de una traducción generalmente exitosa que nosotros podemos inteligentemente hablar de “fallas de traducción”. Sin embargo, Davidson apunta, estas fallas no parecen ser tan extremas como para que no podamos dar cuenta de las supuestas “diferencias conceptuales” que las motivaron, utilizando nuestros propios recursos lingüísticos. Como un buen ejemplo de que esto es en verdad así, Davidson nos presenta el



caso de Thomas Kuhn, quien es capaz de decirnos cómo eran las cosas en la ciencia del pasado en términos de un lenguaje “pos-revolucionario”; y el caso de Whorf, quien sostiene que dadas las diferencias entre la metafísica de los Hopi y la de los estadounidenses, el Hopi y el Inglés no pueden “ser calibrados”; Whorf, sin embargo, intenta mostrar esto expresando en inglés algunos contenidos de oraciones Hopi.

Ante casos como estos el cuestionamiento de Davidson es el siguiente: si mediante nuestros propios recursos lingüísticos nosotros podemos comunicar aquéllos conceptos que supuestamente no eran traducibles a nuestro lenguaje, entonces ¿dónde queda la diversidad conceptual? Lo que Davidson concluye, dadas las consideraciones precedentes, es que “todo intento por establecer no-intertraducibilidad (o diversidad conceptual) se autorrefuta”.

Es así como Davidson descarta la idea de “esquema conceptual” y el relativismo conceptual que ésta conlleva. Pero sus argumentos tienen un alcance mucho mayor, pues también permiten descartar cualquier tipo de relativismo cognoscitivo. Es central en Davidson la tesis de que no existe “diversidad cognoscitiva radical”.

La negación de la “diversidad cognoscitiva radical” nos permitirá ahora entender porqué para Davidson la idea de un sujeto que razone de acuerdo con principios normativos de razonamiento radicalmente diferentes de los nuestros es ininteligible. En la visión davidsoniana, si un sujeto es racional, entonces él debe compartir con nosotros un núcleo importante de estándares de racionalidad.<sup>14</sup> De no existir este núcleo de racionalidad compartida, entonces simplemente no habría posibilidad alguna de interpretarlo; por ende, no habría ningún elemento o evidencia que nos permitiera afirmar que ese sujeto es un agente “racional”. El problema aquí es quién podría establecer dicha racionalidad dada la premisa de que no es posible interpretar al sujeto en cuestión. Por lo tanto, la idea de un sujeto que sea racional y al mismo tiempo no podamos interpretar es ininteligible.

## **2. EN CONTRA DE LA UNIVERSALIDAD DE LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO.**

El objetivo en este apartado será presentar dos argumentos filosóficos en contra de la tesis davidsoniana de que hay un conjunto básico de principios normativos de razonamiento que son compartidos por todo agente racional. Estos argumentos abren la posibilidad para pensar que, en principio, pueden existir sujetos racionales que razonen de acuerdo con principios normativos de razonamiento fundamentalmente distintos de los nuestros.

El primero de estos argumentos en contra de la universalidad de los principios normativos de razonamiento puede extraerse desde los argumentos de Christopher Cherniak a favor de la tesis de *racionalidad mínima* sustentados en *Minimal Rationality*.

El segundo de estos argumentos es presentado por Stephen Stich en *The Fragmentation of Reason*.

## 2.1 RACIONALIDAD MÍNIMA.

Empecemos pues presentando la propuesta de Cherniak a favor de lo que él llama *racionalidad mínima*. Como se recordará, Davidson establece una conexión conceptual entre creencia y racionalidad, él nos dice que ser racional es una condición necesaria para tener creencias y por tanto para contar como un agente cognoscitivo. Si ninguna conexión lógica puede ser establecida entre los supuestos “estados intencionales” de un sujeto, entonces simplemente no hay manera de atribuirle creencia alguna, pues una creencia se individua por sus conexiones lógicas con otras creencias (y demás estados intencionales).

Si aceptamos la tesis davidsoniana de que ser racional es una condición necesaria para contar como un agente cognoscitivo, debemos entonces hacernos una pregunta que es fundamental responder a fin de establecer las condiciones que un sujeto debe satisfacer para cumplir con este requisito de racionalidad. Dicha pregunta es la siguiente: *¿qué tan racional debe ser un sujeto para poder calificar como un agente cognoscitivo?*

En *Minimal Rationality* Cherniak propone tres posibles respuestas a esta pregunta:

- 1) Para calificar como un agente cognoscitivo un sujeto no necesita ser racional en lo absoluto (racionalidad nula).
- 2) Para calificar como un agente cognoscitivo un sujeto debe ser idealmente racional (racionalidad ideal)
- 3) Para calificar como un agente cognoscitivo un sujeto debe ser mínimamente racional (racionalidad mínima).

A esta lista podemos agregar un cuarto tipo de respuesta, no explícitamente considerada por Cherniak:

- 4) Para calificar como un agente cognoscitivo, los estados mentales de un sujeto deben “*manifestar un subconjunto especial de inferencias esenciales*”.

Analicemos lo que implica cada una de estas posiciones para la posibilidad misma de interpretar y predecir las acciones de otras personas mediante la atribución de estados intencionales.

### Racionalidad Nula.

Una teoría cognoscitiva que no establezca restricción alguna sobre los tipos de relaciones inferenciales que vinculan entre sí y con el mundo a los estados intencionales de un sujeto, no aportaría ninguna base para considerar a ninguna acción de dicho sujeto como más

probable que otra. Tal teoría carecería en absoluto de poder predictivo. De modo que, en lo que respecta a la racionalidad nula podemos concluir con Cherniak “*No rationality, no agent*”. (Cherniak, 1986, p. 3)

Racionalidad ideal.

Una posición como esta asume que un agente racional debe adherirse *siempre* a los principios normativos de razonamiento que se derivan de la lógica clásica, teoría de la elección racional y teoría de la probabilidad. Un agente que falle, aunque sea mínimamente, en apearse a tales principios, no puede ser considerado como un agente racional. Básicamente, un agente racional ideal debe cumplir con las siguientes condiciones:

- 1) **Condición de consistencia ideal:** por un lado, un agente racional ideal debe eliminar cualquier inconsistencia que surja en su sistema de creencias.
- 2) **Condición de inferencia ideal:** por otro lado, un agente racional ideal debe ser capaz de inferir todas y solamente las creencias implicadas por su sistema original de creencias.

Como bien señala Cherniak, la racionalidad ideal es inaceptable porque impone restricciones que ningún ser humano puede satisfacer. Dado que somos criaturas con recursos cognoscitivos finitos, sólo podemos atender a la vez a un cierto rango de la estimulación que impacta nuestros sentidos, tenemos una capacidad limitada de memoria y las operaciones cognoscitivas que realizamos disponen de un tiempo finito. Por tanto, es humano cometer errores y algunas veces caer en inconsistencias, olvidos y descuidos. Si nosotros exigimos racionalidad ideal como una condición necesaria para la descripción intencional, entonces se sigue la inaceptable consecuencia de que ningún ser humano puede contar como un agente cognoscitivo. Puesto que esto último es absurdo, debemos entonces concluir que la condición de racionalidad ideal no es una condición necesaria para la descripción intencional. Debemos buscar una condición menos idealizada y aplicable a seres con capacidades cognoscitivas finitas como nosotros.

Racionalidad Mínima.

Esta es la postura asumida por Christopher Cherniak. Según él, para que la descripción intencional sea posible y un sujeto pueda contar como un agente cognoscitivo, el conjunto de creencias-deseos de dicho sujeto debe manifestar *algunas*, más no todas, de las inferencias que serían requeridas de un agente racional ideal. Es importante aclarar que cuando se habla aquí de “*algunas*” inferencias, no se hace alusión a algún conjunto específico de inferencias, simplemente se establece que el conjunto de creencias-deseos del sujeto debe manifestar *algunas*, cualesquiera que éstas sean, de las inferencias que serían requeridas

de un agente racional ideal. Básicamente, un agente mínimamente racional debe cumplir con las siguientes condiciones:

- 1) **Condición de consistencia mínima:** en el caso de que surjan inconsistencias en el sistema de creencias de un agente, entonces él debe eliminar algunas de ellas.
- 2) **Condición de inferencia mínima:** un agente mínimamente racional debe realizar algunas -pero no necesariamente todas- de las inferencias aparentemente apropiadas que se siguen desde su sistema de creencias.

Cherniak considera que la condición de racionalidad mínima es la que de hecho satisfacemos los seres humanos cuando atribuimos creencias, pues tanto la condición de racionalidad nula como la condición de racionalidad ideal excluyen toda posibilidad de descripción intencional.

Inferencias esenciales.

Una última vía de respuesta a la cuestión relativa al grado de racionalidad requerida para que la descripción intencional sea posible, sostiene que hay un *subconjunto específico* de creencias y patrones de inferencia que deben presuponerse en cualquiera sujeto que pueda ser considerado un agente cognoscitivo. Es decir, el conjunto de creencias-deseos de un sujeto debe manifestar algunas *inferencias esenciales* a fin de que la descripción intencional sea posible. Por ende, dichas inferencias esenciales deben ser universales para los seres humanos y para cualquiera criatura a la que se pueda interpretar sobre la base de la atribución de creencias, deseos y demás. Así, por ejemplo, según esta postura, toda criatura a la que se pueda describir intencionalmente debe ser capaz, pongamos por caso, de adherirse a la regla *modus ponens* (aunque en realidad nadie ha especificado cuáles serían estas inferencias esenciales).

### 2.1.1 LA POSICIÓN DE DAVIDSON.

Las consideraciones acerca de la aceptabilidad de la postura de *inferencias esenciales* se verán un poco más adelante. Permítaseme antes tratar de ubicar con cual de las posturas anteriores quedaría comprometido Davidson, a fin de poder ir viendo cómo los argumentos de Cherniak pueden dirigirse contra él.

De entrada es claro que la posición de Davidson se contrapone totalmente a la posición de *racionalidad nula*, pues como se ha dicho ya en páginas precedentes, Davidson defiende una conexión necesaria entre racionalidad y creencia, por tanto una conexión necesaria entre ser racional y contar como un agente cognoscitivo. Ahora bien, ¿cómo casa la posición de Davidson con la idea de que para contar como un agente cognoscitivo un sujeto debe ser racionalmente ideal? La respuesta a esta pregunta es un poco menos clara. Por su parte

Cherniak considera que Davidson sostiene una posición acorde con la racionalidad ideal. En *Minimal Rationality*, Cherniak señala que hay un compromiso de Davidson con la condición de consistencia ideal:

*“But on the other hand , the rest of Davidson’s discussion strongly suggests that he thinks the possession of beliefs and desires requires perfect consistency, rather than just “a large degree” of it”. (Cherniak, 1986; p.17).*

en seguida, Cherniak cita el siguiente pasaje de Davidson con el fin de mostrar la presuposición davidsoniana de una condición de consistencia ideal en el ámbito de las preferencias humanas al exigir que el “principio de transitividad” nunca sea violado.

*“I do not think we can clearly say what should convince us that a man at a given time (or without a change of mind) preferred a to b, b to c, and c to a. The reason for our difficulty is that we cannot make good sense of an attribution of preference except against a background of coherent attitudes.” (Davidson; citado en Cherniak, 1986; pp.17-18)*

Por mi parte considero que la posición de Davidson no demanda una racionalidad tan estricta como para ser identificada con la posición de racionalidad ideal. Si bien Davidson demanda un alto grado de coherencia y consistencia en nuestros sistemas de creencias para que la descripción intencional sea posible, considero que él reconoce que cierto grado de error es factible. Por ejemplo, en *Incoherence and Irrationality* Davidson declara que *“An agent can not fail to comport most of the time with the basic norms of rationality, and it is this fact that makes irrationality possible.” (Davidson, 1985; p.352)*

Esta cita nos sugiere que, en la visión de Davidson, lo que se requiere para que la descripción intencional sea posible es que el agente se comporte “la mayor parte del tiempo” –más no *todo* el tiempo- de acuerdo con las normas básicas de racionalidad. Por lo tanto, a mi modo de ver, la postura de Davidson no está comprometida con la racionalidad ideal.

La cita precedente también sugiere fuertemente que, en la perspectiva de Davidson, ser un agente mínimamente racional no sería una condición suficiente para calificar como un agente cognoscitivo, dada la exigencia davidsoniana de que un agente racional debe comportarse “la mayor parte del tiempo” de acuerdo con las normas básicas de racionalidad. Recordemos que la racionalidad mínima demanda, de un agente cognoscitivo, solamente la capacidad de comportarse algunas veces de acuerdo con los principios normativos de razonamiento.<sup>15</sup> Este grado de racionalidad sería demasiado mínimo para Davidson.

Toda vez que hemos encontrado razones para descartar la racionalidad nula, la racionalidad ideal y la racionalidad mínima como posiciones con las que Davidson pudiera estar de acuerdo, el lector puede ya anticipar que la tesis que aquí de sustenta es que es con la posición de inferencias esenciales con la que mejor casa una concepción como la de Davidson. Según la posición de inferencias esenciales, para que un agente pueda ser considerado como racional, el sistema de creencias de dicho agente debe manifestar un conjunto específico de inferencias esenciales. Este requisito para la racionalidad es claramente consistente con la demanda de Davidson de que “hay principios compartidos por todas las criaturas que actúan intencionalmente”, así como con la demanda de que “es una condición necesaria de tener pensamientos y, por ende, de contar como un agente cognoscitivo, el que ciertos estándares básicos de racionalidad tengan aplicación.” Demandas que se expresan con claridad en el siguiente extracto, el cual me permito citar nuevamente:

*“These are principles shared by all creatures that have propositional attitudes or act intentionally; and since I am (I hope) one of those creatures, I can put it in this way: all thinking creatures subscribe to ‘my’ basic standards or norms of rationality. This sounds sweeping, even authoritarian, but it comes to no more than this, that it is a condition of having thoughts, judgments, and intentions that the basic standards of rationality have application.”*  
(Davidson, 1985; p. 351)

### 2.1.2 EN CONTRA DE LA POSICIÓN DE INFERENCIAS ESENCIALES.

Una vez establecido que la respuesta que Davidson daría a la pregunta *¿qué tan racional debe ser un sujeto para poder calificar como un agente cognoscitivo?* casa mejor con la posición de *inferencias esenciales* que con cualquiera de las tres restantes opciones, podemos ahora sí abocarnos totalmente a las objeciones, que pueden extraerse desde los argumentos de Cherniak, en contra de la posición de *inferencias esenciales*.<sup>16</sup> Como veremos, estas objeciones constituyen en sí mismas un argumento importante en contra de la tesis de Davidson de que existe un conjunto básico de principios de razonamiento que son compartidos por todo agente racional. Veamos pues los argumentos de Cherniak.

Reconociendo el hecho de que los seres humanos somos criaturas con capacidades cognoscitivas finitas no puede esperarse que una persona sea capaz de realizar todas las inferencias que se siguen desde su sistema de creencias. Según la propuesta de racionalidad mínima, dicha persona debe ser capaz sólo de realizar correctamente algunas de las inferencias que le serían útiles para actuar apropiadamente en una situación dada<sup>17</sup>.

Asumir que un individuo es un agente mínimo nos confiere cierta capacidad para interpretar y predecir las acciones de ese individuo. Sin embargo, Cherniak señala que nuestra capacidad predictiva puede verse incrementada si nos apoyamos en algunas teorías auxiliares relativas a la psicología cognoscitiva particular del sujeto que estamos interpretando. Específicamente hablando, Cherniak considera que para que nuestras atribuciones de creencia y predicciones sean exitosas, es fundamental apoyarse –entre otras- en una teoría empírica acerca de las inferencias que son factibles para la persona que estamos interpretando, esto es, una teoría acerca de las inferencias que esa persona puede realizar confiablemente. Dicha teoría establecería un ordenamiento de las inferencias factibles para un sujeto de acuerdo con el grado de dificultad de dichas inferencias para ese sujeto particular. De este modo, una teoría de las inferencias factibles nos proveería información acerca de cuáles son las inferencias que el agente debe realizar, a saber, aquéllas que son las más fáciles de acuerdo con el ordenamiento de factibilidad establecido por la teoría; y también nos proveería información sobre cuáles son las inferencias que, dado su grado de dificultad, no pueden ser confiablemente realizadas por el agente. Por ejemplo, inferir  $Q$  a partir de  $P \rightarrow Q$  y  $P$ , puede ser una inferencia mucho más fácil que inferir  $(\forall x)Fx \rightarrow (\forall x)Gx$  a partir de  $(\exists x)(\forall y)(Fx \rightarrow Gy)$ . Por tanto, es de esperarse que el agente mínimamente racional realice apropiadamente y con mayor propensión inferencias del primer tipo.

Ahora bien, un punto fundamental en la discusión de Cherniak en torno a la teoría de las inferencias factibles tiene que ver con la cuestión acerca de la universalidad de dicha teoría. Cherniak se pregunta si hay algunas habilidades lógicas descritas por una teoría de las inferencias factibles que deban ser poseídas por cualquiera criatura que pueda ser considerada un agente mínimamente racional. Como una primera parte de la respuesta a esta pregunta, Cherniak señala que los ordenamientos de factibilidad pueden diferir considerablemente entre los seres humanos, dependiendo por ejemplo del grado de entrenamiento en lógica formal, incluso para un mismo individuo dicho ordenamiento puede variar de un tiempo a otro en la medida en que él puede aprender nuevas estrategias deductivas. Por tanto, es un hecho de la psicología humana que ciertas inferencias pueden ser más difíciles para ciertas personas que para otras. No obstante, cabe todavía la pregunta: ¿es posible que haya un subconjunto de inferencias que sean fáciles o intuitivas para cualquier ser humano normal que no tenga ningún entrenamiento en lógica formal? Antes de responder a esta pregunta, es importante destacar un punto sobre el cual Cherniak llama la atención; él señala que es esencial aclarar el estatus de cuestionamientos como el precedente, es decir, debe quedar claro qué elementos de la respuesta a la pregunta

planteada pueden ser aportados por investigación empírica, y qué elementos pueden ser aportados por algún argumento trascendental *a priori*. En otras palabras, Cherniak nos conmina a evitar confusiones entre aquello que puede ser establecido como un hecho contingente de la psicología empírica y aquello que puede ser establecido *a priori* como universal y necesario. Lo que centralmente interesa a Cherniak es dilucidar si hay un conjunto de inferencias esenciales que deban ser obvias e intuitivas para todo agente racional, y en que medida esto puede ser establecido como una verdad necesaria, más que como una cuestión de generalización empírica.

Para Cherniak, es muy claro que la condición de racionalidad mínima como un requisito para contar como un agente cognoscitivo es una condición que se establece como una verdad *a priori*. Sin embargo, la pregunta acerca de si hay algún conjunto de inferencias que sean intuitivas para todos los seres humanos debe ser respondida fundamentalmente mediante investigación empírica. De cualquier manera, independientemente de cuáles sean los resultados de esta investigación, debemos evitar tratar a estos hechos de la psicología empírica sobre el razonamiento humano como verdades necesarias universalmente aplicables a todo agente racional posible. Lo que Cherniak concluye es que, más allá de la investigación empírica, no hay razones conceptuales que anulen la posibilidad de un agente mínimamente racional cuyas capacidades inferenciales difieran sustancialmente de las nuestras. Veamos cuáles son sus argumentos en relación a este punto.

Cherniak considera que, entre los seres humanos sin entrenamiento alguno en lógica formal, los ordenamientos de factibilidad difieren en detalles mínimos. No obstante, es posible pensar en criaturas con ordenamientos de factibilidad que difieran significativamente de los ordenamientos de factibilidad de seres humanos típicos. Considérese, por ejemplo, el caso de una criatura para la cual el ordenamiento de inferencias factibles –desde las más fáciles hasta las más difíciles- fuera invertido con respecto a nuestro propio ordenamiento. Es decir, las inferencias más fáciles para dicha criatura serían las más difíciles para nosotros y a la inversa. Así las cosas, para una criatura como esta, sería muy difícil inferir  $Q$  a partir de  $P \rightarrow Q$  y  $P$ , pero inferir  $(\forall x)Fx \rightarrow (\forall x)Gx$  a partir de  $(\exists x)(\forall y)(Fx \rightarrow Gy)$  sería una tarea sencilla. La pregunta pertinente a continuación es si una criatura como esta podría contar como un agente mínimamente racional, pregunta a la que Cherniak responde afirmativamente. Más adelante se aclarará porqué.

A fin de mostrar la plausibilidad lógica de un agente mínimamente racional con un ordenamiento de factibilidad invertido, Cherniak debe rebatir dos argumentos que pretenden mostrar que la posibilidad de una criatura con un ordenamiento de factibilidad invertido queda conceptualmente excluida. Estos argumentos son: 1) el argumento axiomático y 2) el argumento de inferencias constitutivas. Veamos cada uno por vez.



1) El argumento axiomático.

Este argumento demanda que una criatura que no es capaz de realizar una inferencia tan simple como, pongamos por caso, *modus ponens*, no puede ser capaz de realizar inferencias mucho más complejas. Los que sustentan este argumento presuponen que la capacidad de un razonador para realizar tareas complejas se explica por la capacidad de ese sujeto para realizar una secuencia de tareas simples. De alguna manera se asume que la capacidad deductiva de un agente debe conformarse a la forma en que los sistemas deductivos formales han sido generados: a partir de ciertos axiomas y reglas sencillas se prueban teoremas simples; luego, a partir de dichos teoremas simples (en adición con las reglas y los axiomas) es posible probar teoremas mucho más complejos. A la luz de esta concepción, el problema con una criatura con un ordenamiento de factibilidad invertido es que su capacidad para realizar tareas complejas, mientras que es incapaz de realizar tareas simples, parece misteriosa.

Para responder al argumento axiomático Cherniak plantea el caso hipotético, aunque lógicamente posible, de una criatura con un ordenamiento de factibilidad invertido la cual puede, no obstante, conformarse a la metodología axiomática arriba señalada. La criatura imaginaria que Cherniak nos propone considerar puede hacer inferencias desde sus creencias usando un sistema deductivo de axiomas y reglas de inferencia, sólo que dichos axiomas y reglas son sumamente complejos y no obvios para nosotros, por ejemplo, uno de estos axiomas podría ser algo como  $((\exists x)Fx \rightarrow p) \leftrightarrow (\forall x)(Fx \rightarrow p)$ . Otra peculiaridad de dicha criatura es que ésta no puede usar teoremas previamente probados dada su incapacidad de recordarlos, incapacidad que se explica por las características específicas de los procesos de memoria de esta criatura, la cual asumimos es muy distinta de nosotros en lo que a procesos psicológicos se refiere. De este modo, las inferencias realizadas a partir de los axiomas y las reglas de inferencia de esta criatura serían demasiado difíciles para nosotros aunque muy sencillas para ella. De igual modo, las inferencias que son triviales para nosotros podrían ser realizadas por esta criatura pero se requerirían tantos pasos que la realización de la inferencia llegaría a ser una tarea muy complicada. Como vemos, la competencia lógica de la criatura que estamos considerando se aproximaría a un ordenamiento de factibilidad invertido, sin embargo, declara Cherniak, esto no impide que la criatura realice tareas difíciles (aunque fáciles para nosotros) por medio de la realización de tareas más fáciles (aunque difíciles para nosotros). Así las cosas, el argumento axiomático no elimina la posibilidad de una criatura con un ordenamiento de factibilidad invertido.

En una segunda respuesta al argumento axiomático, Cherniak, propone considerar otra criatura hipotética la cual utiliza un sistema deductivo peculiar. Dicha criatura utiliza una lista finita de teoremas de la lógica ordenados de modo tal que ellos pueden ser buscados de una

manera muy eficiente. Además, dicha lista contiene solamente teoremas que rebasan un cierto nivel de complejidad –aquellos que para seres humanos típicos son prácticamente infactibles. Cuando la criatura se pregunta si cierta oración  $p$  es una consecuencia lógica de ciertas premisas, lo único que tiene que hacer es buscar el condicional asociado a la inferencia en cuestión en su lista de teoremas. Si el condicional se encuentra en la lista entonces la criatura inferirá  $p$  desde las premisas, en caso contrario la criatura no realizará la inferencia. Cherniak afirma que una criatura con estas características tendría la habilidad lógica necesaria para resolver muchas de las tareas de razonamiento de que es capaz un ser humano inteligente y normal, no obstante que su ordenamiento de factibilidad se aproximaría a un ordenamiento de factibilidad invertido, y no obstante no adherirse a la metodología axiomática presupuesta en el argumento axiomático, pues una criatura como ésta no sigue en absoluto el procedimiento de probar teoremas complejos mediante la utilización de teoremas más simples. La inteligibilidad de una criatura como la recién descrita permite, nuevamente, mostrar que el argumento axiomático no elimina la posibilidad de una criatura con un ordenamiento de factibilidad invertido.

Una conclusión importante que Cherniak quiere sustentar con ejemplos como los anteriores es que la competencia lógica de una agente mínimamente racional tiene la estructura de una colección o agregado (*cluster structure*). Para dejar claro que quiere decir Cherniak con esto permítaseme citarlo:

*“The logical competence required of a minimally rational agent seems rather to have a cluster structure: it cannot be the case that a minimal agent is able to make no inferences, but the agent can be unable to make any particular one.” (Cherniak, 1986; p. 28)*

## 2) El argumento de inferencias constitutivas.

Este argumento declara que un individuo que, o bien niegue, o bien suspenda la creencia en una regla lógica obvia, como por ejemplo  $\neg(p \wedge \neg p)$ , simplemente no puede entender el sentido de los conectivos lógicos que aparecen en la regla. Así, por ejemplo, si un individuo es incapaz de inferir  $p \vee q$  desde  $p$  (adición), o de inferir  $q$  desde  $p \vee q$  y  $\neg p$  (silogismo disyuntivo), entonces dicho individuo no entiende el sentido de la “disyunción”; no puede por tanto atribuírsele creencias que involucren este conectivo, pues ser capaz de realizar inferencias que involucren reglas como la adición y el silogismo disyuntivo es constitutivo de entender el sentido de “ $\vee$ ”. Una razón para demandar lo anterior es que, si un individuo fuera incapaz de realizar estas inferencias básicas relativas a la disyunción, entonces no habría elementos suficientes para determinar que lo que ese individuo significa por “o” es lo mismo

que lo que nosotros significamos por “o” y diferente de lo que nosotros significamos con algún otro conectivo. Por ejemplo, imaginemos una criatura que utiliza el conectivo ‘#’ el cual tratamos de traducir. Observamos que esta criatura no realiza las inferencias constitutivas para la disyunción, es decir, ella no infiere  $p\#q$  desde  $p$  (adición), ni infiere  $q$  desde  $p\#q$  y  $\neg p$  (silogismo disyuntivo). También observamos que siempre que esta criatura asiente a  $p\#q$  asiente también a  $q\#p$ , pero no así a  $r$ . Sobre la base de estas observaciones nosotros podemos hipotetizar que ‘#’ puede traducirse como “ $\vee$ ” en tanto que la inferencia de  $q\vee p$  desde  $p\vee q$  es correcta y en tanto que la inferencia de  $r$  desde  $p\vee q$  no lo es. El problema aquí es que estas observaciones no constituyen una base suficiente para confirmar nuestra hipótesis de que ‘#’ puede ser traducido como “ $\vee$ ”, pues las habilidades lógicas observadas son también consistentes con la hipótesis de que ‘#’ puede ser traducido como “ $\wedge$ ”, toda vez que la inferencia de  $q\wedge p$  desde  $p\wedge q$  también es correcta y la inferencia de  $r$  desde  $p\wedge q$  también es incorrecta. De este modo, será muy difícil establecer concluyentemente que lo que esa criatura significa con ‘#’ es lo mismo que lo que nosotros significamos con “ $\vee$ ”, a menos que esa criatura sea capaz de realizar inferencias básicas como la adición y el silogismo disyuntivo.

En su respuesta al argumento precedente, Cherniak destaca el hecho de que los seres humanos normales generalmente aceptan reglas lógicas obvias y niegan o suspenden el juicio ante reglas lógicas muy complicadas. Sin embargo, él nos recuerda que puede haber agentes cognoscitivos cuyas capacidades inferenciales sean diferentes de las nuestras, como es el caso de criaturas con un ordenamiento de factibilidad invertido. Dichas criaturas serían incapaces de realizar inferencias que implicaran el uso de reglas lógicas que para nosotros son obvias, no obstante podrían realizar inferencias ante las cuales, dada su complejidad, nosotros simplemente tendríamos que suspender nuestras creencias. Tras este planteamiento Cherniak se pregunta ¿por qué no la aceptación -por parte de una criatura con un ordenamiento de factibilidad invertido- de reglas lógicas complejas implica un entendimiento de los conectivos lógicos implicados, del mismo modo en que ocurre en el caso de nuestra aceptación de reglas lógicas obvias? ¿por qué hemos de asumir que las reglas lógicas que son obvias para nosotros tienen un estatus especial?

Para Cherniak simplemente no hay una base objetiva para conferir a las inferencias obvias para nosotros un estatus privilegiado, dado que la dificultad de una inferencia no es algo que pueda ser determinado por alguna característica sintáctica intrínseca a la inferencia, al margen de la psicología particular del razonador.

Además, Cherniak afirma que no es cierto que en el caso de una criatura que no pueda hacer las dos inferencias constitutivas señaladas para la disyunción no se tengan elementos

suficientes para determinar que el sentido que esa criatura da a ‘#’ es el mismo sentido que nosotros damos a “ $\vee$ ”. Siempre podrán hacerse más observaciones que nos muestren otras capacidades inferenciales de la criatura y a partir de las cuales podrá determinarse que ‘#’ quiere decir “ $\vee$ ” y no “ $\wedge$ ”. Aquí Cherniak plantea la posibilidad de una observación como la siguiente: siempre que la criatura asiente a  $\neg(\neg p \# q)$  ésta asiente a  $p\#q$  (se presupone que previamente ya ha sido establecido que la criatura significa con “ $\neg$ ” lo mismo que nosotros). Esta evidencia es suficiente para determinar que ‘#’ debe ser traducido como “ $\vee$ ” y no como “ $\wedge$ ”, pues la inferencia de  $p\vee q$  desde  $\neg(\neg p \vee q)$  es correcta, en tanto que la inferencia de  $p\wedge q$  desde  $\neg(\neg p \wedge q)$  es incorrecta.

Las consideraciones de Cherniak muestran que la idea de un agente que sea mínimamente racional y que no sea capaz de realizar las llamadas inferencias constitutivas es inteligible. Esta conclusión nuevamente sugiere que la competencia lógica de un agente mínimamente racional tiene una estructura de agregado. En palabras de Cherniak:

*“We have found that, although a deducer cannot be incapable of making all of the inferences governing a given constant, it can be incapable of making any particular inference involving the constant (...) without changing the identity, and our translation, of the constant.” (Cherniak, 1986; p. 42)*

Podemos finalizar esta sección estableciendo algunas conclusiones importantes que se derivan de los argumentos de Cherniak: en primer lugar, una criatura con un ordenamiento de factibilidad invertido, y por ende, con capacidades inferenciales muy diferentes de las nuestras, es lógicamente posible. En segundo lugar, de la demanda de que la competencia lógica de un agente mínimamente racional tiene una estructura de agregado, se sigue que siempre es posible que haya alguna inferencia particular que sea infactible para un sujeto, no obstante que esta pueda ser factible para otros. Por último, de todo lo anterior se sigue que no hay inferencias esenciales, no hay un conjunto básico de reglas de inferencia a las que deba adherirse todo sujeto que pueda ser considerado un agente cognoscitivo. Quedan así asentados los argumentos de Cherniak y cómo estos pueden ser dirigidos contra la tesis davidsoniana de que hay un conjunto básico de principios normativos de razonamiento que son compartidos por todo agente racional.

## 2.2 EL ARGUMENTO DE STEPHEN STICH.

El presente argumento es construido por Stich en su intento de atacar la tesis de que existen límites conceptuales al mal razonamiento. Esta tesis, que Stich atribuye principalmente a Davidson y a Dennett, sostiene que dada la conexión conceptual entre racionalidad y

atribución de creencias, es incoherente suponer que los procesos inferenciales de un agente cognoscitivo podrían desviarse sin límite de los estándares normativos de racionalidad. La estrategia de Stich para socavar esta tesis será mostrar que, si bien es cierto que tales límites conceptuales al mal razonamiento pueden ser establecidos, dichos límites no son interesantes en tanto que ellos son relativos al observador y dependientes de contexto. Como veremos más adelante, el argumento de Stich no sólo mostrará que el mal razonamiento es posible, sino además, y eso es lo interesante aquí, este argumento permitirá abrir la puerta a la posibilidad de diversidad cognoscitiva radical. Es decir, a la posibilidad de que haya agentes cognoscitivos que razonen de acuerdo con principios de razonamiento fundamentalmente distintos de los nuestros.

Partiendo de la aceptación de la existencia de una conexión conceptual entre racionalidad y atribución de creencias, Stich busca explorar un poco más y dilucidar porqué esta conexión conceptual se sostiene. Esta búsqueda lleva a Stich a concluir que el “principio de humanidad” propuesto por Richard Grandy puede dar una explicación de la interdependencia entre racionalidad y creencia. Dicho principio sostiene que cuando se busca interpretar a alguien y nos vemos en la encrucijada de decidir cuál es la traducción (o la interpretación) correcta siempre debe elegirse aquélla en la cual *“el patrón imputado de relaciones entre creencias, deseos, y el mundo sea tan similar al nuestro como sea posible”* (Grandy, 1973; p. 443)

En seguida veremos cómo el principio de humanidad puede dar cuenta de la conexión entre racionalidad y creencia. Sólo permítaseme antes dejar establecidos algunos puntos pertinentes en relación a las bases sobre las que es posible sustentar este principio de humanidad.

La enunciación del principio de humanidad seguramente nos recuerda la demanda asentada en el *principio de caridad*: debe elegirse aquélla interpretación según la cual la mayoría de las aserciones del hablante sean verdaderas y la mayoría de sus inferencias sean racionales, desde nuestros propios estándares de verdad y racionalidad. No obstante, debe notarse que el principio de humanidad es un principio mucho más general que el principio de caridad, por lo que este último queda subsumido en el primero. Por esta razón el dictamen último sobre cual es la traducción correcta que puede ser emitido apegándose al principio de caridad puede coincidir con el dictamen resultante de la adhesión al principio de humanidad. Aunque también es posible que haya casos en los que los dictámenes resultantes de la aplicación de ambos principios no coincidan, tal como lo muestra Grandy en el siguiente pasaje:

*“Suppose Paul has arrived at a party and asserts “The man with the martini is a philosopher.” And suppose the facts are that there is a man in plain view who is drinking water from a martini glass and that he is not a philosopher. Suppose also that in fact there is only one man at the party drinking a martini, that he is a philosopher, and that he is out of sight in the garden. Under the circumstances the charitable thing to do would be to take Paul’s remark at face value (homophonically), since that is simple and makes his remark true. But the natural thing to do is to understand him as having said something false, or at least to view the situation as one in which his utterance shows he has a false belief. That is what is predicted by the principle of humanity....Since no reason could be given as to why Paul would have a belief about the philosopher in the garden, it is better to attribute to him an explicable falsehood than a mysterious truth....*

*The principle of humanity directs us to bear in mind that the speaker is a person and has certain basic similarities to ourselves when we are choosing between translations.” (Grandy, 1973; p. 445)*

Con el ejemplo anterior Grandy busca dar sustento a la idea de que es más bien el principio de humanidad –y no precisamente el principio de caridad- el que debe gobernar el proceso de traducción. Ahora bien, en la visión de Grandy, las razones por las que un intérprete debe adherirse al principio de humanidad son de carácter pragmático. Grandy considera que uno de los propósitos de la traducción es la predicción y la explicación de la conducta del interpretado. Para la consecución de esta meta, el intérprete se apoya en las creencias y deseos imputados. El problema aquí es que esta información no es suficiente para una predicción exitosa de la conducta, pues nuestro conocimiento de las creencias y deseos del hablante es de cualquier modo limitada para el fin que se persigue. Por lo tanto, nos dice Grandy, el traductor/intérprete debe apoyarse en un modelo del agente. Un modelo que puede obtenerse usándonos a nosotros mismos. Lo que esto quiere decir es que cuando nosotros buscamos predecir la conducta de una persona, siempre tenemos el recurso de considerar lo que nosotros haríamos si estuviéramos en la situación de esa persona y tuviéramos las creencias y deseos que se presupone que ella tiene. En otras palabras nosotros podemos intentar predecir la conducta de alguien a través de una “simulación”, mediante nuestro propio aparato cognoscitivo, de ese alguien. Que la simulación sea o no exitosa, y por ende la predicción, es algo que depende de la similaridad del sistema de creencias-deseos de la persona en cuestión con nuestro propio sistema de creencias-deseos. De aquí la importancia de adherirse al principio de humanidad y escoger aquellas traducciones en las que “el patrón imputado de relaciones entre creencias, deseos y el mundo sea tan similar al nuestro como sea posible”, de otro modo la traducción no serviría al propósito de la predicción de la conducta del otro.

Vemos pues que para Grandy la adherencia al principio de humanidad es forzada sobre nosotros sólo porqué dado nuestro conocimiento limitado de las creencias y deseos de otra persona nos vemos forzados a utilizar una estrategia de “simulación” a la hora de intentar predecir sus acciones, de modo que de no existir tal limitación en nuestro conocimiento podríamos prescindir del principio de humanidad. Es así que –en la visión de Grandy- las razones por las que debemos adherirnos al principio de humanidad son de carácter pragmático. En este último punto, Stephen Stich está en total desacuerdo con Grandy. Stich sostiene que aún sin estas limitaciones de conocimiento sobre las creencias y deseos de otros, la restricción impuesta por el principio de humanidad sigue vigente. Él argumenta que, de cualquier manera, se requiere de similaridad entre los sistemas de creencias de intérprete e interpretado para que la interpretación sea posible:

*“If a subject’s mental state does not interact with other mental states in a pattern which approximates the pattern exhibited by our own conditional beliefs, it does not count as a conditional belief. Similarly, if a subject mental state is not caused by stimuli similar to the ones which lead me to believe that there is an elephant in front of me, and if it does not interact with other mental states in a way similar to the way mine would, then it does not count as a belief that there is an elephant in front of him.” (Stich, 1990; p.47)*

La razón de esta conexión entre similaridad y descripción intencional se encuentra en la manera en qué de hecho los seres humanos procedemos cuando atribuimos estados intencionales. A este respecto Stich coincide plenamente con Grandy en que lo que nosotros hacemos cuando atribuimos creencias, deseos y demás a otras personas es asumirnos a nosotros mismos en sus roles y situaciones. Esta asunción de rol o “simulación” del otro nos permite determinar lo que nosotros creeríamos y haríamos en esa situación. El paso siguiente es imputar esas mismas creencias a la persona que interpretamos y esperar que se comporte de las maneras previstas, desde nuestra propia percepción de su situación. No obstante, para Stich, el seguimiento de esta estrategia no es forzado sobre nosotros dado el limitado conocimiento que tenemos sobre los estados cognoscitivos de otras personas, como asume Grandy, sino que más bien, esta es la estrategia que de hecho seguimos los seres humanos cuando interpretamos a otros.

En tanto que esta última tesis requiere tener un fundamento empírico, la aceptaremos aquí con cierta reserva, pues la investigación empírica sobre la manera en que de hecho las personas atribuyen estados intencionales a otras personas es actualmente un campo de investigación en incipiente desarrollo donde no faltan las cuestiones polémicas y las teorías propuestas son ampliamente debatibles.<sup>18</sup>

Una vez establecida la razón por la cual, según Stich, estamos forzados a adherirnos al principio de humanidad, procedamos ahora a ubicar cómo es que dicho principio puede explicar la conexión entre racionalidad y descripción intencional. Puesto que la posibilidad de descripción intencional presupone similitud con nosotros mismos, y puesto que nosotros razonamos suficientemente bien; entonces aquéllos a quienes nosotros podemos atribuir estados intencionales deben igualmente razonar suficientemente bien. En conclusión, la conexión entre racionalidad y descripción intencional deriva de la conexión entre similitud y descripción intencional.

Como hemos visto, para Stich la similitud es una condición necesaria de la descripción intencional. Podemos ahora formular una pregunta análoga a la que ya nos habíamos planteado en relación al grado de racionalidad requerido para que la descripción intencional fuera posible: ¿Qué tanta similitud se requiere para que la descripción intencional sea posible? Si la respuesta a esta pregunta fuera que el grado de similitud requerido para poder atribuir estados intencionales a otro individuo es de similitud total o completa, entonces tendríamos que aceptar la absurda consecuencia de que no podríamos interpretar absolutamente a nadie y nadie podría contar como un agente cognoscitivo, pues no existe otro individuo que comparta con nosotros absolutamente todas y las mismas creencias con justamente los mismos patrones de relaciones inferenciales. Por lo tanto, el grado de similitud requerido para que la descripción intencional sea posible no puede ser de similitud total o completa. La opción que nos queda es asumir que se requiere similitud en ciertos aspectos cognoscitivos entre nosotros y aquellos que intentamos interpretar, para que la atribución de estados intencionales sea posible. Esta última es justamente la posición que Stich asume. Ahora bien, ¿en qué aspectos cognoscitivos y en qué grado debe haber similitud? Esto es algo que, según propone Stich, sólo puede ser establecido por el contexto.

Lo interesante de la posición de Stich es que abre la posibilidad de pensar inteligentemente en una sucesión de personas cada una de las cuales es muy similar a la persona inmediatamente anterior en el sentido de que comparten justamente las mismas creencias y el mismo patrón de relaciones entre ellas, excepto en el caso de una creencia particular. Es decir, la segunda persona cree todo lo que cree la primera persona salvo en el caso de una creencia específica la cual sólo es creída por la segunda persona pero no por la primera. Relaciones análogas se sostienen entre la segunda persona y la tercera, y para las personas restantes en esta sucesión. En una situación como esta las personas no tendrían problemas en atribuir estados intencionales a aquéllos que les son próximos en la sucesión, sin embargo, la primera y la última persona no podrían interpretarse, dado que llegarían a ser cognoscitivamente muy disímiles. Los mismos resultados se producirían si el aspecto mínimo



en que estas personas difieren fuera el tipo de relación inferencial que una creencia particular mantiene con otra.

Si las cosas son como Stich propone, entonces ¿qué pasa con los límites al mal razonamiento presupuestos por Davidson? Dado el panorama recién bosquejado puede esperarse que en la medida en que las creencias y patrones de inferencia de un sujeto sean más y más diferentes de los nuestros, en esa misma medida nuestra posibilidad de atribuirle creencias, deseos y demás se vaya perdiendo. De ahí que un sujeto muy diferente de nosotros pueda parecer poco o nulamente racional y por lo tanto no lo consideremos un agente cognoscitivo. Sin embargo, Stich nos hace notar que esta línea de demarcación entre estados intencionalmente caracterizables y aquellos que no lo son es totalmente vaga y relativa al observador, dado que depende totalmente del grado de similitud entre intérprete e interpretado. Por lo mismo, los límites que pueden ser puestos al mal razonamiento son vagos y relativos al observador. Consecuentemente, estos límites no son teóricamente interesantes y no tienen por tanto porque preocuparnos.

Como podemos darnos cuenta, Stich, en comparación con Davidson, establece una condición menos estricta para que la descripción intencional sea posible. Él sólo demanda similitud en algunos aspectos cognoscitivos, en contraste con la condición mucho más fuerte que propone Davidson, a saber, que haya un conjunto básico de principios de razonamiento compartido por todos aquellos que puedan ser considerados agentes cognoscitivos. En este contexto, la crítica de Stich contra Davidson sería la siguiente: el análisis de los requerimientos de la descripción intencional no nos compromete a otra cosa que a la "similitud" en algunos aspectos cognoscitivos como una condición necesaria para la atribución de creencias. La demanda mucho más estricta de Davidson de que debe haber un conjunto básico de principios de razonamiento compartidos por todo agente al que se le pueda describir intencionalmente se sigue de la teoría davidsoniana del significado y de toda su visión de la interpretación. Sin embargo no tenemos porque aceptar la semántica de Davidson. El análisis de la descripción intencional, al margen de la teoría davidsoniana de la interpretación, dicta solamente que la similitud es una condición necesaria de la descripción intencional y no más. Si el argumento de Stich es correcto, entonces podemos concluir que desde el análisis de los requerimientos de la descripción intencional no se desprende ningún tipo de restricción conceptual que anule la posibilidad de diversidad cognoscitiva radical.

### 3. CONCLUSIÓN

En el presente capítulo hemos visto la teoría de la interpretación de Donald Davidson, así como su concepción de la racionalidad. El objetivo ha sido mostrar cómo de éstas se desprende la tesis de que debe haber un conjunto básico de principios normativos de razonamiento compartido por todas las criaturas racionales y, de ahí, la tesis de que los principios normativos de razonamiento tienen que ser universales.

Con el interés de rebatir la tesis anterior, hemos visto, además, dos concepciones alternativas en lo que respecta a los requisitos de la descripción intencional. Por un lado, vimos que, según Christopher Cherniak, para que la descripción intencional de un agente sea posible, éste debe ser mínimamente racional. Ahora bien, del análisis que hace Cherniak acerca de las capacidades cognoscitivas de un agente mínimamente racional, se desprende una importante consecuencia para la tesis de que los principios normativos de razonamiento son universales, pues la plausibilidad de agentes con ordenamientos de factibilidad invertidos, abre la posibilidad para pensar en sujetos racionales que razonen de acuerdo con principios de razonamiento diferentes de los nuestros y que suscriban, por tanto, principios normativos de razonamiento diferentes de los nuestros.

Por otro lado, vimos que, según Stephen Stich, una condición necesaria para que la descripción intencional sea posible es la *similaridad* entre intérprete e interpretado. Sin embargo, esta condición no implica que todos los agentes racionales posibles tengan que ser estrictamente similares entre sí. Esta condición, sólo exige cierto grado de similaridad en ciertos aspectos cognoscitivos. Puesto que dicho grado de similaridad puede ser variable, en el sentido de que puede haber sujetos muy similares a nosotros, del mismo modo que puede haber sujetos no tan similares a nosotros; podemos perfectamente imaginar el caso de un agente al que no podamos interpretar dadas sus diferencias con nosotros, no obstante que dicho agente sí pueda ser interpretado por un individuo al cual nosotros sí podamos interpretar y el cual pueda a su vez interpretarnos. Así las cosas, la diversidad cognoscitiva parece perfectamente plausible. Por lo tanto, la posibilidad de sujetos que razonen de acuerdo con principios de razonamiento diferentes de los nuestros queda abierta.

## NOTAS.

---

<sup>13</sup> Para una presentación esclarecedora de los argumentos, y en general de la concepción de la racionalidad sustentada por Donald Davidson, véase: Eraña Lagos, 2003.

<sup>14</sup> En otras palabras, debe haber un acuerdo masivo entre sus principios normativos de razonamiento y los nuestros. Sólo en el marco de este acuerdo será posible establecer algunas diferencias, si es que las hubiera, entre nuestros respectivos estándares de racionalidad.

<sup>15</sup> Cabe señalar que Cherniak también establece un requisito negativo para la racionalidad mínima. Dicho requisito sostiene que un agente mínimamente racional debe “no solamente intentar algunas de las acciones que son apropiadas dado su conjunto de creencias-deseos, él debe además, no intentar suficientes de las acciones que son inapropiadas dado su conjunto de creencias-deseos”. (Cherniak, 1986, pp. 9-10). Un requisito análogo se sostendría en el caso de la condición que requiere de un agente mínimamente racional, la capacidad de razonar algunas veces de acuerdo con los principios normativos de razonamiento. Dicho requisito negativo establecería que un agente mínimamente racional no debe razonar suficientes veces de maneras que van en contra de los principios normativos de razonamiento.

<sup>16</sup> Como ya se ha señalado, Cherniak no hace alusión explícita a la posición de *inferencias esenciales* como otra posibilidad de respuesta a la pregunta sobre qué tan racional debe ser un sujeto para contar como un agente cognoscitivo, sin embargo, de las consideraciones que él hace en aras de precisar cual sería la competencia lógica de un agente mínimamente racional, se desprenden importantes consecuencias para la posición de *inferencias esenciales*.

<sup>17</sup> Para Cherniak, “una acción es aparentemente apropiada si y sólo si, de acuerdo a las creencias de A, ésta tendería a satisfacer los deseos de A”. (Cherniak, 1986; p. 7)

<sup>18</sup> Ver Apéndice B para una descripción breve de las dos principales teorías psicológicas de la interpretación.

## **CAPITULO CUATRO.**

### **LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO EN EL MARCO DE UNA EPISTEMOLOGÍA NATURALIZADA.**

En tanto que el presente trabajo se ubica plenamente en el marco de una epistemología naturalizada, se dedica el primer apartado de este capítulo a la presentación de los argumentos esgrimidos por Quine a favor de una epistemología que pueda echar mano de los datos aportados por la ciencia empírica y en contra de la epistemología que tradicionalmente se había venido desarrollando, epistemología para la cual los fundamentos de todo principio normativo sólo pueden ser establecidos *a priori*.

Posteriormente, en conformidad con un enfoque naturalizado de la epistemología, argumentaré, sobre la base de argumentos tanto conceptuales como empíricos, a favor de dos tesis principales:

1) la primera es que, los principios normativos de razonamiento no son necesariamente principios universales, esto es, principios aplicables a la evaluación del razonamiento de todo sujeto en todo tiempo y lugar; 2) la segunda es que, los principios normativos de razonamiento no son necesariamente principios formales, pues al menos algunos de ellos son principios dependientes de contenido o específicos a un dominio.

#### **1. LA JUSTIFICACIÓN DE LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO.**

Como ya se ha señalado en el capítulo precedente, uno de los proyectos centrales de la epistemología ha sido la evaluación de nuestros métodos de investigación y de los procesos cognoscitivos mediante los cuales llegamos a consolidar nuestros sistemas de creencias. El objetivo de esta evaluación es determinar si nuestras creencias son epistémicamente aceptables, es decir, si hay realmente buenas razones para sustentarlas y estamos, por ende, justificados en creer lo que creemos. Si esta es la tarea de la epistemología es entonces esencial para el epistemólogo dar respuesta a preguntas como ¿qué constituye una *buen razón* para creer en algo? o ¿qué condiciones debe satisfacer una creencia para ser epistémicamente aceptable? pues las respuestas dadas a estas cuestiones constituirán la base sobre la cual el epistemólogo podrá luego establecer los principios normativos que han de guiar sus evaluaciones. Estos principios normativos dictaminarán cómo debemos proceder en asuntos de cognición. Pero las cuestiones anteriores no son las únicas

cuestiones relevantes a las que el epistemólogo debe responder en aras de realizar su trabajo evaluativo. Otras cuestiones pertinentes que demandan una respuesta son: ¿por qué hemos de aceptar la autoridad del epistemólogo y adherirnos a los principios normativos que él establece? ¿cómo sabemos que esos son los principios normativos correctos? Preguntas como estas obedecen a una preocupación por la justificación misma de los principios normativos o, en otros términos, por los fundamentos de la normatividad epistémica.

En el presente apartado veremos las dos respuestas que a lo largo del desarrollo de la epistemología, han sido dadas a la pregunta por los fundamentos de la normatividad epistémica. Veremos también cómo el compromiso con una u otra de estas respuestas ha dado lugar al surgimiento de dos concepciones muy distintas de la epistemología como disciplina filosófica.

### 1.1 EPISTEMOLOGIA TRADICIONAL Y EPISTEMOLOGIA NATURALIZADA.

¿De dónde proviene la justificación de las normas epistémicas? La respuesta que tradicionalmente ha dado la epistemología a esta cuestión es que las normas epistémicas se justifican *a priori*, es decir, con independencia de la experiencia sensorial. Dicho en otros términos, determinar qué cuenta como epistémicamente aceptable involucra un trabajo de reflexión al margen de cualquier indagación empírica. Esto es así en tanto que las normas epistémicas se conciben como verdades analíticas y toda verdad analítica tiene una justificación *a priori*. Una proposición *analítica* es aquella cuya verdad puede ser establecida exclusivamente -sin importar cuales sean los hechos- sobre la base de los significados de los símbolos de que se compone. Es así que las proposiciones analíticas no proveen información alguna acerca de cómo es el mundo, ellas carecen enteramente de contenido factual. Por tal razón ninguna experiencia puede invalidarlas. Esto último les concede un carácter de universalidad y necesidad. Es decir, las proposiciones analíticas son verdaderas en cualquier mundo posible y son necesarias en el sentido de que su negación resulta contradictoria.

Ilustremos estos puntos con un ejemplo típico de proposición analítica: “Todo hombre soltero es no casado”. Esta proposición es analítica en tanto que su verdad se establece en virtud únicamente de los significados de las palabras *soltero* y *hombre no casado*. Además, esta proposición es verdadera independientemente de cómo sea el mundo y su negación se nos presenta como contradictoria.

La razón por la que la negación de “Todo hombre soltero es no casado” -y en general la negación de cualquier proposición analítica- resulta contradictoria radica en el hecho de que dicha negación contraviene nuestras convenciones y usos lingüísticos. Al señalar este punto

quiero enfatizar dos supuestos básicos de la epistemología tradicional: primero, que si bien las proposiciones analíticas no tienen contenido empírico, ellas pueden darnos conocimiento en tanto que ellas registran las maneras en que hemos decidido usar las palabras; ellas también pueden contribuir a develar ciertas implicaciones de nuestras creencias de las que de otro modo no llegaríamos a ser conscientes; segundo, que el carácter necesario de las proposiciones analíticas tiene como fundamento las reglas tácitas del lenguaje que subyacen a las maneras en que usamos las palabras.

Dada la concepción previa sobre la analiticidad, la demanda de la epistemología tradicional de que las normas epistémicas son verdades analíticas tiene las siguientes implicaciones: 1) las normas epistémicas no son susceptibles de ser revisadas a la luz de la experiencia, 2) la validez de las normas epistémicas se desprende de su vinculación lógica con verdades universales y necesarias y, 3) su validez se deriva, además, del análisis de nuestras intuiciones –intuiciones que se presupone son universalmente compartidas- respecto a las significaciones y usos de nuestros conceptos de aceptabilidad epistémica, tales como los de “justificación” y “racionalidad”. Por último, 4) otra implicación importante que se desprende de la concepción de las normas epistémicas como verdades analíticas es una cierta concepción de la epistemología según la cual ésta, como disciplina, tiene un carácter autónomo respecto de las ciencias empíricas. Desde la perspectiva de la epistemología tradicional, es este carácter autónomo lo que hace posible que la epistemología pueda determinar los criterios de validación de la ciencia empírica desde fuera de la ciencia misma, evitando así caer en algún tipo de razonamiento o justificación circular.

En contraste con la visión de la epistemología tradicional, una segunda respuesta a la pregunta por los fundamentos de la normatividad epistémica, es que la aceptabilidad de las normas epistémicas no puede ser establecida con independencia de la experiencia sensorial, en consecuencia ellas pueden ser revisadas a la luz de consideraciones empíricas. Esta visión tiene su origen en las tesis sustentadas por W. V. O. Quine en *Dos Dogmas del Empirismo* (1953). En este texto, Quine hace un cuestionamiento serio a la distinción analítico-sintético, esto es, a la distinción entre proposiciones que son verdaderas en virtud de los significados de las palabras de que se componen, y proposiciones que son verdaderas no sólo en virtud del significado de sus términos sino además en virtud de cómo es el mundo. El cuestionamiento de esta distinción tuvo consecuencias drásticas para la manera en que tradicionalmente se había venido concibiendo el quehacer epistemológico. En primer lugar, reconocer el hecho de que las normas epistémicas pueden ser revisables en función de cuestiones de naturaleza empírica, implica reconocer que la epistemología ya no

puede tener ese carácter autónomo que se presuponía que ella tenía respecto de las ciencias fácticas. La epistemología no puede ser concebida ya como una disciplina independiente, muy por el contrario, la epistemología llega ahora a ser concebida como parte misma de la ciencia. Consecuentemente, el estudio empírico de la cognición humana llega – entre otros estudios empíricos- a considerarse relevante en la determinación y justificación de las normas epistémicas, de modo que la separación tajante entre epistemología y psicología desaparece. Es así que la epistemología se naturaliza.

Una segunda consecuencia importante de esta concepción naturalizada de la epistemología es que las normas epistémicas pierden su carácter de universalidad y necesidad, pues si ellas pueden ser evaluadas empíricamente, entonces se abre la posibilidad de que las normas epistémicas que alguna vez resultaron aceptables dejen de serlo en la medida en que nuestras visiones acerca de cómo es el mundo se modifican, o en la medida en que nuestras visiones del mundo pueden diferir y posiblemente conflictuar con las de otras culturas. De este modo, las normas epistémicas no pueden tener aplicación para todo sujeto, en todo tiempo y lugar. Vemos pues que esta posibilidad de revisar las normas epistémicas a la luz de la experiencia, abre espacio a la posibilidad de diversidad normativa.

Por supuesto, varias objeciones han surgido en contra de una epistemología naturalizada, objeciones que serán abordadas en su momento a lo largo del presente capítulo. Por ahora me abocaré a presentar, aunque sea de una manera somera, los argumentos en los que Quine se apoya para demandar no sólo que la distinción analítico-sintético no se sostiene, sino además –como veremos- que la epistemología debe echar mano de la psicología si es que pretende lograr su propósito de dilucidar qué vínculos de justificación hay entre teoría y evidencia.

Dos tesis centrales defendidas por Quine son la base de toda su argumentación posterior. En primer lugar, Quine se compromete con una *teoría verificacionista del significado*. Según esta teoría, el significado de un enunciado está dado por el método de confirmación o refutación empírica de ese enunciado. De aquí que un enunciado sea significativo si tiene ciertas consecuencias observacionales susceptibles de ser contrastadas en la experiencia. En segundo lugar, Quine se compromete con una *teoría holista de la verificación*, teoría originalmente propuesta por Pierre Duhem. Según ésta, los enunciados singulares que conforman una determinada teoría científica no tienen por ellos mismos un conjunto de consecuencias observacionales que les sean propias. Sólo de teorías completas pueden desprenderse consecuencias observacionales susceptibles de ser contrastadas. Por lo tanto, la unidad de verificación son las teorías, no enunciados aislados. De esto último y de la

teoría verificacionista del significado se sigue que la unidad de significación son las teorías, no enunciados aislados. De modo que si tiene algún sentido hablar del significado de enunciados singulares, es siempre teniendo en cuenta que ese significado depende de las relaciones que el enunciado en cuestión mantiene con los demás enunciados que conforman la teoría así como con los enunciados que conforman la base evidencial de dicha teoría. De aquí que Quine rechace la idea de que el significado oracional sea algo fijo que pueda prevalecer cuando hay cambios teóricos importantes. En otras palabras: el significado oracional está indeterminado.

En la visión de Quine, una teoría científica puede ser considerada como una especie de campo de fuerza en cuyo centro se encuentran enunciados no-observacionales los cuales suelen ser más resistentes a la revisión (tales como leyes teóricas y las leyes de la lógica), mientras que en la periferia se encuentran enunciados observacionales, los cuales son más susceptibles de ser revisados en caso de darse un conflicto con la experiencia. De cualquier manera, cualquier reajuste en la periferia del campo puede provocar reajustes al interior del campo dadas las interconexiones lógicas que los enunciados mantienen entre sí. De aquí la demanda de Quine de que no hay enunciados que sean inmunes a la revisión. Además, ningún enunciado es inmune a la revisión porqué, cuando se dan conflictos con la experiencia, también es posible intentar preservar los enunciados en la periferia mientras se hacen reajustes al interior del campo. No obstante que este tipo de reajustes puedan implicar cambios más drásticos en el sistema, llegan realmente a darse cuando hay cambios sustanciales en nuestros sistemas de conceptos.

Aquí es importante destacar que las revisiones de nuestros sistemas de creencias pueden darse en cualquier punto de dicho sistema, ya sea central o periférico. La razón por la que esto es así es que, cuando hay desacuerdos entre la teoría y la evidencia nosotros podemos saber que muy probablemente algo anda mal en la teoría, pero dada la teoría holista de la verificación (y del significado), nosotros no podemos saber en qué enunciados se encuentra justamente el problema. Esto hace que haya un amplio rango de elección respecto a qué enunciados deben ser eliminados o qué enunciados deben ser preservados. Lo anterior nos lleva a darnos cuenta de que no hay ninguna conexión unívoca entre un enunciado singular (aislado de una teoría) y una experiencia en particular, y de que hay por tanto más de una manera de conectar la teoría con la evidencia. Esto último queda claramente especificado en una tercera tesis sustentada por Quine. Me refiero aquí a la *tesis de la subdeterminación de la teoría por la evidencia*. Específicamente, esta tesis sostiene que, dado un cierto cuerpo de evidencia siempre será posible encontrar dos o más teorías –las cuales pueden incluso ser incompatibles entre sí- que se adecuen igualmente bien a dicha evidencia.



Frente a este panorama delineado por Quine podemos ahora ubicar porqué la distinción analítico-sintético no se sostiene. Por una parte, se ha establecido ya que para Quine no existen enunciados que no puedan ser revisados a la luz de la experiencia sensorial. Por lo tanto no existen enunciados justificables a priori. Si no hay enunciados que puedan justificarse con total independencia de la experiencia; entonces, partiendo del hecho de que todo enunciado analítico se justifica a priori y de que todo enunciado que se justifica a priori es analítico, podemos también concluir que no hay enunciados analíticos.<sup>19</sup>

Por otra parte, en el análisis que hace Quine de la noción de analiticidad él señala que dicha noción presupone un cierto concepto de “*significación*”, por lo tanto, si se quiere esclarecer la noción de analiticidad será importante desarrollar una teoría de la significación. Quine señala además que “*el objeto primario*” de una teoría de la significación es “*la sinonimia de las formas lingüísticas*” (Quine, 1953; p. 52). Esclarecer la noción de *sinonimia* llega entonces a ser esencial para el esclarecimiento de la noción de analiticidad, dado que nosotros podríamos llegar a definir una oración analítica como aquella que puede ser construida desde una verdad lógica<sup>20</sup> simplemente sustituyendo sinónimos por sinónimos, tal como lo muestra el mismo Quine (Quine, 1953; pp. 52-53) con el siguiente ejemplo:

- 1) Ningún hombre no casado es casado
- 2) Ningún hombre soltero es casado

la oración 1 es una verdad lógica (también es una oración analítica) que puede dar lugar a otra oración (como la oración 2) -que aunque no es una verdad lógica sí es una verdad analítica- simplemente sustituyendo ‘no casado’ por su sinónimo ‘soltero’.

Sin embargo, el problema con esta concepción de la analiticidad radica en que presupone que *es posible hablar de la significación de palabras o de oraciones aisladas, lo cual no es posible si se acepta un holismo del significado*. En realidad, si se acepta un holismo del significado, entonces no tiene sentido decir que dos oraciones aisladas son sinónimas, por tanto tampoco tiene sentido decir que hay oraciones que son analíticas, esto es, verdaderas en virtud de sus significaciones.

En *Dos Dogmas del Empirismo*, Quine no sólo cuestiona la distinción analítico-sintético, sino que también pone en tela de juicio el proyecto reductivista que empiristas lógicos como Carnap habían venido desarrollando. El presupuesto básico del reductivismo es que cualquier enunciado no observacional que sea significativo debe poder traducirse a algún enunciado de nivel observacional, esto es, que haga alusión a la experiencia inmediata.

Quine sostiene que este proyecto reductivista está destinado al fracaso. Sus razones para sostener esto se derivan de su compromiso con las tesis arriba mencionadas, a saber, su compromiso con una teoría verificacionista del significado, con una teoría holista de la

verificación y con la tesis de la subdeterminación de la teoría por la evidencia. A grosso modo, Quine considera que la reducción de enunciados no observacionales a enunciados de experiencia inmediata no es posible porque, como ya se ha mencionado, existe más de una manera de conectar la teoría con la evidencia. Quine considera por tanto que no es posible derivar estrictamente la teoría desde la evidencia, por lo que no puede haber una justificación lógica que vincule necesariamente a un cierto cuerpo de evidencia con determinada teoría científica.

El rechazo de Quine a la distinción analítico-sintético y al reduccionismo le llevaron posteriormente a sostener, en *Naturalización de la Epistemología* (Quine, 1969), que la epistemología debía apelar a la psicología si es que en realidad buscaba dilucidar los fundamentos de la ciencia. En lugar de pretender establecer *a priori* cuáles son los vínculos de justificación entre teoría y evidencia, el epistemólogo debía buscar establecer estos vínculos a partir de un análisis de las maneras cómo de hecho los científicos llegan a construir sus teorías a partir de la experiencia sensorial. En la medida en que este proceso de construcción es un proceso cognoscitivo cuyo estudio compete a la psicología, la propuesta de Quine implica un desdibujamiento de las fronteras entre epistemología y psicología. Consecuentemente, una característica esencial de una epistemología naturalizada es la afirmación de que las maneras cómo de hecho los seres humanos generan y revisan sus sistemas de creencias son relevantes al establecimiento de las maneras como deben hacerlo.

## 1.2 LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO ESTÁN “INDEXADOS” A LA COMPETENCIA HUMANA DE RAZONAMIENTO.

Para la epistemología tradicional determinar las formas en que los seres humanos de hecho razonamos y determinar las condiciones que deben satisfacerse para conferir justificación y validez a dichos razonamientos, son dos proyectos diferentes y totalmente independientes. El primero es un proyecto propio de la psicología, el segundo es un proyecto esencialmente epistemológico. Sin embargo, como ya vimos, para una epistemología naturalizada estos dos proyectos no son independientes. Muy por el contrario, para una epistemología naturalizada, el establecimiento y justificación de los principios normativos de razonamiento no puede proceder al margen de consideraciones empíricas concernientes a las maneras cómo los seres humanos de hecho razonamos.

El presente trabajo se inscribe plenamente en el marco de una concepción naturalizada de la epistemología. De ahí que una premisa fundamental para las tesis que más adelante se intenta sustentar –a saber: el carácter no necesariamente universal ni necesariamente formal de los principios normativos de razonamiento- sea que *no hay principios normativos de*

*razonamiento que puedan ser establecidos puramente a priori, esto es, no hay principios normativos de razonamiento que puedan justificarse con total independencia de la investigación empírica sobre el razonamiento humano.*

Siendo pues congruentes con una postura epistemológica que no ignora cuestiones de naturaleza empírica, considero que debemos reconocer que no tiene sentido proponer estándares de racionalidad que, por su complejidad, rebasen las capacidades de razonamiento de los seres humanos. Pues simplemente no parece correcto pedir que las personas razonen en formas en que -dadas las limitaciones de sus recursos cognoscitivos- no pueden hacerlo. Si hemos aceptado que “debe” implica “puede”, aceptaremos el absurdo que hay en exigir a una persona la realización de una acción que simplemente ella no puede realizar. Por ejemplo, sería absurdo exigir que los seres humanos razonemos de maneras que preserven la consistencia de nuestros sistemas de creencias, si por consistencia se entiende consistencia ideal. Cumplir con esta condición de consistencia ideal implicaría eliminar absolutamente todas las inconsistencias de nuestros sistemas de creencias, requisito, éste último, que ningún ser humano podría satisfacer. La razón básicamente es que el conjunto de nuestras creencias, más el conjunto de todas las posibles implicaciones que se siguen de ellas, constituye un conjunto infinito, por lo cual, la tarea de checar que cada una de nuestras creencias, y sus posibles implicaciones, sea compatible con todas las demás sería una tarea infinita y requeriría de un tiempo infinito. Si aceptamos con Cherniak que los seres humanos somos seres con capacidades cognoscitivas finitas (predicamento finitario), hemos de reconocer entonces el absurdo que hay en exigir la adhesión a principios normativos de razonamiento que son infactibles para criaturas como nosotros. Absurdo en el que caen ciertas posturas filosóficas que asumen concepciones de racionalidad sumamente idealizadas, tal es el caso de la visión estándar de la racionalidad. Hemos de recordar que, según la visión estándar de la racionalidad, ser racional es razonar de acuerdo con los principios postulados por la lógica clásica, teoría de la probabilidad y teoría de la elección racional. A la luz del predicamento finitario la visión estándar de la racionalidad se nos presenta como una concepción demasiado estricta de la racionalidad en tanto que no puede ser alcanzada por criaturas como nosotros. Que esto es así es algo que también puede ser establecido por los resultados de los experimentos de razonamiento, los cuales, como ya hemos visto, parecen demostrar que el razonamiento efectivo de los seres humanos se desvía sistemáticamente de muchos de los principios normativos de razonamiento que la visión estándar de la racionalidad postula. Así las cosas, si los humanos no podemos razonar sistemáticamente de acuerdo con muchos de los principios de la lógica clásica, teoría de la probabilidad, etc. en tanto que ellos son infactibles para nosotros, podemos concluir

entonces que la visión estándar de la racionalidad debe ser rechazada y reemplazada por una concepción de la racionalidad que postule principios normativos de razonamiento que sean factibles para criaturas con recursos cognoscitivos limitados como nosotros. Pero ¿cuál puede ser esta concepción alternativa? Algunas respuestas posibles han sido ya propuestas por filósofos como Stein (1996) y Stich (1990) entre otros. Por mi parte, yo no intentaré proponer aquí ninguna respuesta posible a esta pregunta. Baste sólo con dejar en claro que cualesquiera que sean los principios normativos de razonamiento correctos, éstos no son independientes de las habilidades reales de razonamiento exhibidas por los seres humanos. En este sentido, sostengo que los principios normativos de razonamiento están *indexados* a la competencia de razonamiento humana.

### 1.3 UNA OBJECCIÓN A LA CONCEPCIÓN NATURALIZADA DE LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO.

Recién se ha dicho que los principios normativos de razonamiento están “indexados” a la competencia humana de razonamiento. Esto quiere decir que no es posible determinar cuáles son los principios normativos de razonamiento correctos sin el apoyo de una teoría auxiliar relativa a las maneras como los seres humanos de hecho razonamos. Esta aseveración ha levantado una importante objeción a la concepción naturalizada de los principios normativos de razonamiento. Objeción que analizaremos y a la que se buscará dar respuesta en el presente apartado.

La objeción es básicamente que una concepción naturalizada de los principios normativos de razonamiento elimina la posibilidad de evaluar las capacidades de razonamiento de las personas en tanto que abandona el aspecto normativo de la epistemología y reduce la empresa epistemológica a una mera descripción. En términos más específicos la objeción reza como sigue: si es absurdo pretender que los seres humanos razonen en formas en que simplemente no pueden hacerlo -dado el predicamento finitario-, entonces los principios normativos que sean infactibles para nosotros no pueden ser los principios normativos correctos. Dado lo anterior, los principios normativos correctos deben ser principios factibles para nosotros; por ende dichos principios deben estar incorporados en nuestra competencia de razonamiento. Por lo tanto, nosotros debemos conocer cuál es nuestra competencia de razonamiento para así poder determinar, a partir de este conocimiento, cuáles son los principios normativos de razonamiento correctos. El problema con este proceder es que cae en el error de pretender derivar el “deber ser” del “ser” y reduce este “deber ser” a una mera descripción de las habilidades de razonamiento efectivas de las personas. Este modo de proceder es característico de aquel investigador del razonamiento humano que, ante resultados experimentales que muestran una desviación sistemática de los sujetos de un

determinado principio normativo de razonamiento previamente concebido, decide –antes que atribuir irracionalidad a sus sujetos- rechazar el principio normativo en cuestión y postular el principio subyacente a la ejecución de los sujetos como el principio normativo correcto. El investigador adopta así la llamada *estrategia del rechazo de la norma* (Stein, 1996, p. 233), la cual consiste en que, siempre que se presenta una divergencia entre la competencia de razonamiento humana y un supuesto principio normativo de razonamiento, el investigador debe rechazar la norma y postular el principio de razonamiento que los humanos de hecho están usando como el principio normativo correcto.

Es fácil ver porqué este modo de proceder elimina la posibilidad de evaluar el razonamiento: si se asume que cualesquiera que sean los principios de razonamiento incorporados en nuestra competencia de razonamiento, estos principios son los principios normativos correctos, entonces no hay nadie que pueda razonar de maneras distintas a las prescritas por los principios normativos de razonamiento (en el caso en que esto llegara a ocurrir siempre será posible apelar a ciertos errores de ejecución). Si toda persona normal razona de acuerdo con la competencia humana de razonamiento, toda persona normal razona de acuerdo con los principios normativos correctos. Se elimina así la posibilidad de mal razonamiento, todo razonamiento es razonamiento correcto. ¿Qué sentido tiene entonces la pretensión epistemológica de evaluar nuestros modos de generar y revisar creencias?

La objeción arriba presentada es una objeción importante e implica un cuestionamiento serio a la concepción naturalizada de los principios normativos de razonamiento, pues si el trabajo del epistemólogo implica determinar cuáles son los principios normativos de razonamiento correctos partiendo exclusivamente de la base de una descripción de las maneras como los seres humanos de hecho razonamos, entonces sería efectivamente cierto que el trabajo del epistemólogo quedaría reducido a una tarea descriptiva. Y puesto que este quehacer descriptivo compete propiamente a la psicología, la epistemología, cuyo interés es esencialmente normativo, sería eliminada como disciplina y enteramente reemplazada por la psicología. Así las cosas, la preocupación de aquéllos que ven en la naturalización de la epistemología una amenaza para la epistemología misma es perfectamente entendible. Sin embargo, como veremos en seguida, no hay en verdad razón de preocupación, pues es posible asumir una concepción naturalizada de la epistemología sin abandonar el carácter normativo propio y esencial de ésta.

La tesis de que una epistemología naturalizada elimina toda posibilidad de evaluación del razonamiento se sustenta esencialmente en un presupuesto erróneo, a saber, en el presupuesto de que una concepción naturalizada de los principios normativos de razonamiento queda forzosamente comprometida a asumir la estrategia del rechazo de la

norma. Efectivamente, el compromiso con dicha estrategia eliminaría, como ya hemos visto, la posibilidad de evaluar el razonamiento. No obstante, una concepción naturalizada de los principios normativos de razonamiento no está necesariamente vinculada a la estrategia del rechazo de la norma. Veamos porqué.

Reconociendo lo impropio que resulta demandar que los seres humanos razonen de maneras en que, dadas sus capacidades cognitivas, simplemente no pueden hacerlo; aquí se ha aceptado plenamente la demanda de que en materia de razonamiento *debe implicar puede*. Lo cual quiere decir que, para que un determinado principio X pueda contar como un principio normativo para un determinado sujeto S, se requiere, como una condición necesaria, que S pueda razonar de acuerdo con X. Si un sujeto S no puede en el presente razonar de acuerdo con el principio de razonamiento X, ni puede tampoco, mediante proceso de entrenamiento alguno, aprender a razonar de acuerdo con X, entonces X no puede ser un principio normativo de razonamiento para S. Podemos concluir entonces que el hecho de que S pueda razonar de acuerdo con el principio de razonamiento X es una condición necesaria para que X pueda fungir como un principio normativo de razonamiento para S. Sin embargo, en ningún momento se ha dicho que el hecho de que S pueda razonar de acuerdo con el principio de razonamiento X es una condición suficiente para que X se constituya en un principio normativo de razonamiento para S. S puede sin duda razonar de maneras muy particulares e incluso extrañas, pero ese hecho no basta para aceptar esas maneras singulares de razonar como principios normativos correctos, pues *puede* no implica *debe*. Vemos pues que, en modo alguno, la concepción naturalizada de los principios normativos de razonamiento está asumiendo la estrategia del rechazo de la norma. Por lo tanto la posibilidad de evaluación del razonamiento humano no queda eliminada.

Hasta aquí he intentado dejar en claro cual es la concepción epistemológica con que aquí me comprometo, así como el hecho de que el compromiso con una concepción naturalizada de la epistemología no representa ninguna amenaza para el quehacer epistemológico. En lo que resta de este capítulo me abocaré a la presentación de dos consecuencias importantes para nuestra concepción de la naturaleza de los principios normativos de razonamiento. Consecuencias que se desprenden no sólo del análisis filosófico, sino también del análisis de los resultados de la investigación en psicología del razonamiento. Dichas consecuencias son: 1) el cuestionamiento de la pretendida universalidad de los principios normativos de razonamiento y, 2) el cuestionamiento al carácter estrictamente formal de los mismos. Veamos pues.

## **2. LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO NO SON NECESARIAMENTE UNIVERSALES.**

Tradicionalmente la epistemología ha tendido a concebir a los principios normativos de razonamiento como principios universales. Esto es, como principios que prescriben las maneras correctas de razonar para toda persona normal en todo tiempo y lugar. Por ende, para la epistemología tradicional hay un único conjunto de principios normativos correctos y no puede haber sujetos racionales que suscriban principios normativos sustancialmente diferentes de éstos. Un objetivo central del presente trabajo es minar esta concepción tradicional y proveer argumentos a favor de la idea de que puede haber principios normativos de razonamiento sustancialmente distintos de los que nosotros consideramos los principios normativos de razonamiento correctos, aunque igualmente aceptables.

En tanto que Donald Davidson es uno de los principales filósofos en el grupo de aquéllos que sustentan esta concepción universalista de los principios normativos de razonamiento, el ataque aquí será dirigido específicamente en contra de la concepción davidsoniana de la racionalidad.

De manera muy general, un primer punto que puede ser señalado en contra de la concepción davidsoniana, lo constituye la presuposición de que los principios normativos de razonamiento correctos pueden ser establecidos puramente *a priori*. Como ya se ha visto, desde la perspectiva de una epistemología naturalizada, esta presuposición es errónea.

Vayamos ahora a argumentos más específicos en contra de la concepción universalista de los principios normativos de razonamiento sustentada por Davidson. Permítaseme dividir estos argumentos en dos clases: 1) argumentos conceptuales y, 2) argumentos empíricos en contra de la universalidad de los principios normativos de razonamiento.

### **2.1 ARGUMENTOS CONCEPTUALES EN CONTRA DE LA UNIVERSALIDAD DE LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO.**

En este apartado me limitaré a recapitular, aunque de una manera más breve, los argumentos presentados en el capítulo precedente. Como el lector recordará estos argumentos se los debemos básicamente a Christopher Cherniak (1986) y a Stephen Stich (1990). Empecemos con los argumentos de Cherniak.

Cómo se recordará Davidson nos ha señalado que existe una conexión entre racionalidad y creencia en el sentido de que un sujeto cuyos estados intencionales no puedan conectarse entre sí de maneras lógicamente apropiadas no puede ser considerado, en última instancia, como teniendo creencias. Pues una creencia se identifica como una creencia específica

dadas las relaciones lógicas que mantiene con otros estados intencionales. Ahora bien, en relación a la cuestión sobre qué tan racional deber ser un sujeto para que la atribución de creencias sea posible, la posición de Davidson es que el sistema de creencias de dicho sujeto debe ser masivamente racional y manifestar un conjunto básico de reglas esenciales de inferencia. En contraste, la posición de Cherniak ante esta cuestión es que para que la atribución de creencias sea posible y un sujeto pueda contar como un agente cognoscitivo, éste debe ser mínimamente racional, esto es, debe ser capaz de realizar solamente algunas de las inferencias que serían apropiadas para un agente racional ideal.

Una posición como la de Davidson, comprometida con la visión de que existe un núcleo básico de inferencias esenciales compartido por todo agente racional, queda irremisiblemente comprometida con una posición universalista en lo que respecta a la aplicabilidad de los principios normativos de razonamiento. En contraste, la posición de Cherniak abre espacio a la posibilidad de agentes racionales que razonen de acuerdo con principios de razonamiento sustancialmente distintos de los nuestros. Posibilidad que queda asentada en los ejemplos sobre criaturas hipotéticas que Cherniak presenta con el fin de mostrar que la idea de una criatura con un ordenamiento de factibilidad invertido, es perfectamente inteligible. Recuérdese que “ordenamiento de factibilidad” refiere a un ordenamiento de las inferencias que un determinado individuo puede realizar confiablemente, según su grado de dificultad para ese individuo, desde las más fáciles hasta las más difíciles. Cherniak argumenta que, aún frente a una teoría psicológica sobre las inferencias que son factibles para seres humanos típicos, siempre es posible pensar en la existencia de agentes mínimamente racionales cuyos ordenamientos de factibilidad sean invertidos con respecto al nuestro, de modo que las inferencias que son fáciles para nosotros sean sumamente difíciles para ellos, y a la inversa. En un caso extremo, es incluso plausible la idea de un agente con un ordenamiento de factibilidad completamente invertido, en el sentido de que las inferencias que son factibles para él son totalmente infactibles para nosotros, y a la inversa. Para sustentar la plausibilidad de agentes racionales con ordenamientos de factibilidad invertidos con respecto al nuestro, Cherniak debe rebatir dos importantes argumentos conceptuales que intentan mostrar que debe haber un conjunto de habilidades lógicas poseídas por cualquier criatura que pueda ser considerada un agente mínimamente racional.

El primero de estos argumentos, conocido como el argumento axiomático, señala que la realización de una inferencia compleja implica siempre la realización de una secuencia de inferencias más simples; por lo tanto, cualquier agente mínimo debe ser capaz de realizar esas inferencias simples. De no ser así, resultaría altamente problemático explicar la capacidad de una criatura para la realización de inferencias sumamente complejas, dado el



caso de que dicha criatura fuera incapaz de realizar inferencias que son simples y obvias para nosotros. Cherniak responde a este argumento proponiéndonos imaginar una criatura la cual realiza inferencias sobre la base de un sistema deductivo singular. En lugar de poseer un conjunto de axiomas y reglas de inferencia a partir de los cuales se van deduciendo teoremas cada vez más complejos, dicha criatura posee una lista finita de teoremas los cuales se encuentran ordenados de manera tal que su búsqueda resulta altamente eficiente. Hay que señalar que esta lista esta constituida por teoremas que resultan sumamente difíciles e incluso infactibles para seres como nosotros. Así las cosas, cuando dicha criatura intenta realizar una inferencia, por ejemplo, determinar si cierta oración  $p$  se sigue de ciertas premisas, lo único que ella tiene que hacer es buscar el condicional asociado a dicha inferencia en su lista de teoremas. En caso de que dicho condicional se encuentre efectivamente en su lista, entonces la criatura realizará la inferencia en cuestión. En caso contrario la criatura no realizará la inferencia.

Cherniak sostiene que una criatura como esta, no obstante no realizar tareas complejas sobre la base de la realización de una secuencia de tareas simples –tal como exige el argumento axiomático-, sería capaz de enfrentarse satisfactoriamente a muchas de las tareas de razonamiento que puede realizar un ser humano normal. Por tanto, dicha criatura, cuyo ordenamiento de factibilidad se aproximaría a un ordenamiento de factibilidad invertido, podría ser considerada como un agente mínimamente racional. De este modo, el argumento axiomático no elimina la posibilidad de una criatura con habilidades lógicas muy distintas de las nuestras.

El segundo argumento que intenta mostrar que debe haber un conjunto de habilidades lógicas poseídas por cualquier criatura que pueda ser considerada un agente mínimamente racional es el argumento de “inferencias constitutivas”. Dicho argumento sostiene que la capacidad para realizar ciertas inferencias que son obvias es constitutivo de entender el significado de los conectivos lógicos implicados en esas inferencias. Así, por ejemplo, si un sujeto es incapaz de realizar ciertas inferencias que se consideran obvias en el caso de la disyunción, tales como: inferir  $p \vee q$  desde  $p$  (adición), o inferir  $q$  desde  $p \vee q$  y  $\neg p$  (silogismo disyuntivo), entonces ese sujeto no entiende el significado del conectivo “o”. Según quienes sustentan este argumento esto tiene que ser así, pues si una criatura, a la que nosotros buscamos interpretar, fuera incapaz de realizar estas inferencias constitutivas, entonces nosotros no tendríamos elementos suficientes para poder establecer que lo que esa criatura significa con “o” es lo mismo que lo que nosotros significamos con “o” y diferente de lo que nosotros significamos con algún otro conectivo.

La respuesta de Cherniak a este argumento nuevamente consiste en apelar a una criatura hipotética con un ordenamiento de factibilidad invertido la cual es incapaz de realizar las

llamadas “inferencias constitutivas”. Sin embargo, esta criatura sería capaz de realizar inferencias muy complejas las cuales serían infactibles para nosotros; y aquí Cherniak se pregunta ¿por qué hemos de asumir que el uso de reglas lógicas complejas (para nosotros), por parte de esta criatura, no implica un entendimiento del sentido de los conectivos involucrados, y por qué hemos de asumir que nuestro uso de reglas lógicas obvias (para nosotros) si implica un entendimiento del sentido de los conectivos lógicos involucrados? Parece que la única respuesta que el defensor del argumento de inferencias constitutivas puede dar a esta pregunta, es que las reglas lógicas que son obvias para nosotros deben tener un estatus especial. Pero, ¿sobre qué base objetiva podemos aseverar esto? En realidad, nos dice Cherniak, no hay ninguna base objetiva que nos permita conferir a las inferencias que son obvias para nosotros un estatus privilegiado, dado que no hay una característica sintáctica intrínseca a una inferencia que nos permita determinar su grado de dificultad al margen de consideraciones sobre la psicología de los individuos particulares que han de realizar dicha inferencia. Por otra parte, Cherniak señala que no es cierto que en el caso de una criatura que no pueda realizar las llamadas inferencias constitutivas, no tendríamos elementos suficientes para determinar que lo que esa criatura significa con, pongamos por caso “o”, es lo mismo que lo que nosotros significamos con “o”. Pues si bien es cierto que la criatura es incapaz de realizar inferencias que son básicas para nosotros, también es cierto que ella será capaz de realizar un sin número de inferencias que nos darán la base suficiente para determinar el sentido en que esta criatura está utilizando sus símbolos de constantes lógicas. Por lo tanto, la idea de un agente que sea incapaz de realizar las inferencias constitutivas y que sea mínimamente racional es perfectamente inteligible.

Hemos visto hasta aquí como los argumentos de Cherniak abren espacio a la posibilidad de diversidad cognoscitiva radical. Veamos ahora como, algunos argumentos presentados por Stich, pueden encaminarnos en esa misma dirección.

Reconociendo el hecho de que hay una conexión conceptual entre racionalidad y creencia, los argumentos de Stich se originan en su intento de indagar un poco más y tratar de dilucidar cuál es la razón de esta conexión conceptual. Los esfuerzos de Stich le llevan finalmente a proponer que la clave de la interdependencia entre racionalidad y descripción intencional se encuentra en el *principio de humanidad* propuesto por Richard Grandy. Según dicho principio, el criterio que debemos seguir a fin de determinar cuál es la traducción o interpretación correcta cuando buscamos atribuir estados intencionales y predecir la conducta de otros, es el de elegir siempre aquella traducción en la cual “*el patrón imputado*

*de relaciones entre creencias, deseos, y el mundo sea tan similar al nuestro como sea posible” (Grandy, 1973; p. 443)*

Para Stich, es este principio, más que el principio de caridad, el que debe guiar nuestras interpretaciones. Pues más que el presuponer que las creencias de aquellos a quienes interpretamos son verdaderas e interactúan de acuerdo a patrones lógicos de relación, la condición verdaderamente necesaria para que la descripción intencional sea posible es la similaridad entre nosotros y aquéllos a quienes interpretamos. Un elemento a favor de esta tesis puede encontrarse en el ejemplo propuesto por Grandy del hombre con el martini. Recordemos rápidamente este ejemplo: “Paul llega a una fiesta y dice “el hombre con el martini es un filósofo”. No obstante, el hombre que está a la vista de Paul sólo está bebiendo agua en un vaso de martini y no es filósofo. Por otra parte hay en la fiesta sólo un hombre que está bebiendo un martini y que es filósofo, pero este hombre se encuentra en el jardín, fuera de la vista de Paul”. Este ejemplo pretende ilustrar un caso en el que, la adhesión al principio de caridad no nos dará una interpretación correcta de la aserción de Paul de que “el hombre con el martini es un filósofo”, mientras que la adhesión al principio de humanidad, por el contrario, sí lo hará. La razón es la siguiente: si nos apegamos al principio de caridad tendríamos que asumir que la creencia de Paul es verdadera y que él se está refiriendo al hombre que se encuentra en el jardín. Sin embargo, nuestra intuición nos señala que esta no puede ser la interpretación correcta, pues ¿cómo podría Paul estarse refiriendo a alguien que está fuera de su vista y del que no sabe siquiera que está en el jardín? Por el contrario, si nos apegamos al principio de humanidad tendremos que asumir que Paul está teniendo una creencia falsa en relación al hombre que se encuentra al alcance de su vista. Esta creencia atribuida a Paul a la luz del principio de humanidad, aunque falsa, resulta explicable y factible en tanto que nosotros podemos asumírnos a nosotros mismos en la situación de Paul y sabemos que en esas mismas circunstancias podríamos llegar a creer lo mismo que Paul.

Otro elemento, destacado por Stich, a favor de la tesis de que es la similaridad más que la racionalidad, la condición necesaria para la descripción intencional, es el hecho de que en realidad lo que parece que nosotros hacemos cuando interpretamos a otros, es imaginarnos a nosotros mismos en el lugar de ese otro y considerar lo que nosotros creeríamos, o desearíamos, o haríamos justamente en esa situación. En tanto que ésta última es una tesis empírica, le corresponde a la psicología determinar su verdad. Pero, independientemente de que ésta sea verdadera o falsa, e independientemente de que la teoría rival, a saber la teoría-teoría (ver apéndice B), sea verdadera o falsa; es importante destacar aquí que *ninguna de estas teorías se compromete con la tesis de que debe haber un conjunto de principios de razonamiento compartidos por todos los sujetos para que la descripción*

*intencional sea posible, además ninguna de estas teorías podría comprometerse con algo como esto a priori.*

Una vez que Stich establece que la similaridad es una condición necesaria para la descripción intencional, él procede a responder a la interrogante que dio origen a sus planteamientos, a saber, ¿cuál es la razón de la interdependencia conceptual entre racionalidad y creencia? La respuesta de Stich es que si la descripción intencional presupone que nosotros somos similares -en ciertos aspectos cognoscitivos- a aquéllos a quienes interpretamos; entonces, puesto que nosotros somos más o menos buenos en la tarea de razonar, aquéllos a quienes interpretamos y que son similares a nosotros, deben también ser más o menos buenos en la tarea de razonar. En conclusión, Stich señala que la conexión entre racionalidad y creencia es “*un subproducto*” del principio de humanidad. En sus propias palabras Stich nos dice:

*“The minimal rationality condition is a by-product of the principle of humanity; and the latter principle obtains because in intentional description we characterize other people’s cognitive states via their similarity to our own.” (Stich, 1990; p. 50)*

Una vez dejado en claro que la conexión entre racionalidad y creencia es debida a la conexión entre similaridad y descripción intencional, Stich procede entonces a destacar que la posibilidad de describir intencionalmente a otros es una cuestión de grado: a aquellos más similares a nosotros nos será más fácil atribuirles estados intencionales y, por ende, tenderán a parecernos más racionales que aquellos cuyos procesos cognoscitivos sean muy diferentes de los nuestros. Si esto que nos propone Stich es correcto, entonces podemos inteligentemente imaginar una sucesión de personas ordenadas de acuerdo con sus similitudes cognoscitivas de modo tal que, la primera persona en la sucesión es muy similar a la segunda persona, la segunda persona en la sucesión es muy similar a la primera y a la tercera persona y así sucesivamente, de modo que para cada persona en esta sucesión sea muy fácil atribuir estados intencionales a las personas adyacentes. Sin embargo, la facilidad con que es posible interpretar a otros se irá perdiendo a medida en que éstos se encuentren más distantes hasta llegar al caso extremo de la primera y la última persona en esta sucesión, las cuales posiblemente no podrían interpretarse entre sí en lo absoluto. La diversidad cognoscitiva radical parece entonces plausible.

Tras la exposición previa podemos concluir esta sección señalando que las respuestas que Cherniak da a las objeciones planteadas por el argumento axiomático y por el argumento de inferencias constitutivas, así como el argumento de Stich a favor del principio de humanidad

en lugar del principio de caridad como guía apropiada para la interpretación, nos brindan elementos para rechazar la tesis, presupuesta por Davidson y por la epistemología tradicional, de que es necesario que haya un núcleo básico de reglas de inferencia compartido por todo individuo que pueda ser considerado un agente racional. Podemos establecer entonces la factibilidad de agentes cognoscitivos que razonen de acuerdo con principios muy distintos de los nuestros y que suscriban, consecuentemente, principios normativos de razonamiento sustancialmente distintos de los nuestros.

## 2.2 ARGUMENTOS EMPÍRICOS EN CONTRA DE LA UNIVERSALIDAD DE LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO.

En aras de mostrar la no universalidad de los principios normativos de razonamiento, una concepción naturalizada de la epistemología nos autoriza a echar mano de los resultados de la investigación empírica sobre el razonamiento. Es por esto que se dedica este apartado al análisis de las consecuencias que pueden tener los resultados de los experimentos de razonamiento para la tesis de la universalidad de los principios normativos de razonamiento.

Aún cuando los resultados de la investigación experimental sobre el razonamiento nos muestran, en ciertos casos, que hay tendencias en las maneras en que los sujetos responden a los diversos problemas de razonamiento que se les plantean; estos mismos resultados nos muestran también divergencias importantes en los tipos de respuestas emitidas por los sujetos. Este hecho nos sugiere que puede haber gran variabilidad en las maneras en que los seres humanos razonamos. Además, los resultados de experimentos realizados con el fin de evaluar los efectos de la enseñanza y el entrenamiento en el uso de ciertos principios de razonamiento, sugieren que también puede haber diferencias sustantivas en lo que los seres humanos *podemos* hacer en lo que al trabajo de razonar se refiere. Si esto último es en verdad el caso, y si retomamos el compromiso asumido en los apartados 1.2 y 1.3 con el principio de que *debe* implica *puede*, entonces hemos de concluir que; puesto que hay diferencias en lo que los sujetos *pueden* hacer en materia de razonamiento, debe entonces haber también diferencias en los estándares normativos con los que las personas deben ser evaluadas. Por lo tanto, no todo estándar normativo es aplicable a todo sujeto y los principios normativos de razonamiento no son necesariamente universales.

Para dar fuerza al argumento precedente revisaremos en seguida los resultados de algunos experimentos de razonamiento que pueden darnos evidencia a favor de la tesis de que hay diferencias importantes en las capacidades que los sujetos exhiben para razonar de acuerdo

con determinados principios o reglas de razonamiento; de modo que, mientras algunos sujetos *pueden* razonar de acuerdo con cierto principio de razonamiento, otros sujetos simplemente no *pueden* hacerlo.

Antes de entrar de lleno en la descripción de estos experimentos es importante precisar el sentido que aquí se está dando a la palabra *puede*. Se entenderá aquí que un determinado sujeto S *puede* razonar de acuerdo con cierto principio normativo de razonamiento X, si:

- O bien, la ejecución actual exhibida por S no diverge de la ejecución que sería esperada en el caso en que el principio X fuera el principio subyacente en la competencia de razonamiento de S, y hay además suficiente evidencia para descartar la posibilidad de que principios diferentes de X sean los principios subyacentes a la ejecución de S.
- O bien, es el caso que S no cumple la condición anterior, no obstante, S es capaz – tras un proceso de aprendizaje- de cumplir la condición anterior, es decir, de llegar a razonar de acuerdo con X.

En otras palabras, asumiremos que un sujeto puede razonar de acuerdo con cierto principio de razonamiento no sólo en el caso en que de hecho su ejecución sea acorde con dicho principio (y haya evidencia para descartar otros principios), sino también en el caso en que dicho sujeto pueda beneficiarse de un proceso de enseñanza o entrenamiento tras el cual pueda llegar a razonar de acuerdo con el principio en cuestión. Consecuentemente, asumiremos que un sujeto no puede razonar de acuerdo con cierto principio de razonamiento cuando la ejecución de dicho sujeto no se adhiere sistemáticamente a dicho principio y, además, ningún proceso de enseñanza resulta efectivo para lograr que ese sujeto llegue a razonar de acuerdo con el principio en cuestión.

En este tenor, los experimentos que describiré en seguida son interesantes en tanto que nos muestran la medida en que las personas pueden o no pueden mejorar su competencia de razonamiento, como resultado de haberse sometido a diferentes tipos de procesos de enseñanza.

Cheng, Holyoak, Nisbett & Oliver (1993), realizaron tres experimentos con el fin de probar las siguientes hipótesis:

- 1) Si la gente razona utilizando esquemas de razonamiento pragmático, entonces cualquier indicio semántico que pueda elicitar dichos esquemas ayudará a los sujetos en la solución apropiada (desde la perspectiva de la lógica estándar) de problemas condicionales, siempre y cuando sea el caso que la solución que el esquema genera

coincida con la solución generada mediante el sistema de reglas asociadas con el condicional material.

- 2) En tanto que el sistema de reglas asociadas con el condicional material no tiene aplicabilidad en contextos naturales (Ver pp. 190-191), la gente no posee la habilidad para interpretar problemas concretos en términos del condicional material. Por lo tanto, el entrenamiento abstracto o formal en el sistema de reglas asociadas con el condicional material no producirá mejoría significativa en la habilidad de las personas para solucionar problemas concretos; problemas que supuestamente requieren, para su solución, de la aplicación de reglas relativas al condicional material.
- 3) En contraste con el entrenamiento abstracto en el sistema de reglas asociadas con el condicional material, el entrenamiento abstracto en las reglas asociadas con un esquema de razonamiento pragmático (específicamente el esquema de “obligación”), producirá una mejoría significativa en la habilidad de los sujetos para aplicar dicho esquema. La razón para presuponer esto es que las reglas propias del esquema de razonamiento ya forman parte de la competencia de razonamiento de los sujetos.

#### Experimento 1.

En un primer experimento Cheng et.al. quisieron explorar hasta que punto los sujetos podían mejorar su ejecución en la tarea de selección como resultado de un entrenamiento previo en el condicional material y sus reglas asociadas. Ellos asignaron a 80 sujetos, los cuales no tenían ningún conocimiento previo de lógica, a cuatro diferentes grupos: 1) un grupo recibió un entrenamiento formal en la lógica del condicional material: se explicó a los sujetos la equivalencia entre el condicional material y su contrapositiva, así como las falacias más comunes de afirmar el consecuente y de negar el antecedente; 2) otro grupo no recibió entrenamiento formal en la lógica del condicional material, pero recibió un entrenamiento basado en ejemplos: se pidió a los sujetos intentar solucionar dos problemas de selección, inmediatamente después se les indicaron sus errores y se les hizo ver cuáles eran las respuestas correctas, todo esto en los términos específicos del problema en cuestión; 3) un tercer grupo recibió tanto el entrenamiento formal en la lógica del condicional material como el entrenamiento basado en ejemplos; 4) un cuarto grupo no recibió entrenamiento alguno. Después de que los primeros tres grupos pasaron por esta fase de entrenamiento, se presentó a todos los sujetos con ocho problemas de selección: seis de ellos involucrando una regla condicional, mientras que dos de ellos involucraban una regla bicondicional. De entre las seis reglas condicionales dos eran reglas arbitrarias, es decir, no se vinculaban con el conocimiento previo de los sujetos; dos eran reglas que –dado el conocimiento previo de

los sujetos- tenían la peculiaridad de sugerir la conversa (i.e. 'si  $q$  entonces  $p$ ' dado que 'si  $p$  entonces  $q$ '); y dos eran reglas fácilmente interpretables como permisos.

Los resultados de Cheng et. al. mostraron básicamente dos cosas. 1) Por una parte se observó un efecto del tipo de problema: el porcentaje de sujetos que cometió a lo menos un error<sup>21</sup> fue mucho menor para los problemas de permiso ( $34 \pm 6\%$ ), algo mayor para los problemas que sugerían la conversa ( $66 \pm 6\%$ ), y el porcentaje más alto de sujetos que cometió a lo menos un error fue para los problemas arbitrarios ( $81 \pm 6\%$ ). Los autores consideran que este resultado es consistente con su hipótesis de que las personas razonamos mediante esquemas de razonamiento pragmático. 2) Por otra parte se observó el efecto que la condición de entrenamiento tuvo sobre la ejecución de los sujetos; los porcentajes de sujetos que cometieron a lo menos un error son los siguientes:

<b>Condición de entrenamiento.</b>	<b>Porcentaje de sujetos que cometieron a lo menos un error en los problemas de selección condicional.</b>
Entrenamiento formal en el condicional material & Entrenamiento basado en ejemplos.	$39 \pm 12$
Entrenamiento formal en el condicional material.	$65 \pm 12$
Entrenamiento basado en ejemplos.	$62 \pm 12$
Control: No entrenamiento	$75 \pm 12$

Estos datos muestran que tanto el entrenamiento formal en el condicional material por sí sólo, como el entrenamiento basado en ejemplos por sí sólo, no produjeron ninguna mejoría significativa en la ejecución de los sujetos. En contraste, en el caso en que el entrenamiento formal estuvo asociado con el entrenamiento basado en ejemplos, sí pudo observarse una mejora significativa en la ejecución de los sujetos en comparación con el grupo control. Cheng et. al. declaran que el entrenamiento formal (por sí sólo), resulta inefectivo para producir una mejoría en la ejecución de los sujetos, porque los sujetos no tienen la habilidad para aplicar la regla condicional formal a situaciones concretas; por otra parte, ellos declaran que el entrenamiento basado en ejemplos (por sí sólo) es inefectivo porque los sujetos no



tienen la habilidad de captar cuál es la regla que subyace en los ejemplos que les están siendo presentados. Para Cheng y sus colegas este resultado confiere sustento a su hipótesis de que el sistema de reglas asociadas con el condicional material no es un sistema de reglas que la gente usa naturalmente.

### Experimento 2.

En un segundo experimento Cheng et. al. quisieron descartar la posibilidad de que la falta de efectividad del entrenamiento en la lógica del condicional (por sí sólo) y del entrenamiento basado en ejemplos (por sí sólo), se hubiera debido al hecho de que el “micro-curso” sobre la lógica del condicional que ellos impartieron, hubiera sido demasiado corto e insuficiente para producir alguna mejoría. Por esta razón, en el experimento 2 ellos examinan el efecto que un curso mucho más amplio –un curso de un semestre en lógica estándar- puede tener en la habilidad de los sujetos para solucionar problemas condicionales.

En este experimento participaron 53 estudiantes que debían tomar un curso introductorio de lógica. Dichos estudiantes fueron sometidos a un pretest y a un pos-test. El pretest se aplicó en la primera semana de clases, antes de que los sujetos recibieran algún tipo de instrucción en la lógica del condicional. El pos-test se aplicó en la última semana de clases. Los problemas condicionales utilizados fueron los mismos del experimento 1, sólo que en este caso, los ocho problemas fueron divididos en dos grupos de cuatro, incluyendo cada grupo un problema condicional arbitrario, un problema condicional que sugería su conversa, un problema condicional de permiso y un problema bicondicional. Los dos conjuntos resultantes de problemas fueron asignados como material del pretest o como material del pos-test en un orden contrabalanceado.

El resultado observado en este experimento fue que no hubo ninguna diferencia significativa en el porcentaje de sujetos que cometieron a lo menos un error en los problemas condicionales en el pretest ( $75 \pm 5$ ), y en el pos-test ( $72 \pm 5$ ). Ahora bien, considerando individualmente cada uno de los posibles tipos de errores cometidos por los sujetos, pudo observarse una mejoría solamente para un tipo de error, el error de seleccionar  $q$ . Este error decreció un  $10 \pm 8\%$  del pretest al pos-test.

### Experimento 3.

En este experimento Cheng et. al. examinan la posibilidad de producir una mejoría en la ejecución de los sujetos, no ya mediante un entrenamiento abstracto en la lógica del condicional material, sino mediante un entrenamiento abstracto en una regla que los autores

consideran, sí forma parte del repertorio de razonamiento intuitivo de los sujetos. Me refiero aquí a una regla de obligación.

El entrenamiento en la regla de obligación consistió en lo siguiente. Los sujetos recibieron un texto de dos páginas en que se detallaba la naturaleza de las obligaciones, especificando que éstas pueden asumir la siguiente forma: “si se da la situación x, entonces la acción A debe ser realizada”. Se les daba también el ejemplo de una oración que expresa obligación presentada en la forma si-entonces, así como procedimientos para determinar si la violación de una obligación ha ocurrido. Estos procedimientos especifican reglas a seguir en cada uno de los cuatro posibles casos: 1) se da una cierta situación x, 2) no se da la situación x, 3) la acción A se realiza, 4) la acción A no se realiza. Se dio diez minutos a los sujetos para revisar este material e inmediatamente después se les pidió resolver ocho problemas de selección. Cuatro de estos problemas eran problemas relativamente arbitrarios y cuatro eran problemas fácilmente interpretables como obligaciones.

Los resultados arrojados por este experimento muestran un efecto favorable altamente significativo debido al entrenamiento: el porcentaje de sujetos que cometieron a lo menos un error en el grupo que recibió el entrenamiento fue de  $27 \pm 12$ , en comparación con un porcentaje de  $54 \pm 12$  de sujetos que cometieron a lo menos un error en un grupo control. Cheng et. al. también observaron que el entrenamiento en la regla de obligación resultó en una mejoría en la ejecución de los sujetos, no sólo para los problemas de obligación, sino también para el caso de los problemas arbitrarios algunos de los cuales pudieron ser interpretados como obligaciones. En estos resultados, Cheng et. al. encuentran nuevamente sustento para su tesis de que los seres humanos razonamos mediante esquemas de razonamiento pragmático tales como el esquema de obligación.

Hasta aquí la descripción de los experimentos. Veamos ahora como ellos pueden apoyar la tesis de que hay diferencias en los principios de razonamiento que los sujetos *pueden* llegar a utilizar. Fundamentalmente lo que estos experimentos muestran es, no sólo que los sujetos no razonan de acuerdo con las reglas puramente sintácticas de la lógica estándar asociadas con el condicional material, sino además, que la mayoría de los sujetos *no pueden* (en el sentido de “*pueden*” aquí utilizado) razonar de acuerdo con dichas reglas -en el sentido de aplicarlas provechosamente en contextos de la vida práctica cotidiana- toda vez que ellos no pueden beneficiarse de ningún tipo de proceso de enseñanza, puramente abstracto, encaminado a promover su aplicación. Para poner lo anterior en términos más concretos, pensemos en el caso del error más frecuente que se comete en la tarea de selección, la no selección de la carta correspondiente a  $\neg q$ . En tanto que el entrenamiento puramente

abstracto en la lógica del condicional material resulta casi totalmente inefectivo en la reducción de este tipo de error y en tanto que dicho error está asociado a la no aplicación de la regla *modus tollens*, tenemos buenos elementos para concluir que el *modus tollens* no es una regla subyacente al razonamiento intuitivo de muchos de los sujetos. Como consecuencia, en el ámbito normativo no podemos asumir que *modus tollens* sea un principio normativo de razonamiento para esos sujetos. Sin embargo, los resultados de los experimentos de Cheng et. al. no excluyen la posibilidad de que ciertos sujetos sí puedan razonar de acuerdo con dicha regla. Pues para los tres experimentos siempre hubo sujetos que acertadamente seleccionaron la carta correspondiente a  $\neg q$ . Por ejemplo, en el caso del primer experimento, Cheng et. al. reportan los siguientes porcentajes de sujetos que cometieron el error de no seleccionar  $\neg q$  como una función del tipo de problema:

<b><i>Tipo de problema.</i></b>	<b><i>Porcentaje de sujetos que cometieron el error de no seleccionar <math>\neg q</math>.</i></b>
Condicional Arbitrario.	62
Condicional que sugiere su conversa.	42
Condicional de permiso.	24

lo cual significa que para el primer tipo de problema un 38% de sujetos seleccionó acertadamente la carta  $\neg q$ , para el segundo tipo de problema un 58%, y para el tercer tipo de problema un 76% de los sujetos seleccionaron  $\neg q$ . En el caso de los problemas de permiso dicho resultado favorable se explica por la elicitación del esquema de permiso, cuyo sistema de reglas conduce a soluciones que coinciden con las soluciones que resultarían de la aplicación del sistema de reglas asociadas al condicional material. Sin embargo, en el caso de los condicionales arbitrarios, en los que no se espera la elicitación del esquema de permiso, podemos pensar que los sujetos (38%) que seleccionaron  $\neg q$  pudieron valerse del *modus tollens*. Estos resultados muestran que algunos sujetos de los que participaron en este experimento *pueden* de hecho razonar de acuerdo con la regla lógica *modus tollens*, mientras que la mayor parte de los sujetos no razonan *ni pueden razonar* de acuerdo con dicha regla.

La conclusión anterior puede también ser apoyada por los resultados de los experimentos presentados en los dos primeros capítulos de este texto. Recordemos que, en esos

experimentos siempre hay un porcentaje, aunque sea mínimo, de sujetos que dan la solución correcta cuando se les presenta un problema condicional arbitrario.

Así pues, parece haber diferencias individuales sustantivas en la capacidad de los sujetos para razonar de acuerdo con determinados principios de razonamiento. En los términos de Cherniak, podría decirse que hay diferencias en los ordenamientos de factibilidad de distintas personas. Si esto último es correcto, podemos entonces afirmar que hay diferencias en lo que los distintos seres humanos *pueden* hacer en materia de razonamiento. Es posible, por lo tanto, concluir que los principios normativos de razonamiento no pueden ser aplicados indistintamente a todo sujeto. Tampoco pueden ser indistintamente aplicados en todo tiempo y lugar, pues las capacidades inferenciales de las personas pueden variar, no sólo entre individuos, sino incluso para el mismo individuo de un momento a otro. Tal cosa puede ocurrir cuando una persona puede beneficiarse de un proceso de enseñanza que le permite mejorar su competencia en la solución de determinado tipo de problemas. Es posible también que haya diferencias en la competencia de razonamiento de personas que pertenecen a distintas culturas. Tal situación puede darse en la medida en que diferentes ambientes pueden favorecer -dados los diferentes tipos de problemas a que los grupos de individuos deben enfrentarse en diferentes contextos físicos y sociales- que diferentes tipos de esquema de razonamiento sean inducidos.

Si, dadas las razones arriba apuntadas, aceptamos que los principios normativos de razonamiento no pueden ser aplicados a todo sujeto en todo tiempo y lugar, simplemente porque es impropio exigir a un sujeto que haga lo que, dadas sus capacidades cognitivas, es incapaz de hacer; entonces tendremos que aceptar que los principios normativos de razonamiento no son necesariamente universales.

### 2.3 RELATIVISMO.

La epistemología tradicionalmente se ha comprometido con la idea de que los estándares normativos de racionalidad deben ser universalmente aplicables. Es decir, dichos estándares deben tener fuerza normativa para todo sujeto en todo tiempo y lugar. De no ser así, caeríamos en un relativismo desenfrenado en el que cada quién podría regirse por sus propios estándares, cayendo a final de cuentas en la posición de "todo vale", y si todo vale, no tiene sentido entonces el hablar de normas de razonamiento.

Las opciones para la epistemología tradicional se reducen entonces a dos: o absolutismo normativo o relativismo desenfrenado. En tanto que lo segundo es totalmente inaceptable, pues implica que no hay normas, lo primero debe ser entonces el caso.

En contra de la epistemología tradicional, la tesis que aquí se ha defendido es que los principios normativos de razonamiento no son absolutos ni universales, por ende se ha

asumido una posición relativista. Sin embargo, puede ser dicho, a favor de la posición aquí asumida, que el relativismo que ésta implica no es un relativismo desenfrenado del tipo “todo vale”, no es tampoco un relativismo demasiado extremo.

Un claro ejemplo de un relativismo que sí es desenfrenado es el que se sigue de la noción de justificación implícita en las tesis de Protágoras. Dicha noción de justificación es la siguiente:

Si un sujeto *S* cree que *p* en *t*, entonces *S* está justificado en creer que *p* en *t*.

Para esta concepción de la justificación cualquier creencia que un sujeto pueda tener es epistémicamente aceptable, y puesto que diferentes personas pueden tener creencias muy diferentes e incluso contradictorias, lo que es válido y lo que no depende del punto de vista de cada quien, de modo que todo vale. Por tanto, un relativismo de corte protagórico es en verdad inaceptable.

Otra postura epistémica que conduce al relativismo es el pragmatismo epistémico de Stephen Stich. Según Stich, en la evaluación de sistemas o de procesos cognoscitivos, un sistema será considerado como mejor que otros en la medida en que éste contribuya, de la manera más satisfactoria, al logro de cosas que son intrínsecamente valiosas para un sujeto *S* cuyos procesos cognoscitivos están siendo evaluados.

Esta posición es relativista en tanto que las cosas que cada ser humano valora, sus deseos y metas, pueden de hecho ser muy diferentes. De aquí que, lo que es bueno o aceptable para una persona puede no ser bueno o aceptable para otra.

A favor del relativismo sticheano puede ser dicho que este no es un relativismo desenfrenado para el cual todo vale, pues cuando un sujeto intenta lograr la satisfacción de sus deseos mediante procesos cognoscitivos o mecanismos inferenciales que no constituyen la mejor forma para alcanzar sus metas, podemos decir que dicho sujeto está procediendo de manera irracional. Sin embargo, el relativismo de Stich conduce a una concepción de los principios normativos de razonamiento demasiado individualista en el sentido de que cada persona tendrá su propio conjunto de principios normativos, y los principios que tengan fuerza normativa para una persona difícilmente la tendrán para otras. Así pues, aunque el relativismo de Stich no es un relativismo desenfrenado, sí es un relativismo demasiado extremo.

En contraste con los dos relativismos previos, el relativismo que se desprende de la posición que aquí se ha asumido no es, como ya se ha dicho, ni desenfrenado, ni demasiado extremo. Veamos porqué.

La posición aquí asumida en relación a los principios normativos de razonamiento es la siguiente:

Si un principio de razonamiento  $P$  tiene fuerza normativa para un sujeto  $S$  en  $t$ , entonces  $S$  *puede* usar  $P$  en  $t$ .

Puesto que los experimentos de razonamiento han mostrado diferencias individuales en lo que los sujetos *pueden* hacer en materia de razonamiento, entonces debe haber también diferencias en los principios normativos de razonamiento de acuerdo con los cuales diferentes individuos *deben* razonar. De aquí que los principios normativos de razonamiento no sean universales.

Ahora bien, si nosotros clasificamos a las personas de acuerdo con lo que *pueden* hacer en lo que a la tarea de razonar refiere, nosotros podremos formar varios conjuntos de personas. Cada uno de dichos conjuntos estará constituido por personas con competencias de razonamiento muy similares, mientras que personas pertenecientes a distintos conjuntos o grupos tendrán competencias de razonamiento diferentes.

Una vez hecho esto se podría, sobre la base de la información sobre las distintas capacidades de razonamiento de las personas (aunque no exclusivamente sobre esta base), determinar cuáles son los principios normativos de razonamiento correctos que deben normar el razonamiento para cada uno de los grupos. De modo que, los principios normativos de razonamiento serán los mismos para personas pertenecientes a un mismo grupo, y diferentes para personas pertenecientes a distintos grupos.

Claramente, esto que acabo de decir conlleva un relativismo. Pero este relativismo no es desenfrenado, pues no implica que "todo vale". Y no todo vale porqué, como ha sido dicho en la sección 1.3, no todos los principios conforme a los cuales un grupo de personas puede razonar son principios normativos correctos. Puede darse el caso de que un principio subyacente en el razonamiento de cierto grupo de sujetos, sea inaceptable desde un punto de vista epistémico, no obstante que dichos sujetos puedan razonar conforme a él.

Este relativismo no es tampoco tan extremo como el de Stich. El relativismo de Stich es demasiado individualista en el sentido de que los principios normativos correctos son diferentes para cada uno de los distintos individuos, pues es muy difícil que distintos individuos coincidan en sus jerarquías de valores, consecuentemente será muy difícil que coincidan en los procesos que serán apropiados para alcanzar sus diferentes metas. En cambio, en el relativismo que aquí se sustenta, es muy plausible encontrar grupos de individuos que, en tanto que miembros de un mismo grupo cultural o de una misma sociedad, compartan principios de razonamiento. Por ejemplo, una comunidad de Físicos, en tanto que han recibido una formación profesional similar, comparten un conjunto de principios metodológicos y de racionalidad que son aceptables para la realización de su trabajo. De

este modo puede haber un conjunto de estándares normativos que tengan fuerza normativa para un gran número de personas. Por lo tanto, el relativismo que se desprende de la posición aquí defendida no es tan individualista y, por tanto no es tan extremo como el de Stich.

### **3. LOS PRINCIPIOS NORMATIVOS DE RAZONAMIENTO NO SON, NECESARIAMENTE, PRINCIPIOS FORMALES.**

Según la visión estándar de la racionalidad, los principios normativos correctos conforme a los cuales los seres humanos debemos razonar, son los principios de la lógica clásica, teoría de la probabilidad y teoría de la elección racional. Dichos principios son principios formales, es decir, son principios de razonamiento que no hacen alusión específica a los contenidos de los problemas sobre los que versa el razonamiento, sino simplemente a su estructura o forma. De ahí que un principio formal de razonamiento pueda ser aplicado a un sin número de problemas de razonamiento independientemente de su contenido. Por esto es que puede decirse que los principios formales de razonamiento son en sí mismos reglas de *propósito general o de dominio general*.

Durante los primeros años de la investigación psicológica del razonamiento humano, la visión estándar de la racionalidad era la única visión prevaleciente. Consecuentemente la mayor parte de los psicólogos presuponían que los principios subyacentes en la competencia de razonamiento humano eran principios formales, por lo tanto, principios demasiado generales como para poder ser aplicados en la solución de un gran número de problemas de razonamiento sin importar cual fuera su contenido.

Sin embargo, conforme la investigación del razonamiento se ha ido desarrollando, se ha acumulado evidencia que muestra la relevancia del contenido de los problemas para la solución de los mismos. Esto sugiere fuertemente que muchos de los principios subyacentes en la competencia humana de razonamiento no son independientes de contenido, sino más bien principios de dominio específico, por tanto, principios cuyo rango de aplicabilidad está acotado a ciertos tipos de problemas cuyos contenidos son propios de un ámbito temático específico.

Si consideramos los resultados de los experimentos que han sido descritos en el presente trabajo, hay tres cosas que podemos notar: 1) primero, que en general, la ejecución de los sujetos mejora significativamente cuando los problemas planteados manipulan contenidos concretos y/o contextos realistas, lo cual es indicativo de que principios de dominio específico están en juego. 2) segundo, que siempre hay un número, aunque sea pequeño,

de sujetos que resuelven correctamente los problemas con contenidos abstractos, lo cual sugiere que algunos sujetos pueden, posiblemente, valerse de principios formales. 3) tercero, que siempre hay sujetos, aunque pocos, que no pueden resolver correctamente los problemas que se les presentan ni aún cuando dichos problemas manejan contenidos concretos y/o contextos realistas, ni aún cuando han recibido un entrenamiento para el manejo de reglas más específicas (como en el caso del experimento 3 de Cheng et.al. 1993). Este resultado sugiere que ciertos principios de razonamiento de dominio específico, aún cuando subyacen en la ejecución de la mayoría de los sujetos, no son del todo universales. A partir de aquí podemos hacernos de una imagen de la competencia humana de razonamiento como constituida por un sin número de mecanismos inferenciales específicos de dominio, y en menor medida y para ciertos sujetos, mecanismos inferenciales de dominio general.

Si, en términos generales, los seres humanos nos valemos de principios inferenciales de dominio específico, entonces, cuando se trata de evaluar nuestros mecanismos inferenciales, los principios normativos de razonamiento deben estar necesariamente relativizados a dominios específicos. Imagínese, por ejemplo, el mecanismo inferencial encargado de “detección de tramposos” que Cosmides propone. Dicho mecanismo cuenta con reglas que pueden aplicarse con éxito en la detección de posibles tramposos, sin embargo, dicho mecanismo no puede entrar en operación en el caso de otro tipo de problemas; consecuentemente, dicho mecanismo no sólo no puede llevar a la respuesta correcta cuando se trata de resolver un problema distinto del de detectar tramposos, sino que no puede llevar a respuesta alguna. Si esto es así, es entonces absurdo demandar que el mecanismo de detección de tramposos es deficiente desde un punto de vista epistémico, en tanto que no puede darnos la respuesta correcta a un problema, pongamos por caso, aritmético. Hacer esto sería tan absurdo como quejarnos de un horno de microondas que no nos sirve como lavaplatos. Para evaluar el funcionamiento de un horno de microondas nosotros debemos considerar su ejecución en un rango de tareas para cuya realización fue específicamente diseñado y, obviamente, dentro de este rango no queda incluida la tarea de lavar platos. Recordemos aquí que *debe* implica *puede*. Sostendré por tanto que, si como la investigación del razonamiento humano sugiere, las reglas que constituyen la competencia de razonamiento humana son de dominio específico; entonces la evaluación epistémica de dichas reglas debe también estar relativizada a esos dominios. De modo que tiene sentido considerar a un determinado mecanismo inferencial encargado de detectar tramposos como mejor que otro mecanismo encargado de la misma función, en tanto que el primero nos lleva efectivamente a evitar intercambios sociales desventajosos la mayor parte de las veces, mientras que el segundo nos permite evitar dicho tipo de intercambios sólo algún número



mínimo de veces. Sin embargo, es francamente absurdo asumir que un mecanismo inferencial cuya función es la detección de tramposos es peor que otro mecanismo cuya función sea, por poner un caso, la solución de problemas matemáticos, pues parece muy difícil pensar en un criterio más general, aplicable a ambos dominios, que nos permitiera emitir dicho juicio. Así pues, los criterios con los que evaluamos un cierto mecanismo inferencial deben ser criterios específicos al dominio de problemas al que dicho mecanismo se aplica.

Aceptar que los principios normativos con los que evaluamos el razonamiento humano deben estar relativizados a dominios específicos, nos lleva forzosamente a reconsiderar lo que ha de entenderse por “ser racional”. A la luz de las consideraciones precedentes, ser racional no puede ya ser estrictamente identificado con “razonar de acuerdo con los principios de la lógica clásica, teoría de la probabilidad y teoría de la elección racional” como la visión estándar de la racionalidad postula.

#### NOTAS.

---

<sup>19</sup> Está implícito aquí el presupuesto empirista de que no existen enunciados “sintéticos *a priori*”.

<sup>20</sup> Para Quine una ‘verdad lógica’ “es un enunciado que es verdadero y sigue siéndolo para cualquier interpretación de sus componentes que no sean partículas lógicas” (Quine, 1953; p.53)

<sup>21</sup> Se considera que para problemas condicionales hay cuatro posibles tipos de errores: fallar en seleccionar  $p$  (no se realiza modus ponens); fallar en seleccionar  $\neg q$  (no se realiza modus tollens); seleccionar  $q$  (se comete la falacia de afirmar el consecuente); seleccionar  $\neg p$  (se comete la falacia de negar el antecedente).

## APENDICE A.

### VERSIONES COMPLETAS DE LOS PROBLEMAS PLANTEADOS EN LOS EXPERIMENTOS DE COSMIDES (1989)

Experimentos 1 y 2 . (experimentos 1 y 3 de Cosmides).

Los siguientes cuatro problemas fueron usados tanto en el experimento 1 como en el experimento 2, sólo que las reglas para los dos problemas no familiares se presentaron invertidas en el experimento 2.

#### 1) Problema de contrato social estándar no familiar. (CSE-noF)

Versión “raíz de cassava”:

*“Tú eres un Kaluame, un miembro de una cultura polinesia ubicada en la isla de Maku en el Pacífico. Los Kaluame tienen muchas leyes estrictas las cuales deben ser seguidas, y los más viejos te han confiado la tarea de vigilar que se cumplan. Fallar traería desgracias a ti y a tú familia.*

*Entre los Kaluame, cuando un hombre se casa, él es tatuado en la cara; solamente los hombres casados tienen tatuajes en la cara. Un tatuaje facial significa que un hombre es casado, una cara no marcada significa que un hombre es soltero.*

*La raíz de cassava es un poderoso afrodisíaco –ésta hace al hombre que la come irresistible para las mujeres. Además es deliciosa y nutritiva- y muy escasa.*

*A diferencia de la raíz de cassava, las nueces de molo son muy comunes, pero ellas son un alimento pobre –las nueces de molo saben mal, no son muy nutritivas, y no tienen propiedades “medicinales” interesantes.*

*Aunque todos desean raíz de cassava, comerla es un privilegio que tu gente estrechamente raciona. Tú eres una persona muy sensual, aún sin las propiedades afrodisíacas de la raíz de cassava, pero tú tienes una moral sexual muy estricta. Los más viejos desapruueban enérgicamente las relaciones entre personas no casadas y particularmente desconfían de los motivos e intenciones de los solteros.*

*Por tanto, los más viejos han hecho leyes para regular la obtención de privilegios. La ley que se te ha encomendado a ti hacer cumplir es la siguiente:*

*“Si un hombre come raíz de cassava, entonces él debe tener un tatuaje en la cara”.*

*La raíz de cassava es un afrodisíaco tan poderoso, que muchos hombres tienen la tentación de violar esta ley siempre y cuando los más viejos no se den cuenta. Las cartas abajo tienen*

información acerca de cuatro hombres jóvenes Kaluame sentados en un campamento temporal; no hay ancianos a la redonda. Una bandeja llena con raíces de cassava y nueces de molo ha sido dejada para ellos. Cada carta representa a un hombre. Un lado de la carta te dice que alimento el hombre está comiendo, y el otro lado de la carta te dice si el hombre tiene o no un tatuaje en su cara.

Tú tarea es descubrir al hombre cuyos deseos sexuales le hayan llevado a romper la ley –si alguno te engaña a ti, tú y tu familia serán desgraciados. Indica solamente aquella(s) carta(s) que tú definitivamente necesitar voltear para ver si alguno de estos hombres Kaluame está rompiendo la ley”.

Las cuatro cartas son: “come raíz de cassava”, “no tatuado”, “come nueces de molo”, “tatuado”.

#### Versión “carne de duiker”.

“Tú eres un antropólogo estudiando a los Namka, una cultura de cazadores-recolectores viviendo en los desiertos al suroeste de Africa. Tú estás particularmente interesado en si los muchachos Namka obedecen las leyes de su gente.

Cada luna llena hay un especial festejo en el cual un duiker –un pequeño antílope- es matado y comido. La carne de duiker es completamente escasa y deliciosa –un banquete real. Comer carne de duiker es un privilegio que debe ser ganado.

Para los muchachos este privilegio está gobernado por la siguiente ley:

“Si tú comes carne de duiker, entonces tú has encontrado una cáscara de huevo de avestruz”.

Encontrar cáscaras de huevo de avestruz es una tarea sofisticada y difícil la cual lleva a un muchacho años aprender. Haber encontrado una cáscara de huevo de avestruz por ti mismo es por tanto un signo de que tú has manejado las destrezas más difíciles de la cacería. Para los Namka, esto representa una transición de los muchachos a la edad viril.

Tú te preguntas si los muchachos Namka engañan sobre esta ley cuando nadie los está mirando. Tú decides ocultarte detrás de unos arbustos y observar. Durante el curso del festejo de la luna llena, tú ves cuatro diferentes muchachos acercarse al duiker asado mientras nadie más está mirando.

Las cartas abajo tienen información acerca de estos cuatro muchachos. Cada carta representa a un muchacho. Un lado de la carta dice si el muchacho ha encontrado una cáscara de huevo de avestruz, y el otro lado de la carta dice si ese muchacho tomó algo de la carne asada de duiker.

*El olor de duiker asado es verdaderamente tentador para los muchachos. Tú quieres saber si alguno de ellos violó la ley. Indica solamente aquella(s) carta(s) que tú definitivamente necesitas voltear para ver si alguno de estos muchachos ha roto la ley”.*

*Las cuatro cartas son: “come algo de carne de duiker”, “nunca ha encontrado una cáscara de huevo de avestruz”, “no come nada de la carne de duiker”, “ha encontrado una cáscara de huevo de avestruz”.*

## **2) Problema descriptivo no familiar. (D-noF)**

### Versión “raíz de cassava”

*“Tú eres un antropólogo estudiando a los Kaluame, una cultura polinesia ubicada en la isla de Maku en el Pacífico. Antes de partir para la Isla de Maku tú lees un reporte que dice que algunos hombres Kaluame tienen tatuajes en sus caras, y que ellos comen raíces de cassava o nueces de molo, pero no ambos. El autor del reporte, quien no hablaba la lengua Kaluame, dijo que la siguiente relación parecía sostenerse:*

*“Si un hombre come raíz de cassava, entonces él debe tener un tatuaje en la cara”*

*Tú decides investigar la peculiar aseveración de tú colega. Cuando tú llegas a la isla de Maku, tú aprendes que la raíz de cassava es un alimento básico encontrado al sur de la isla. Las nueces de molo son muy altas en proteína, y crecen en los árboles de molo, los cuales son principalmente encontrados sobre la costa norte de la isla.*

*Tú también aprendes que los solteros viven principalmente sobre la costa norte, pero cuando los hombres se casan, ellos usualmente se mudan al sur de la isla. Cuando un hombre Kaluame se casa, él es tatudo en la cara; solamente los hombres casados tienen tatuajes en la cara. Un tatuaje facial significa que un hombre es casado, una cara no marcada significa que un hombre es soltero. Quizá los hombres están simplemente comiendo los alimentos que tienen más disponibles.*

*Las cartas abajo tienen información acerca de cuatro hombres Kaluame sentados en un campamento temporal al centro de la isla. Cada hombre está comiendo raíz de cassava o nueces de molo las cuales él ha traído desde su hogar. Cada carta representa a un hombre. Un lado de la carta dice qué comida el hombre está comiendo y el otro lado de la carta dice si o no el hombre tiene un tatuaje en la cara. La regla mencionada por tu colega puede no ser verdadera; tú quieres ver por ti mismo. Indica solamente aquella(s) carta(s) que tú definitivamente necesitas voltear para ver si alguno de estos hombres Kaluame está rompiendo la regla”.*

*Las cuatro cartas son: “no tatuado”, “tatuado”, “come raíz de cassava”, “come nueces de molo”.*

### Versión “carne de duiker”

*“Tú eres un antropólogo estudiando a los Namka, una cultura de cazadores-recolectores en los desiertos al suroeste de África. Una y otra vez tú escuchas a varios Namka repetir el siguiente dicho:*

*“Si tú comes carne de duiker, entonces tú has encontrado una cáscara de huevo de avestruz”.*

*Los duikers son pequeños antílopes encontrados en la parte oriental de la extensión del hogar Namka. Ambas, carne de duiker y cáscaras de huevo de avestruz, son buscadas por los Namka: ellos comen la carne y usan las cáscaras como cantinforas porque ellas son claras y sostienen mucha agua. Además, los duikers frecuentemente se alimentan de huevos de avestruz.*

*Como antropólogo, tú no sabes si ese dicho es metafórico, refiriendo, por ejemplo, al territorio del clan o a prácticas rituales, o si el dicho refleja una relación real que los Namka usan para guiar su conducta de forrajeo. ¿Significa éste que si tú encuentras el primero tú encuentras el segundo? Esto es lo que tú estás tratando de descubrir.*

*¿Es esto un hecho o folklore? ¿el Namka quiere decir cáscara de huevo y carne de duiker, o son estas cosas meramente símbolos para algo enteramente distinto? Desafortunadamente tú no sabes su lenguaje lo suficientemente bien como para preguntarles a ellos. Es así que tú decides investigar si la regla establecida en este dicho tiene alguna base factual.*

*Muchas especies de aves pueblan el área, y en tus caminatas tú has encontrado varios escondites de huevos de varias clases. Las cartas abajo tienen información acerca de cuatro diferentes localizaciones de escondites de huevos. Cada carta representa una localización, y cada localización tiene las huellas de un mamífero asociado con ésta. Un lado de la carta dice que clase de cáscara de huevo tú encontraste en una localización y el otro lado de la carta dice las huellas de qué mamífero tú encontraste ahí.*

*Quizá el dicho de los Namka no tiene bases factuales. Indica solamente aquella(s) carta(s) que tú definitivamente necesitas voltear para ver si alguna de estas localizaciones viola la regla expresada en el dicho Namka”.*

*Las cartas son: “cáscara de huevo de codorniz”, “cáscara de huevo de avestruz”, “duiker”, “comadreja”.*

### **3) Problema descriptivo familiar. (D-F)**

*“Parte de tú nuevo trabajo para la Ciudad de Cambridge es estudiar las demografías de transportación. Tú lees un reporte previamente hecho sobre los hábitos de los residentes de Cambridge el cual dice:*

*“Si una persona va a Boston, entonces toma el metro”*

*Las cartas abajo tienen información sobre cuatro residentes de Cambridge. Cada carta representa una persona. Un lado de la carta dice a donde fue la persona y el otro lado de la carta dice cómo la persona llegó ahí.*

*Indica solamente aquella(s) carta(s) que tú definitivamente necesitas voltear para ver si alguna de estas personas viola esta regla”.*

Las cartas son: “metro”, “Arlington”, “taxi”, “Boston”.

#### **4) Problema Abstracto (ABS)**

*“Parte de tú nuevo trabajo clerical en la preparatoria local es asegurar que los documentos de los estudiantes han sido procesados correctamente. Tú trabajo es asegurar que los documentos se conforman a la siguiente regla alfanumérica:*

*“Si una persona es clasificada ‘D’, entonces sus documentos deben ser marcados código ‘3’”. Tú sospechas que la secretaria que reemplazaste no categorizó los documentos de los estudiantes correctamente. Las cartas abajo tienen información acerca de los documentos de cuatro personas quienes han ingresado a esta preparatoria. Cada carta representa a una persona. Un lado de la carta dice la letra de clasificación de una persona y el otro lado de la carta el número de código de la persona.*

*Indica solamente aquella(s) carta(s) que tú definitivamente necesitas voltear para ver si los documentos de alguna de estas personas viola esta regla.*

Las cartas son: “D”, “F”, “3”, “7”.

#### Experimentos 3 y 4. (experimentos 5 y 8 de Cosmides).

Los siguientes dos problemas se usaron tanto en el experimento 3 como en el experimento 4, sólo que en el experimento 4 las reglas fueron invertidas.

#### Versión: “preparatoria Grover”

##### **1) Problema de permiso con la estructura de un contrato social estándar. (P-CSE)**

*“Tú supervisas a cuatro mujeres quienes fueron voluntarias para ayudar a la Junta Local de Educación. Cuando tú viniste a trabajar hoy, encontraste que había algunos rumores y murmuraciones. Se supone que tus voluntarias siguieron ciertas reglas para asignar estudiantes de varios pueblos a la escuela de distrito apropiada. Cada voluntaria es madre de un joven quien está por entrar a la preparatoria, y cada una procesó los documentos de su propio hijo. Por lo tanto ahora corren rumores de que tus voluntarias hicieron trampa cuando ellas asignaron a sus propios hijos a una escuela. Esta es la situación:*

Los estudiantes van a ser asignados ya sea a la preparatoria Grover, la cual se localiza en Grover City, o a la preparatoria Hannover, la cual se localiza en el pueblo de Hannover. La preparatoria Grover es una gran escuela con un excelente registro de estudiantes colocados en buenos colegios. En contraste, la preparatoria Hannover es una escuela mediocre con malos maestros y facilidades decrepitas.

La razón de que las escuelas sean tan diferentes es la voluntad de los padres de cada comunidad para apoyar económicamente a sus escuelas a través de impuestos. Aunque ambas comunidades son igualmente prósperas, los padres en Grover City se han preocupado siempre por la calidad de sus escuelas, incluyendo la preparatoria Grover, y han estado dispuestos a pagar por ésta. En contraste, los padres en el vecino pueblo de Hannover y Belmont nunca han querido gastar dinero, y se han opuesto a cualquier impuesto para mejorar la preparatoria Hannover.

La Junta de Educación tomó estos factores en consideración cuando creó reglas para determinar a qué escuela debe un estudiante ser asignado; la más importante de estas reglas es:

*“Si un estudiante es asignado a la escuela preparatoria Grover, entonces ese estudiante debe vivir en Grover City”*

Se supone que tus voluntarias siguieron esta regla cuando procesaron todos los documentos de los estudiantes -¡Incluyendo los documentos de sus propios hijos!- Tú debes descubrir si los rumores son verdaderos: ¿alguna de tus voluntarias hizo trampa cuando procesó los documentos de su propio hijo?

Las cartas abajo tienen información acerca los documentos de los hijos de las cuatro voluntarias. Cada carta representa el hijo de una voluntaria. Un lado de la carta dice a qué escuela la voluntaria asignó a su propio hijo o hija, y el otro lado de la carta dice en qué pueblo vive el estudiante.

La mayoría de los padres quiere que sus hijos reciban la mejor educación posible, sin embargo, algunos no están dispuestos a pagar por ella. Es fácil imaginar que tus voluntarias, siendo madres ambiciosas, puedan haber estado tentadas a hacer trampa. Indica solamente aquella(s) carta(s) que tú definitivamente necesitas voltear para ver si los documentos de cualquiera de estos estudiantes viola la regla”.

Las cartas son: “Preparatoria Grover”, “pueblo de Hannover”, “Preparatoria Hannover”, “Grover City”.

## **2) Problema de permiso sin la estructura de un contrato social. (P-noCS)**

*“La secretaria que tú reemplazas en la Junta local de Educación pudo haber cometido algunos errores cuando ella procesó los documentos de los estudiantes. Es importante que*

*ciertas reglas para asignar estudiantes de varios pueblos a la escuela de distrito apropiada sean seguidas, porque las estadísticas de población que ellas proveen permiten a la junta de educación decidir cuantos maestros necesitan ser asignados a cada escuela. Si estas reglas no son seguidas, algunas escuelas podrían terminar con demasiados maestros, y otras escuelas con muy pocos.*

*Los estudiantes van a ser asignados ya sea a la preparatoria Grover o a la preparatoria Hannover.*

*Algunos estudiantes viven en el pueblo de Grover City, algunos viven en Hannover, y algunos viven en Belmont. Hay reglas que determinan a qué escuela será asignado un estudiante; la más importante de estas reglas es:*

*“Si un estudiante es asignado a la preparatoria Grover, entonces ese estudiante debe vivir en Grover City”.*

*Poco antes de retirarse, la secretaria que tú reemplazaste empezó a revisar los documentos que especifican en que pueblo viven los estudiantes, y a hacer las asignaciones de escuela de acuerdo con esta regla. Ella era una mujer dulce, pequeña, de cierta edad, quien había llegado a ser más bien distraída, y quien frecuentemente cometía errores cuando categorizaba los documentos de los estudiantes.*

*Las cartas abajo tienen información acerca de los documentos de cuatro estudiantes. Cada carta representa a un estudiante. Un lado de la carta dice a qué escuela la secretaria retirada asignó al estudiante, y el otro lado de la carta dice en que pueblo vive ese estudiante.*

*Tú sospechas que la secretaria retirada pudo haber categorizado inadvertidamente algunos de los documentos de los estudiantes incorrectamente, así que tú decides ver por ti mismo si ella violó la regla. Indica solamente aquella(s) carta(s) que tú definitivamente necesitas voltear para ver si los documentos de cualquiera de estos estudiantes viola la regla”.*

*Las cartas son: “Preparatoria Grover”, “pueblo de Hannover”, “Preparatoria Hannover”, “Grover City”*

#### Versión “preparatoria Milton”

*Los problemas son los mismos que para la versión “Grover” sólo que los términos usados fueron: “Preparatoria Milton”, “pueblo de Milton”, “Preparatoria Crandon”, “Crandon City” y “Appleton” en lugar de “Belmont”.*



## APENDICE B.

### TEORÍAS PSICOLÓGICAS DE LA INTERPRETACION.

En el campo de la psicología hay actualmente dos teorías que pretenden dar cuenta de la capacidad humana para la interpretación de otros seres humanos: la teoría de la simulación y la teoría-teoría.

La teoría de la simulación sostiene que el mecanismo subyacente a nuestra habilidad de atribuir estados intencionales y predecir la conducta de otros es una clase especial de *simulación mental* en la que nosotros nos usamos a nosotros mismos como un *modelo* de la persona a quien queremos interpretar. La manera en que opera este mecanismo, cuando de lo que se trata es de predecir la conducta de otros, es la siguiente: supóngase que Juan quiere predecir la conducta de un sujeto S. Lo que Juan hace entonces es alimentar su sistema de toma de decisiones con un conjunto “pretendido” de creencias y deseos (esto es, creencias y deseos que Juan de hecho no tiene). Estas creencias y deseos pueden ser imputados a S dado que Juan puede tener un conocimiento previo de algunas otras de las creencias y de algunos otros de los deseos de S, así como un conocimiento actual de la situación perceptual de S. Si el sistema de toma de decisiones de Juan es similar al de S, y si las pretendidas creencias y deseos con que ha sido alimentado el sistema son cercanas a las que realmente tiene S, entonces el sistema de toma de decisiones de Juan generará respuestas muy similares a las que el sistema de toma de decisiones de S genera. Juan podrá entonces predecir exitosamente la conducta de S. Es importante hacer ver que la simulación de S que Juan lleva a cabo no finaliza con la realización de la conducta en cuestión (la conducta que Juan predice que S realizará), pues, se presupone que el sistema de toma de decisiones puede, en estos casos, trabajar desconectado (off-line) de los mecanismos encargados del control de la acción.

La manera en que opera la simulación mental cuando lo que se busca es la interpretación o atribución de estados intencionales a otra persona es la siguiente: imaginemos nuevamente que Juan está atribuyendo estados intencionales a un sujeto S. Lo que Juan hace en este caso es observar la situación perceptual de S así como sus acciones. Juan entonces hipotetiza algún conjunto de creencias-deseos el cual, al ser alimentado en su propio sistema de toma de decisiones, producirá una decisión para ejecutar acciones del mismo tipo de las que S ejecuta. Seguramente ocurrirá que halla más de un conjunto de creencias-deseos que satisfaga la condición anterior y que no halla suficiente evidencia para elegir uno de estos conjuntos y excluir los otros. En estos casos, Alvin Goldman, uno de los defensores de la

teoría de la simulación, sostiene que Juan -y cualquier intérprete- debe simplemente asumir que el agente es psicológicamente muy similar a él mismo e imputarle las creencias-deseos que, desde su propia perspectiva, son más naturales.

En contraste con la teoría de la simulación, la teoría-teoría sostiene que, cuando los seres humanos interpretamos a otros y hacemos predicciones de su conducta, nosotros nos valemos de una “estructura de conocimiento” internamente representada la cual consiste de un conjunto de reglas o principios relativos a las maneras en que funciona la mente de los otros. Estas reglas o principios constituyen en sí mismos una “teoría” sobre el agente. Puesto que en general esta teoría no es accesible a la conciencia, nuestro conocimiento de ella es considerado como un conocimiento tácito.

## BIBLIOGRAFIA

- Bracewell, R. J. & Hidi, S. E. (1974): "The solution of an inferential problem as a function of stimulus materials". Quarterly Journal of Experimental Psychology, 26, pp. 480-8.
- Brown, C.; Keats, J. A.; Keats, D. M. & Seggie, I. (1980): "Reasoning about implication: A comparison of Malaysian and Australian subjects". Journal of Cross Cultural Psychology, 11, pp. 395-410.
- Cheng, P. W. & Holyoak, K. J. (1985): "Pragmatic Reasoning Schemas". Cognitive Psychology, 17, pp. 391-416.
- Cheng, P. W.; Holyoak, K. J.; Nisbett, R. E.; & Oliver, L. M. (1993): "Pragmatic versus Syntactic Approaches to Training Deductive Reasoning", In: Richard E. Nisbett (Ed.), Rules For Reasoning. Hillsdale, New Jersey, Lawrence Erlbaum Associates Publishers.
- Cherniak, Christopher; (1986): Minimal Rationality. Cambridge Massachusetts, The MIT Press.
- Clark, H. H. & Chase, W. G. (1972): "On the process of comparing sentences against pictures. Cognitive Psychology, 3, pp. 472-517.
- Cosmides, L. (1989): "The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task". Cognition, 31, pp. 187-276.
- Cosmides, L., & Tooby, J. (1989): Evolutionary psychology and the generation of culture, Part II. Case study: A computational theory of social exchange. Ethology and Sociobiology, 10, pp. 51-97.
- Cox, J. R. & Griggs, R. A. (1982): "The effects of experience on performance in Wason's selection task". Memory and Cognition, 10, pp. 496-502.
- Davidson, Donald; (1984): Inquiries into Truth and Interpretation. Oxford, Clarendon Press.
- Davidson, Donald; (1985): "Incoherence and Irrationality". Dialéctica, Vol. 39, No. 4, pp. 345-354.
- Elman, J.L.; Bates, E.A.; Johnson, M.H.; Karmiloff-Smith, A.; Parisi, D.; Plunkett, K.; (1996): Rethinking Innateness. A connectionist perspective on development. U.S.A., MIT Press, 1998.
- Eraña Lagos, María de los Angeles; (2003): Normatividad Epistémica y Diversidad Cognoscitiva. Tesis de Doctorado, UNAM, México, D.F.
- Evans, J. St. B. T. (1972): "Interpretation and matching bias in a reasoning task". British Journal of Psychology, 24, pp. 193-99.
- Evans, J. St. B. T. (1989): Bias in Human Reasoning: Causes and Consequences. Hillsdale (USA), Lawrence Erlbaum Associates Publishers.

- Fiedler, K. (1988): "The dependency of the conjunction fallacy on subtle linguistic factors". Psychological Research, 50, pp. 123-129.
- García Aguilar, Claudia Lorena (2005): "Especificidad de Dominio y Sesgo en la Psicología del Razonamiento Humano". Signos Filosóficos
- Gigerenzer, G. (1991): "On Cognitive Illusions and Rationality". En: Poznan Studies in the Philosophy of the Sciences and the Humanities. Vol. 21, pp. 225-249.
- Gigerenzer, G. (1993): "The bounded rationality of probabilistic mental models". En: Manktelow, K. I. & Over D. E. (Eds.). Rationality. Psychological and Philosophical Perspectives. London, Routledge, 1993.
- Gigerenzer, G. & Hug, K. (1992): "Domain-specific reasoning: Social contracts, cheating, and perspective change". Cognition, 43, pp. 127-171.
- Giroto, V. & Legrenzi, P. (1989): "Mental representation and hypothetico-deductive reasoning: The case of the THOG problem". Psychological Research, 51, pp. 129-135.
- Golding, E. (1981): The effect of past experience on problem solving. Paper presented at the Annual Conference of the British Psychological Society, Surrey University, April.
- Goodwin, R. Q., & Wason, P. C. (1972): "Degrees of insight". British Journal of Psychology, pp. 205-12.
- Grandy, Richard; (1973): "Reference, Meaning and Belief", Journal of Philosophy. Vol. 70, No. 14, pp. 439-452.
- Griggs, R. A. (1983): "The role of problem content in the selection task and in the THOG problem". In Jonathan Evans (Ed.), Thinking and Reasoning: Psychological Approaches. London, Routledge and Kegan Paul.
- Griggs, R. A. & Cox, J. R. (1982): "The elusive thematic-materials effect in Wason's selection task". British Journal of Psychology, 73, pp. 407-20.
- Hughes, M. A. M. (1966): The use of negative information in concept attainment. Tesis doctoral no publicada. Universidad de Londres.
- Johnson-Laird, P. N. ; Legrenzi, P. & Sonino Legrenzi, M. (1972): "Reasoning and a sense of reality". British Journal of Psychology, 63, pp. 395-400.
- Johnson-Laird, P. N., & Tagart, J. (1969): "How implication is understood". American Journal of Psychology, 82, pp. 367-73.
- Johnson, M.H., & Morton, J. (1991): Biology and Cognitive Development: The Case of Face Recognition. Oxford, Blackwell.
- Kahneman, D., & Tversky, A. (1973): "On the Psychology of Prediction". Psychological Review, 80, pp. 237-251.
- Kupfermann, I. (1991): "Genetic Determinants of Behavior". En Kandel, E. R.; Schwartz, J.H.; Jessell, T. M. (Eds.). Principles of Neural Science. U.S.A., Elsevier Science Publishing.

- Manktelow, K. I. (1979): The role of content in reasoning. Unpublished PhD thesis, Plymouth Polytechnic.
- Manktelow, K. I. & Evans, J. St. B. T. (1979): "Facilitation of reasoning by realism: Effect or non-effect?" British Journal of Psychology, 70, pp. 477-88.
- Newstead, S. E.; Griggs, R. A. & Warner, S. A. (1982): "The effects of realism on Wason's THOG problem". Psychological Research, 44, pp. 85-96.
- Quine, W.V.O. (1953): "Dos Dogmas del Empirismo". En: Desde un punto de Vista Lógico. Barcelona, Ariel, 1962.
- Quine, W.V.O. (1969): "Epistemology Naturalized". In: Ontological Relativity and Other Essays. New York, Columbia University Press.
- Stein, Edward (1996): Without Good Reason. The Rationality Debate in Philosophy and Cognitive Science. Oxford, Clarendon Press.
- Stich, Stephen P. (1990): The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation. Cambridge Massachusetts, The MIT Press.
- Tversky, A., & Kahneman, D. (1983): "Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment". Psychological Review, 90, No.4, 293-315.
- Tweney, R. D.; Doherty, M.E.; Warner, W.J. & Pliske, D.B. (1980): "Strategies of rule discovery in an inference task". Quarterly Journal of Experimental Psychology, 32, pp. 109-24.
- Van Duyne, P. C. (1974): "Realism and linguistic complexity in reasoning". British Journal of Psychology, 65, pp. 59-67.
- Wason, P. C. (1960): "On the failure to eliminate hypotheses in a conceptual task". Quarterly Journal of Experimental Psychology, 12, pp. 129-40.
- Wason, P.C. (1966): "Reasoning". In: Foss, B.M. (ed.), New Horizons in Psychology. Harmondsworth, Penguin.
- Wason, P.C. (1968): "Reasoning about a rule". Quarterly Journal of Experimental Psychology, 20, pp. 273-81.
- Wason, P.C. (1969): "Regression in reasoning?". British Journal of Psychology, 60, pp. 471-80.
- Wason, P. C. & Brooks, P. G. (1979): "THOG: The anatomy of a problem". Psychological Research, 41, pp. 79-90.
- Wason, P.C. & Johnson-Laird, P. N. (1970): "A conflict between selecting and evaluating information in an inferential task". British Journal of Psychology, 61, pp. 509-15.
- Wason, P.C. & Johnson-Laird, P. N. (1972): Psychology of Reasoning: Structure and Content. London, Batsford.

Wason, P.C. & Shapiro, D. (1971): "Natural and contrived experience in a reasoning problem". Quarterly Journal of Experimental Psychology, 23, pp. 63-71.

Yachanin, S. A. & Tweney, R. D. (1982): "The effect of thematic content on cognitive strategies in the four-card selection task". Bulletin of the Psychonomic Society, 19, pp. 87-90.