



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN FILOSOFÍA
FACULTAD DE FILOSOFÍA Y LETRAS
INSTITUTO DE INVESTIGACIONES FILOSÓFICAS**

**EMOCIONES:
HACIA EL DESCUBRIMIENTO DE UNA CLASE NATURAL**

**T E S I S
QUE PARA OPTAR POR EL GRADO DE DOCTOR EN FILOSOFÍA**

**PRESENTA
RUBÉN REYES MORENO**

COMITÉ TUTOR

**DRA. OLGA ELIZABETH HANSBERG TORRES
INSTITUTO DE INVESTIGACIONES FILOSÓFICAS**

**DRA. SALMA SAAB HASSEN
INSTITUTO DE INVESTIGACIONES FILOSÓFICAS**

**DR. JOSÉ LUIS DÍAZ GÓMEZ
DEPARTAMENTO DE HISTORIA Y FILOSOFÍA DE LA MEDICINA
FACULTAD DE MEDICINA**

MÉXICO, D.F., AGOSTO DE 2014



Universidad Nacional
Autónoma de México



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.

Para Ciri
In memoriam

AGRADECIMIENTOS.

Este trabajo versa sobre emociones, y he tenido la suerte de investigarlas, desde siempre y de primera mano, en el mejor laboratorio posible: una gran familia. Por todo lo que me enseñaron y me siguen enseñando sobre el tema, gracias. Gracias a mis padres (Efraín y Ciri), a mis ocho hermanos (Rebeca, Rafael, Ricardo, Raquel, Rodolfo, Roberto, Ime y Doris) y a sus familias por esta maravillosa travesía afectiva que llamamos vida; sin ellos, no sería posible ni tendría sentido. Gracias a Marisol y David. El hermoso relieve emocional con el que adornan mis días es algo de lo que preciso para vivir, y es lo que me impulsa para seguir. Si bien no puedo asegurar que este trabajo esté a la altura del amor, apoyo y comprensión que *toda* mi familia me ha brindado en todo momento, sí puedo decir que, al menos, el trabajo y el esfuerzo detrás de él intentan hacerlo.

En el aspecto teórico y académico, difícilmente podría haber pensado en un mejor respaldo que el que me proporcionó en los últimos años mi comité tutorial. El consejo siempre juicioso y profundamente informado de José Luis Díaz fue, sencillamente, crucial para este trabajo. Representó siempre un edificante placer escuchar sus relatos y experiencias en el estudio de las emociones desde el campo de las neurociencias. Por otro lado, la gran penetración y denuedo con el que Salma Saab examinó mis opiniones, siempre con la intención de hacer de ellas algo más respetable, sólo se parangona con las oportunas recomendaciones y el aliento que me infundía para seguir adelante. La deuda que he adquirido con Olbeth Hansberg en estos últimos años es una para la que no tengo fondos suficientes. Su trabajo, el cual comencé a conocer mientras hacía mis estudios de maestría, es la causa de éste, porque de aquél aprendí que las emociones constituían un tópico filosófico fascinante: pletórico de problemas, tesis, teorías e implicaciones. Desde entonces, me he nutrido de su vasto conocimiento sobre el tema, y ella, por su parte, me ha honrado con su guía y apoyo constante... con la seriedad y dedicación con la que ha atendido y enriquecido mis puntos de vista siempre que se lo he solicitado. Si alguien lee con detenimiento y benevolencia este trabajo notará que, de cuando en cuando, hay cosas con algún valor filosófico. En tales casos, la mayor parte del mérito debe atribuírsele a Olbeth, Salma y José Luis.

También me obliga la amable disposición que mostraron la Dra. Ana Rosa Pérez Ransanz y el Dr. Gustavo Ortiz Millán al aceptar leer mi tesis. Más notable aún fue la cele-

ridad con la que lo hicieron y las favorables opiniones que emitieron sobre la misma. Hubiera querido contar con más tiempo para enriquecer esta investigación con sus incisivas y acertadas observaciones.

Esa rara combinación de amabilidad y eficiencia que caracteriza a Vero Acatitla también merece reconocimiento. Su apoyo en los trámites requeridos por el programa de posgrado siempre fue incondicional y expedito. Creo que no hay trámite que no logre sacar adelante.

Gracias a la Universidad Nacional Autónoma de México. Nunca podré retribuirle a esta inmensa institución todo lo que me ha dado en la vida. Desde el bachillerato hasta el doctorado. De sus aulas, a sus bibliotecas y campos deportivos. Como estudiante, como jugador de futbol americano, como profesor, de cualquier modo, es un orgullo formar parte de la UNAM.

Finalmente, agradezco al CONACyT por la tranquilidad financiera que representó gozar de la beca con la que me distinguió durante cuatro años.

TABLA DE CONTENIDOS

INTRODUCCIÓN GENERAL	9
PARTE I. CLASES NATURALES: REVISANDO LA TRADICIÓN	12
I. INTRODUCCIÓN	13
II. CLASES NATURALES: DEL NOMINALISMO AL ESENCIALISMO	16
II.I. UNA CUESTIÓN PRELIMINAR: ¿POR QUÉ SEGUIR HABLANDO DE CLASES NATURALES?	16
II.II. NOMINALISMO	19
II.III. ESENCIALISMO.....	20
1. <i>KRIPKE Y PUTNAM EN CONTRA DEL NOMINALISMO: ALGUNAS DISTINCIONES</i>	20
2. <i>METAFÍSICA Y SEMÁNTICA DE LAS CLASES NATURALES</i>	25
3. <i>SEMÁNTICA ESENCIALISTA: SE</i>	25
II.IV. ESENCIALISMO, NOMINALISMO Y SEMÁNTICA ESENCIALISTA	27
1. <i>ALGUNAS DIFERENCIAS FUNDAMENTALES ENTRE EL ESENCIALISMO Y EL NOMINALISMO</i>	27
2. <i>ESENCIALISMO</i>	28
3. <i>DE LA NATURALIDAD O QUÉ SON LAS CLASES NATURALES</i>	28
4. <i>ESENCIALISMO DE CLASES NATURALES, REALISMO DE CLASES NATURALES Y SEMÁNTICA ESENCIALISTA (SE)</i>	34
5. <i>ESENCIALISMO: TEORÍAS DESCRIPTIVAS DE LA REFERENCIA Y SE</i>	36
III. IMPLICACIONES DEL ESENCIALISMO PARA EL ESTUDIO DE LAS EMOCIONES.....	38
III.I. LOS ESTADOS MENTALES EN LA MIRA DEL ESENCIALISMO	40
1. <i>LOS ARGUMENTOS DE HANSBERG</i>	41
III.II. CONCLUSIONES: UN PAR DE SUPUESTOS ESENCIALES EN EL ESENCIALISMO DE CLASES NATURALES QUE POSTULA Y CRITICA HANSBERG	55
IV. EN TORNO A LA REALIZACIÓN MÚLTIPLE.....	57
IV.I. REALIZACIÓN MÚLTIPLE: ¿QUÉ PRUEBA?	57
1. <i>¿UNA TESIS A PRIORI?</i>	57
2. <i>UN DILEMA PARA LA REALIZACIÓN MÚLTIPLE</i>	61
3. <i>IDENTIDAD FUNCIONAL Y REALIZACIÓN MÚLTIPLE</i>	64
4. <i>INTEGRANDO LAS OBSERVACIONES A LA LUZ DE LA NEUROFISIOLOGÍA</i>	67
5. <i>CODA: UNA HIPÓTESIS SOBRE LA APARENTE PLAUSIBILIDAD DE TRM</i>	70
V. CLASES NATURALES Y ‘ESTRUCTURA EXPLICATIVA’: DOS VERSIONES.....	72
V.I. LA CONSTRUCCIÓN DE UNA NOCIÓN FUERTE DE ‘ESTRUCTURA EXPLICATIVA’	72
1. <i>ALGO MÁS QUE ‘SEMÁNTICA ESENCIALISTA’ (SE)</i>	74
2. <i>IMPACTO DE LA NOCIÓN FUERTE DE ‘ESTRUCTURA EXPLICATIVA’:</i> <i>ESPECIES Y EMOCIONES</i>	76
V.II. LA CONSTRUCCIÓN DE UNA NOCIÓN MÁS LIBERAL DE ESTRUCTURA EXPLICATIVA: LA PROPUESTA DE BOYD	78
1. <i>HACIA UNA SEMÁNTICA REALISTA DE LOS TÉRMINOS DE CLASE NATURAL</i>	79
2. <i>REFERENCIA: APERTURA INDUCTIVA Y ACCESO EPISTÉMICO</i>	80
3. <i>SOBRE LA NATURALEZA DE LAS DEFINICIONES DE LOS TÉRMINOS DE CLASE NATURAL</i>	82
4. <i>UNA METAFÍSICA REALISTA DE LAS CLASES NATURALES</i>	89
5. <i>UNA CONCEPCIÓN ALTERNATIVA DE CLASE NATURAL (RPH)</i>	92

PARTE II. EMOCIONES Y CLASES NATURALES.....	96
I. INTRODUCCIÓN.....	97
II. TRES TESIS SOBRE EL ESTATUS DE LAS EMOCIONES EN TANTO QUE CLASES NATURALES	101
III. LA INTEGRACIÓN DE LOS BANDOS	104
IV. GRIFFITHS: EMOCIONES BÁSICAS COMO CLASES NATURALES	106
IV.I. NATURALEZA DE LAS EMOCIONES BÁSICAS O PROGRAMAS AFECTIVOS	106
IV.II. ARGUMENTOS, EVIDENCIA Y OBJECIONES EN TORNO A LOS PROGRAMAS AFECTIVOS.....	107
IV.III. ¿SON REALMENTE LAS EMOCIONES BÁSICAS ‘PROGRAMAS’? LA OBJECCIÓN DE LA AGREGACIÓN ARBITRARIA Y LA ALTERNATIVA COMPOSICIONALISTA.....	117
IV.IV. RESPUESTAS A LA OBJECCIÓN DE LA FALTA DE CORRELACIÓN	118
IV.V. RESPUESTA AL RETO COMPOSICIONALISTA: LA DIVERSIDAD DE LOS ESTÍMULOS.....	118
IV.VI. OTRA RESPUESTA: LA MODERADA EFICIENCIA Y ESCASA FLEXIBILIDAD DE LOS PROGRAMAS AFECTIVOS.....	120
IV.VII. LA MODULARIDAD DE LAS EMOCIONES BÁSICAS	121
IV.VIII. PROGRAMAS AFECTIVOS COMO CLASES NATURALES RPH: HOMEOSTASIS Y PROYECTABILIDAD	122
IV.IX. INDETERMINACIÓN Y CARÁCTER DINÁMICO DE LOS PROGRAMAS AFECTIVOS.....	123
V. ARGUMENTOS Y OBJECIONES EN CONTRA DE UNA TEORÍA DE LA EMOCIÓN	125
V.I. EL ARGUMENTO DE LA DESUNIÓN DE LAS EMOCIONES	125
V.II. MÁS OBJECIONES A LA POSIBILIDAD DE UNA TEORÍA GENERAL DE LAS EMOCIONES	128
V.III. RESPUESTAS A ALGUNAS OBJECIONES	130
VI. UN MODELO PARA UNA TEORÍA GENERAL DE LAS EMOCIONES	134
VI.I. MARCO CONCEPTUAL	135
1. <i>CLASES NATURALES DE RACIMOS DE PROPIEDADES HOMEOSTÁTICAS</i>	135
2. <i>PROPIEDADES ‘FÍSICAS’ RELEVANTES</i>	137
VI.II. METODOLOGÍA: SIGUIENDO EL CAMINO DE LA REFERENCIA	146
VI.III. TEORÍA: LA COMPLEJA ESTRUCTURA EXPLICATIVA DE LAS EMOCIONES	150
VII. UNA TEORÍA GENERAL DE LAS EMOCIONES	157
VII.I. NATURALEZA DE LAS EMOCIONES	157
I. <i>PROPIEDADES NEUROFISIOLÓGICAS Y EXPRESIVO/CONDUCTUALES</i>	157
II. <i>PROPIEDADES BIOLÓGICAS</i>	176
III. <i>PROPIEDADES PSICOLÓGICO/FUNCIONALES</i>	185
IV. <i>PROPIEDADES O RASGOS DE SU ARQUITECTURA</i>	192
V. <i>PROPIEDADES FENOMENOLÓGICAS</i>	202
VII.II. ¿QUÉ ES UNA EMOCIÓN?	208
VIII. OBSERVACIONES FINALES SOBRE LA TEORÍA.....	212
VIII.I. PONIENDO A PRUEBA LA TEORÍA	212
VIII.II. EXPANSIÓN Y CONTRACCIÓN DE LA ONTOLOGÍA EMOCIONAL.....	218
CONCLUSIONES	223
REFERENCIAS	227

INTRODUCCIÓN GENERAL

En los últimos veinte años las emociones han recibido mucha atención por parte de la comunidad científica en general; desde psicólogos y neurólogos, hasta antropólogos, filósofos, economistas y políticos, han abonado el ahora fértil terreno de la investigación de los fenómenos afectivos. Este redescubrimiento de las emociones ha traído consigo un conjunto bastante nutrido de teorías, tesis, reinterpretaciones, opiniones y nuevas valoraciones acerca de la naturaleza y función de las emociones. Esta investigación no pretende elaborar un mapa detallado de la nueva y expandida provincia emocional. Si acaso se hace esto, será de forma parcial e incidental. Nos enfocaremos sólo en un expediente derivado de esta nueva geografía de las emociones. Recientemente, diferentes autores han ofrecido diversos argumentos en contra de la idea de que las emociones conforman una clase natural¹. El objetivo de esta investigación consiste en identificar las razones detrás de esta postura. Las razones, por nuestra parte, para llevar a cabo tal empresa, apuntan a la posibilidad de mantener unido, en un sólo dominio explicativo y ontológico a las emociones. En breve, intentaremos hacer lo posible por defender la tesis de que las emociones conforman una sola clase natural.

Con tal objetivo presente, se ha estructurado este trabajo en dos partes con dos propósitos bien definidos. En la primera parte, **Clases Naturales: Revisando la Tradición**, abordaremos una cuestión que, desde nuestra perspectiva, es crucial para decidir el estatus de las emociones en tanto que clases naturales. Se trata de cuestión de cómo se entiende el concepto de ‘clase natural’. Consideramos que el tipo de concepción de ‘clase natural’ que se adopte, es fundamental para decir si las emociones constituyen o no una clase natural. En este sentido, en primer lugar, se intentará hacer evidente que cierta concepción de ‘clase natural’, una que la define de una forma demasiado estricta —en términos de esencias bien delimitadas y estáticas²—, es la responsable de que muchos investigadores —filósofos principalmente— lleguen a concluir que las emociones no conforman una clase natural. En segundo lugar, y en un sentido positivo, expondremos y ofreceremos razones para justificar

¹ Dentro de los varios autores que se pueden incluir en esta lista, el caso de y los argumentos de Griffiths (1997) serán los que con mayor detalle se intentarán contestar en este trabajo.

² *Delimitadas* porque involucran grupos exhaustivos o bien definidos de propiedades. Y *estáticas* porque supone que dichos grupos de propiedades no cambian a través del tiempo. Más especificaciones sobre esta concepción de clase natural y sus implicaciones se ofrecen en I, III.

la adopción de una noción alternativa de ‘clase natural’ mucho más liberal. Una noción que conceda que las clases naturales están constituidas por racimos más bien *indeterminados* y *dinámicos* de propiedades *físicas* que constituyen sus esencias o, como les llamaremos nosotros, estructuras explicativas. Un rasgo crucial de la noción de ‘clase natural’ en cuestión, consiste en poner énfasis en que es la indagación empírica la que decide, no sólo la cuestión de qué fenómenos constituyen una clase natural, sino también la cuestión de qué es una clase natural. La idea es que nuestra noción de ‘clase natural’ debe estructurarse a partir de la forma en la que las buenas teorías, leyes y explicaciones dicen que es la realidad —indeterminada y dinámica—; y no a partir de cómo las idealizaciones filosófico-semánticas dicen que debe ser la realidad —bien delimitada y estática. En resumen, según la noción de ‘clase natural’ que adoptaremos, resultará necesario abandonar la idea de que las clases naturales involucran esencias o estructuras explicativas constituidas de grupos de propiedades bien delimitados y estáticos, a favor de una noción que las entienda como indeterminadas y dinámicas. Este es, de forma compendiosa, el itinerario de la primera parte de este trabajo.

En la segunda parte, **Emociones y Clases Naturales**, teniendo como fundamento la noción de clase natural que se articuló en la primera, nos daremos a la tarea de argumentar a favor y construir un marco conceptual para una teoría de las emociones que permita incluirlas a todas —sean básicas o no— en un solo dominio explicativo y ontológico. Es decir, una teoría en términos de la cual sea posible sostener que las emociones conforman una sola clase natural. El esquema general de dicha teoría es el siguiente. Las emociones conforman una clase natural porque remiten a una esencia o estructura explicativa, es decir, a un racimo indeterminado y dinámico de propiedades físicas vinculadas homeostática y causalmente. La identificación de tal estructura explicativa, instanciada —dentro de ciertos límites de indeterminación— por casos particulares de emociones, apunta a la existencia de un dominio explicativo y ontológico, desde el cual son posibles explicaciones y predicciones interesantes del fenómeno emocional. Las diez propiedades físicas relevantes que configuran, aproximadamente, la estructura explicativa de las emociones, son: (1) Propiedades neuronales. (2) Propiedades fisiológicas. (3) Propiedades expresivo-conductuales. (4) Propiedades adaptativas. (5) Propiedades representacionales. (6) Efectos

psicológicos característicos. (7) Propiedades o rasgos de arquitectura. (8) Sistema de recalibración. (9) Duración breve. (10) Fenomenología.

El título de esta investigación es, *Emociones: Hacia el Descubrimiento de una Clase Natural*, de alcanzar los objetivos que se plantean en cada parte de la misma —en los términos que se han expuesto— podría clarificarse el subtítulo. No sólo nos interesa mostrar que es posible concebir una teoría en términos de la cual podríamos llegar a considerar que las emociones, en general, constituyen una clase natural. Nos interesa poner énfasis en un aspecto metodológico: en la medida en que la noción de ‘clase natural’ que adoptamos se articula en función de lo que las buenas teorías empíricas dicen que es la realidad —indeterminada y dinámica— y ya que las propiedades que conforman la estructura explicativa de las emociones son *físicas*³, resulta evidente que la decisión de si algo es o no una clase natural —en nuestro caso particular, las emociones— tiene que elucidarse, principalmente, por medio de la investigación empírica. Si las emociones son clases naturales, hay que *descubrirlo* empíricamente. Este trabajo intentará mostrar que quizás es demasiado prematura la afirmación que niega que tal descubrimiento sea posible.

³ O, al menos investigables desde una perspectiva empírica-fisicalista.

PARTE I

CLASES NATURALES:

REVISANDO LA TRADICIÓN

I. INTRODUCCIÓN

Uno de los objetivos —rara vez explícito, aunque indiscutiblemente implícito— de la empresa científica en general, es el de proveer un inventario del mobiliario del mundo. Pero lo que hay en el mundo puede ser clasificado de muchas formas diferentes. Existen, por ejemplo, las clases de cosas de las que hablan los distintos tipos de científicos (biológicas, químicas, geológicas); así como los grupos en los cuales esas cosas son, a su vez, agrupadas (mamíferos, metales, rocas ígneas). Pero, ¿qué le confiere el aire de excelencia a las clasificaciones científicas? ¿Qué diferencia existe entre las taxonomías que proponen las ciencias y la que adscribe, por ejemplo, Jorge Luis Borges a un diccionario chino de título *Emporio celestial de conocimientos benévolos*:

En sus remotas páginas está escrito que los animales se dividen en (a) pertenecientes al Emperador, (b) embalsamados, (c) amaestrados, (d) lechones, (e) sirenas, (f) fabulosos, (g) perros sueltos, (h) incluidos en esta clasificación, (i) que se agitan como locos, (j) innumerables, (k) dibujados con un pincel finísimo de pelo de camello, (l) etcétera, (m) que acaban de romper el jarrón, (n) que de lejos parecen moscas.¹

Por lo menos podemos pretender encontrar las siguientes diferencias. Las clasificaciones de la ciencia no se superponen, como (a) y (c) podrían hacerlo. A diferencia de (h) y (l), no dan lugar a problemas lógicos. Y, finalmente, en oposición a (e) y (f), las clasificaciones científicas no incluyen seres fantásticos. Esta lista de diferencias es bastante discutible y poco explicativa. Solo enumera defectos que, supuestamente, las clasificaciones científicas no tienen o aspiran a no tener. Una respuesta de carácter positivo puede concebirse al pensar que, mientras la clasificación de Borges es *arbitraria*, la taxonomía de las ciencias agrupa cosas en *clases naturales*. La apelación a *clases naturales* será tan satisfactoria como lo sea la caracterización de dicha noción. ¡Pero, ahí viene el cartero con malas noticias! Y nos informa que la noción emplazada (*clase natural*) para fundamentar la egregia estirpe de las taxonomías científicas es sospechosa —y para muchos, ciertamente culpable— de un buen número de felonías filosóficas. El caso en contra de la noción de *clase natural* se abre a partir de los siguientes cargos: **a**) se le acusa de implicar una pesada metafísica realista de cuestionable reputación (Dupré, 2001; Hacking, 1991); **b**) la noble-

¹ Borges, “El idioma analítico de John Wilkins”: <http://ndirty.cute.fi/%7Ekarttu/tekstit/wilkins.htm>

za adverbial de este concepto —“natural”— encubre formas de clasificación que no son tan naturales o poco arbitrarias en sí mismas (Goodman, 1978; Hacking, 1991; Kukla, 2000); **c**) es inestable y promiscua —se entiende de diferentes formas en diferentes contextos disciplinarios— (Dupré, 1993); **d**) ninguna de dichas concepciones parece ser capaz de proveer un marco clasificatorio adecuado para entidades teóricas tan fundamentales como las especies (Ghiselin, 1978; Hull, 1978), o capacidades cognitivas o propiedades mentales como las emociones (Hansberg, 1990). **e**) Finalmente, se ha llegado a cuestionar la utilidad y valor metodológico, epistemológico, ontológico y metafísico de la noción de *clase natural* (Dupré, 1993).

En esta primera parte realizaremos un rápido recorrido histórico, pero también crítico, de la noción de clase natural. El objetivo general es llegar a un diagnóstico que sea útil para desarrollar una concepción de *clase natural* que pueda sortear el conjunto de objeciones que se le presentan y, principalmente, que proporcione un marco clasificatorio adecuado para estados psicológicos como las emociones. Para tener una idea más precisa de cómo se pretende satisfacer el objetivo anterior, hablaremos con un poco más de detalle de lo que versará cada uno de los capítulos de esta primera parte.

El primer capítulo comienza con un apartado que pretende mostrar que la noción de ‘clase natural’ sigue siendo fundamental en las pesquisas que realizan tanto científicos como filósofos en relación al estatus ontológico de las entidades con las que se comprometen: leyes, fenómenos, procesos, entidades, sustancias, organismo, etc. Básicamente, este apartado intenta hacer frente al planteamiento (**e**). En él se expone de forma bastante general la historia del nominalismo y sus tesis más importantes. Luego se realizan las siguientes distinciones en torno al esencialismo. En primer lugar, se distinguen las propuestas semántico-metafísicas de Kripke y Putnam en relación a las clases naturales. Se reconoce que comparten algunas tesis generales en relación a la naturaleza histórico-causal de la referencia, pero que puede diferir en torno a cuestiones semánticas y metafísicas de más detalle. En segundo lugar se distingue entre la cuestión semántica sobre el significado y referencia de los términos de clase natural, de la cuestión metafísica de qué son, si existen, las clases naturales. En tercer lugar, se ofrece una caracterización de lo que llamaremos Semántica Esencialista (SE), la cual es un tipo de enfoque filosófico sobre el significado y referencia de los términos de clase natural que abrió la puerta al esencialismo de clases naturales.

El apartado II.IV inicia identificando las diferencias de fondo entre el nominalismo y el esencialismo —cierta interpretación—, y cómo el último parece estar conceptualmente atado a la semántica esencialista. Luego se discute la cuestión que plantea **(b)**, llegando a la conclusión de que lo distintivo de una clase natural, más que el carácter natural, es que involucra una estructura explicativa: esto es, que remite a un conjunto de propiedades físicas relevantes que la caracterizan y cuyo descubrimiento es posible sólo *a posteriori*. En este apartado también se realizan algunos deslindes en torno a distintas formas de metafísica de las clases naturales que parecen asociarse con la semántica esencialista: el *esencialismo metafísico de clases naturales* y el *realismo de clases naturales*. El primero, que asociamos con la postura de Kripke (1981) —y al que bien puede dirigirse el planteamiento **(a)**— sostiene que las estructuras explicativas de las clases naturales consisten en propiedades físicas relevantes que se descubren *a posteriori* y que sirven para establecer condiciones necesarias y suficientes de pertenencia en todos los mundos posibles. El *realismo*, por su parte, sugiere básicamente lo mismo, excepto que no postula la metafísica de los mundos posibles. Las estructuras explicativas sirven para identificar clases naturales, e instancias de sus miembros, sólo en este mundo.

En el capítulo III se exponen algunas de las implicaciones que el esencialismo metafísico de clases naturales tiene para el estudio de las emociones. Al analizar los argumentos que expone Hansberg (1990) en contra de la idea de que los estados mentales y, por lo tanto, las emociones sean clases naturales, identificamos dos supuestos que son cruciales para el éxito de dichos argumentos. Uno es cierta construcción fuerte de la noción de ‘estructura explicativa’, según la cual una estructura explicativa de una clase natural se debe entender como un conjunto *bien delimitado* (exhaustivo) y *estático* (que no cambia a través del tiempo) de propiedades físicas relevantes que existe en todos los mundos posibles en los que existe dicha clase natural. Otro supuesto preeminente en los argumentos en contra de que los estados mentales y las emociones sean clases naturales es el de la Tesis de la Realización Múltiple (TRM), según la cual —por lo menos su versión más fuerte— todo estado mental es susceptible de ser realizado por distintas bases físicas.

En el capítulo IV, nos desviamos un poco de la discusión sobre clases naturales para hacer una revisión crítica de algunos supuestos cuestionables en la idea de Realización Múltiple (RM) y la tesis que lleva su nombre. Básicamente, señalamos que las nociones en las cuales descansan, tanto el supuesto fenómeno de la RM y la TRM, nunca se especifican lo suficiente. En este senti-

do, muy pocas veces se aclara la cuestión de ‘cuándo se trata de un realizador *distinto*’ de un mismo estado o fenómeno. O, por otro lado, no se especifica ‘cuándo sabemos que se trata de un *mismo* estado o fenómeno’ múltiplemente realizado. Regularmente, se detecta cierta asimetría en la forma de determinar cada cuestión. Por un lado, es habitual que cualquier diferencia mínima en los realizadores sea suficiente para hablar de multiplicidad, mientras que, para fijar la identidad del estado o fenómeno realizado, se suele apelar a tipificaciones mucho más abiertas. La falta de especificidad y acuerdo en torno a lo que se pueda considerar como un caso de RM, concluimos, nos debe llevar a tomar con más cautela los argumentos basados en RM y TRM.

En la capítulo V se discuten las implicaciones que tiene una construcción fuerte de estructura explicativa —basada en SE y en el esencialismo metafísico de clases naturales— para cuestiones como la de si las especies y las emociones pueden ser consideradas clases naturales —esto es, la cuestión que plantea (d). Se sugiere que dicha concepción fuerte es responsable, en gran parte, de las posturas que niegan que las especies y emociones se consideren clases naturales. En este sentido, se propone una semántica y metafísica realista de las clases naturales, basada en las ideas de Boyd, que pudiera generalizarse a todas las disciplinas particulares —resolviendo el problema que plantea (c)— y que, mejor aún, pudiera ser útil para generar un marco conceptual apropiado para considerar que las especies y, principalmente, las emociones puedan considerarse, después de todo, clases naturales —con lo que se lograría salvar las dificultades que señala (d).

II. CLASES NATURALES: DEL NOMINALISMO AL ESENCIALISMO

II.1. UNA CUESTIÓN PRELIMINAR: ¿POR QUÉ SEGUIR HABLANDO DE CLASES NATURALES?

Antes de comenzar el estudio en torno a las distintas concepciones de *clase natural*, dedicaremos algunas reflexiones a las críticas que se han expuesto en torno a la utilidad y valor de la noción de ‘clase natural’. Uno de los principales críticos de la noción de ‘clase natural’ es John Dupré (1993). Este autor ha declarado el carácter obsoleto del término técnico ‘clase natural’ en virtud de que no es una noción que juegue un papel importante en la construcción y desarrollo de teorías científicas respetables. En pocas palabras, los científicos (de ninguna rama o disciplina) nos hablan de ‘clases naturales’. También señala (en un tono muy parecido al de Quine, 1969/1975) que no es de mucha utilidad preguntarse si las categorizaciones implícitas en las distintas teorías científicas remiten a clases naturales. Y no tienen caso

cuestiones de esta guisa porque las categorías o entidades a las que aluden las teorías científicas apuntan ya a los *compromisos ontológicos* que asumen los hombres de ciencia que las suscriben. Luego, es irrelevante preguntar si dichas categorías son o no clases naturales. Al figurar en teorías científicas bien establecidas y aceptadas por la comunidad científica, gozan ya de una preeminencia ontológica difícilmente disputable. ¿Para qué embrollar más la cuestión filtrando, innecesariamente, la metafísica de las clases naturales?

En relación a la primera observación de Dupré, aquella que apunta al hecho de que los científicos nunca o rara vez hablan de ‘clases naturales’, hay que decir que está estadísticamente justificada. En efecto, los científicos rara vez hablan de las categorías teóricas que suscriben, en términos de clases naturales. Sin embargo, ocasionalmente lo hacen. Entender cuándo y por qué lo hacen puede ayudar a reivindicar el uso, valor y vigencia de la noción de ‘clase natural’. Aunque no siempre, los científicos llegan a preguntarse por el estatus, en tanto que clase natural, de las categorías teóricas que resultan de su interés. Lo hacen cuando no es claro el estatus ontológico de las categorías en cuestión. Por ejemplo, la controversia entre biólogos en torno al rango ontológico de las especies (si conforman clases o individuos) ha discurrido en términos de si las especies son clases naturales o no (Hull, 1978; Ghiselin, 1978; de Queiroz, 1995; Mayr, 1995). LaPorte (2004) documenta un caso muy similar que afectó a la comunidad de químicos, quienes, a partir precisamente del célebre ejemplo putnamiano del jade, iniciaron una discusión acerca de la historia y naturaleza de los sistemas de clasificación de sustancias en la química. Es importante subrayar que tal discusión se daba, constantemente, en términos de si la jadeíta, la nefrita y/o el jade serían o no una o varias clases naturales. Otro ejemplo relevante se presenta precisamente en nuestra vecindad. Psicólogos (Barrett, 1996; Russell, 2006), antropólogos (Ekman, 1992, 1999, 2003), científicos cognitivos (Evans, 2005) y neurofisiólogos (Rolls, 2005; Panksepp; 1998, 2000; Damasio, 1994, 2000) han labrado una vasta parcela de discusión en torno a la cuestión de si las emociones son o no clases naturales. Los científicos conocen la noción de clase natural, y la usan cuando no queda claro el estatus ontológico de una categoría (especie, sustancia, emoción, etc.) relevante, en términos teóricos, en su campo.

Ahora bien —y en esto radicará la vigencia y valor de la noción de clase natural— cuando los hombres de ciencia echan mano del término ‘clase natural’, lo hacen *porque* se

plantean una cuestión filosófica, acerca del estatus ontológico y metafísico de una categoría que juega un papel relevante en su disciplina. Lo que esto significa es que existe aún cierta incertidumbre en una comunidad de investigadores en torno a la naturaleza del *compromiso ontológico* que asumirán en relación a dicha categoría. A las ciencias y los hombres que las practican no les son indiferentes las cuestiones ontológicas y metafísicas asociadas con las clases naturales. De hecho, como se ha ilustrado con los ejemplos aludidos, resultan tan acuciantes para los científicos cuestiones como la de si un tipo de entidad o fenómeno existe independientemente de nuestras clasificaciones, o como la de si los sistemas de clasificación son adecuados, naturales o arbitrarios, que no reparan en utilizar la noción de ‘clase natural’ para organizar las discusiones y poner a prueba sus intuiciones. Los científicos llegan a plantearse y discutir cuestiones de tipo ontológico sobre sus clasificaciones, en términos de clases naturales, porque en ello les va el rango ontológico y metafísico de las teorías y categorías con las se comprometen o esperan comprometerse. La noción de clase natural, con la metafísica que implica —la idea de que la naturaleza funciona y manufactura una rica y variada realidad, de manera más o menos organizada y uniforme o, por lo menos, que así la concebimos—, es crucial para resolver la naturaleza del compromiso ontológico que adquieren o no, los científicos, con las categorías que postulan. Esto es, reconocer si el compromiso ontológico es de tipo nominalista, instrumentalista o realista, o si renunciarán a la categoría y a las teorías en las que figura por no reunir los requisitos para ser considerada una clase natural, en cuyo caso no tendría lugar ningún compromiso ontológico, sino una eliminación. En la medida en que a las ciencias y los hombres de ciencia les ha resultado relevante reconocer la naturaleza de los compromisos ontológicos que adquieren al postular una categoría y un cuerpo de teorías sobre ésta, les ha sido de mucha utilidad y valor apelar a la noción de clase natural para tipificar la naturaleza de dichos compromisos, o para validarlos.

En resumen, el uso, valor y vigencia de la noción de ‘clase natural’ se reivindica en la medida en que es crucial para definir el estatus ontológico y metafísico de las clasificaciones, categorías y teorías con los que se compromete la ciencia y los científicos. En contra de lo que piensa Dupré, consideramos que la metafísica que infiltra la noción de clase natural en algunas discusiones científicas, no es superflua, ni innecesaria. Es crucial para elucidar el estatus metafísico y ontológico de las categorizaciones científicas. Seguir hablando de clases naturales se justifica, entonces, porque ayuda a construir un marco conceptual adecuado para

reflexionar y discutir en torno al estatus ontológico y metafísico de ciertas clasificaciones, categorías y teorías científicas. En nuestro caso particular, para investigar la naturaleza y ontología de las emociones, así como la metafísica implicada en una teoría y/o metodología para el estudio de las emociones.

II.II. NOMINALISMO

El término ‘clase natural’ es relativamente nuevo, la noción que le subyace tiene mucho más abo-
lengo. Russell introdujo el término en el léxico filosófico en su obra de 1948, *El conocimiento humano*. No obstante, como sugiere Hacking (1991/2007), debe reconocerse que la reflexión en torno a cuestiones relacionadas con la noción de *clase natural* ya está presente en *cierta tradición nominalista* que se remonta a autores como, Locke, Leibniz, Mill y Peirce. Nuestro recuento sobre el concepto de *clase natural* elige el mismo punto arbitrario de partida: el origen *nominalista*. Tal origen se remonta a Locke, quien distingue entre ‘esencia nominal’ y ‘esencia real’ de las cosas. Como la última nos es desconocida, nuestras categorizaciones se realizan a partir de esencias nominales, las cuales se definen como familias de propiedades que encontramos frecuentemente co-ocurriendo en el Mundo. Esta unión recurrente de propiedades se termina de sellar por medio de un término (‘nombre’ es la palabra que utiliza Locke). Por ejemplo, primero encontramos que las propiedades de ser dorado, sólido y soluble en agua regia (esencia nominal) se hayan recurrentemente co-ocurriendo en ciertas muestras de sustancias en el Mundo. Enseguida, asignamos el término ‘oro’ para esta familia de propiedades co-ocurrentes; y, finalmente, llamaremos ‘oro’ a toda muestra que satisfaga las propiedades reconocidas en la esencia nominal que designamos con el término ‘oro’. Aunque Locke reconoce la posibilidad de que las propiedades que constituyen la esencia nominal de una clase de cosas está determinada por su esencia real (micro-estructura, composición química, forma explicativa, etc.), apunta de inmediato que esto no pasa de ser una suposición que, aunque plausible, no ayuda en nada para fijar la extensión de los términos de clase natural, puesto que desconocemos dicha esencia real. La conclusión de Locke es, entonces, que lo que determina la extensión de los términos de clase natural es la esencia nominal: el conjunto de propiedades que vinculamos, por medio de un término o nombre, con una clase de cosas.²

² Cfr. Locke, (1690/1999, libro III, cap. VI).

El nominalismo lockeano entraña lo que puede designarse como **convencionalismo *a priori***. Convencionalismo en tanto que las clasificaciones de las cosas a partir de ‘esencias nominales’, tiene más que ver con las propiedades *que los seres humanos reconocen* en las cosas, que con las propiedades que en realidad tienen. *A priori* en virtud de que las definiciones de estos términos [sus condiciones de aplicación] se estipulan *a priori*: a partir de lo que un usuario competente promedio podría determinar como las características definitorias de la clase de objeto a la cual hace referencia el término. En otras palabras, la determinación del significado de un nombre o término de clase natural es un acuerdo o convenio lingüístico en una comunidad que relaciona *a priori* dicho nombre con un grupo *familiar* de propiedades (esencia nominal).

Es importante señalar que esta tradición empirista sobre las clases naturales ya presenta un rasgo interesante que sobrevivirá en las concepciones más recientes. Es parte de esta concepción lockeana de las clases naturales una alta valoración de la ‘definición en términos de *necesidad y suficiencia*’. Es decir, para Locke, el que un objeto obtenga la membrecía dentro de una clase natural exige previa satisfacción de condiciones *necesarias y suficientes*, las cuales están especificadas en la definición (esencia nominal) de la clase en cuestión.

II.III. ESENCIALISMO

I. KRIPKE Y PUTNAM EN CONTRA DEL NOMINALISMO: ALGUNAS DISTINCIONES

La reacción crítica en contra del nominalismo lockeano se horneó a fuego lento... tomó algunos siglos.³ Fue hasta que aparecieron, en la segunda mitad del siglo XX, los análisis que tanto Kripke (1980/1995) como Putnam (1975/1984) ofrecieron en torno a las nociones de ‘significado’ y ‘referencia’, que tuvo lugar una reconsideración crítica del nominalismo y nació un enfoque distinto y novedoso para pensar la semántica de los términos de clase natural y la metafísica de las clases naturales como tales. Aunque es hasta cierto punto habitual⁴ subsumir las teorías de la referencia de ambos autores bajo la misma rúbrica —a saber, como teorías causales de la referencia—, en virtud de ciertas similitudes programáticas, es importante mencionar que sus enfoques

³ La tradición nominalista en clases naturales, como señala acertadamente Hacking (1991/2007), no careció de seguidores durante todos estos años. Hacking cuenta por lo menos a cinco insignes filósofos decantándose hacia el nominalismo: Mill, Peirce, Russell, Quine y Goodman.

⁴ Autores como Mellor (1977), Donellan (1983), Soames (2002) y Stanford y Kitcher (2000), han llegado a hablar de una teoría de la referencia compartida por Kripke y Putnam.

tanto semánticos como metafísicos, en torno a las clases naturales, pueden llegar a ser sustancialmente distintos.

Vale la pena dedicar un tiempo para identificar los aspectos programáticos que comparten Kripke y Putnam, y los aspectos de detalle que distancian las propuestas de ambos autores en relación a las clases naturales. Podríamos hablar de cuatro ideas compartidas. En primer, niegan que el significado de un *Término de Clase Natural* (TCN) tenga que ver con descripciones que satisfacen sólo los miembros de una clase natural o con racimos de propiedades: idea con la cual están profundamente comprometidos los nominalistas *a la* Locke. En segundo lugar, ambos parecen conceder que la extensión es un elemento crucial del significado de los TCN. Todas las muestras de agua y todos los tigres son, en parte, lo que significan los términos “agua” y “tigre”. Pero la extensión no está dada por el significado (como lo parece suponer Frege). Más bien, la extensión define (parcialmente) el significado. En tercer lugar, tanto Kripke como Putnam reconocen el carácter histórico/colectivo que sustenta el fenómeno de la referencia. La preservación del vínculo que une a un TCN y su referente es posible gracias a que los hablantes, usuarios de un TCN, preservan, a lo largo del tiempo, y transmiten, de generación en generación, el vínculo entre el TCN y su referente. Este rasgo común en las propuestas de ambos autores es lo que ha propiciado que se llame ‘teorías causales de la referencia’ a sus propuestas. Aunque hay que acotar que Putnam ha preferido referirse a su teoría de la referencia como teoría histórica o colectiva de la referencia (Hacking, 2007). Finalmente, ambos autores están de acuerdo en que la investigación empírica es crucial para la especificación de la extensión de los TCN. Cumple agregar que es mayor el énfasis que Putnam pone en la investigación empírica. Esto se refleja en la apelación e importancia que le asigna a la *división del trabajo lingüístico*, la cual apunta a la investigación empírica desarrollada por expertos (científicos) a lo largo del tiempo, y gracias a la cual se precisa la extensión de los TCN. Kripke, por su parte, no le concede muchas reflexiones a la forma en la que tiene lugar y se desarrolla la investigación empírica que resultará esencial para elucidar la extensión de un TCN. Kripke delega el peso de la especificación de la extensión a la *rigidez* de los TCN. Un TCN selecciona a sustancias, organismos, etc., que comparten una *esencia* en todos los mundos posibles.⁵

⁵ Si bien es cierto que Putnam (1975/1984) reconoce que la noción de ‘designador rígido’ puede ayudar a ilustrar las limitaciones de las teorías descriptivas de la referencia, no parece comprometerse con las implicaciones metafísicas (la posibilidad de identidades necesarias *a posteriori*) que Kripke deriva de dicha noción. Hacking (2007), por ejem-

En cuanto a las diferencias, pueden señalarse al menos las siguientes. A primera vista, podría pensarse que otro lugar común entre las teorías de Putnam y Kripke tiene que ver con que el primero habla de especificación deíctica (decir ‘*esto* es agua’ frente a una muestra paradigmática de agua) y el segundo de un acto de bautismo originario (nombrar, por primera vez, ‘agua’ a una muestra paradigmática de agua). Para ambos autores, el fenómeno de la referencia, que involucra un TCN, se funda en estos tipos de procedimientos. Pero la similitud entre ambos procesos y aproximaciones es, de nuevo, limitada. El acto de bautismo kripkeano, a partir del cual se fija la referencia en todos los mundos posibles de un TCN —rigidez—, bien visto, es una especie de idealización cruda de cómo surge el fenómeno de la referencia. Putnam, por su parte, es un poco menos vago. La especificación deíctica de la referencia de un TCN, se apoya, a su vez, en la noción de *estereotipo*. Un *estereotipo* alude a un conocimiento colectivo mínimo —posiblemente equivocado— de las propiedades que se asocian con el referente del TCN. Es decir, el *estereotipo* hace posible, en primer lugar y quizás de una forma imprecisa, la especificación deíctica y el fenómeno de la referencia.

Otro aspecto en el cual difieren Putnam y Kripke consiste en que el primero rara vez habla de esencias. Putnam (1975/1984; 1978, 1990), más bien, alude a estructuras ocultas o de microestructuras física, cuyo descubrimiento empírico ayuda a precisar la extensión y el significado de los TCN. Por su parte, Kripke (1981/1995) no repara en aludir a las esencias como aquello a lo cual refieren los nombres propios y TCN —en los mundos posibles en los que existan. Si bien es cierto que Kripke identifica las esencias con propiedades micro-físicas de las cosas, descubiertas o descubribles empíricamente, por lo que no diría en el fondo algo muy distinto de lo que Putnam dice, se trata de una cuestión de énfasis que para muchos tiene sus implicaciones metafísicas. Evidentemente, Putnam (1975/1984) no quiere ser considerado como un *esencialista*, sino como un *realista* que hace frente a los retos que planteó en los años sesentas la caracterización del conocimiento científico que desarrolló Thomas Khun (1962). Kripke, por su parte, no parece reparar tanto en el peso de los apelativos, quizás porque tampoco teme asumir las consecuencias metafísicas de su semántica, lo cual, como veremos enseguida, representa una diferencia más entre sus ideas y las de Putnam.

plo, considera que la alusión a designadores rígidos, por parte de Putnam, no es más que un gesto de respeto hacia el trabajo de Kripke.

Una diferencia crucial entre los enfoques de Kripke y Putnam consiste en que el primero adopta una noción de *necesidad metafísica* que es mucho más fuerte que la *necesidad física* hacia la que parece decantarse Putnam. Las identidades teoréticas de Kripke (v. gr. Agua = H₂O) son enunciados de identidad *metafísicamente necesarios* (es metafísicamente imposible que el agua no sea H₂O), que sólo se pueden conocer empíricamente. Una vez que se ha fijado la referencia de un TCN como ‘agua’ —por medio del bautismo inicial frente a una muestra paradigmática— nada en el mundo actual o en cualquier otro mundo posible que tenga las propiedades superficiales del agua, y que no tenga la composición química subyacente, será agua o podrá ser llamada así de forma correcta. La *necesidad metafísica* entendida en estos términos parece ser una consecuencia de la *rigidez referencial* que caracteriza a los TCN en la semántica de Kripke (LaPorte, 2004; Beebe y Sabbarton-Leary, 2010). La necesidad metafísica con la cual se compromete Kripke, también es más compatible con una postura esencialista metafísicamente robusta, según la cual las clases naturales se caracterizan por tener esencias: conjuntos de propiedades que dictan condiciones de necesidad y suficiencia para determinar la extensión de la clase (Ellis, 2001). Por su parte, la semántica de Putnam no se basa realmente ni en la rigidez referencial ni en ninguna apelación a mundos posibles, lo cual es fundamental para que Putnam no se comprometa con una noción de necesidad tan fuerte como la que suscribe Kripke. El famoso experimento mental de la tierra gemela apunta a una situación hipotética (algo que podría tener lugar en el mundo en el que vivimos), no contrafáctica (algo que sería posible en otro u otros mundos posibles). En este sentido, cuando Putnam insiste en que lo que llaman ‘agua’ los habitantes de la tierra gemela no es realmente agua, no está implicando que el agua tenga que ser en todos los mundos posibles H₂O. Por lo tanto, para Putnam, ni el enunciado de identidad ‘Agua = H₂O’ instancia un caso de *necesidad metafísica*, ni el TCN ‘agua’ tendría que considerarse como un designador rígido. El tipo de necesidad que, para Putnam, involucra un enunciado de identidad como ‘Agua = H₂O’ es, como él mismo ha llegado a indicar (1983), *física*, no *metafísica*. En nuestro mundo, tal y como lo revelan las investigaciones empíricas, resulta ser físicamente necesario que el agua sea H₂O.

Finalmente, otro aspecto en el que parecen diferir Kripke y Putnam tiene que ver con el tipo de propiedades que constituyen las esencias de las clases naturales. Para Kripke es inviable el reductivismo. Simplemente deja fuera la propiedad esencial de la mente: su fenomenología. Un estado mental es más que las propiedades físicas a las que se pretende reducir. Para apoyar esta afirmación, de nuevo, apela a consideraciones modales. Es posible pensar en un organismo que

instancia las propiedades físicas neuronales con las que se identifica el dolor, sin que dicho organismo experimente dolor. Según Kripke, en tal caso, tendríamos que negar que el organismo en cuestión realmente instancie ‘dolor’. En este sentido, para Kripke, la propiedad esencial de algunos estados mentales no es física. Luego, según el esencialismo de Kripke, las propiedades esenciales de algunas clases naturales no tienen que ser necesariamente físicas. Putnam⁶, por su parte, ataca al reductivismo —la teoría de la identidad tipo— con un argumento distinto: el argumento de la realización múltiple. Un estado mental puede ser realizado por distintas bases físicas. Pero, a diferencia de Kripke, Putnam no intenta exacerbar la importancia de la fenomenología mental, sino el aspecto computacional o funcional de los estados mentales. Si bien el funcionalismo es lógicamente compatible con cualquier forma de dualismo, para Putnam es claro que el aspecto funcional de los estados mentales depende ontológicamente de una base física en la cual se realiza dicho aspecto. Luego, para Putnam incluso las propiedades funcionales de los estados mentales pueden considerarse como propiedades físicas. Finalmente, para Putnam las propiedades que constituyen las esencias de las clases naturales son físicas.

A modo de conclusión, puede decirse que, en efecto, las propuestas de Kripke y Putnam representan una reacción en contra del nominalismo. Tienen en común su rechazo a la especificación de la referencia de los TCN en términos de racimos de descripciones. Ambas representan aproximaciones externalistas a la cuestión de la especificación de la extensión y la naturaleza del significado. Este aspecto se evidencia por la relevancia que le conceden al proceso histórico-colectivo por medio del cual se fija y preserva la referencia de los TCN. Además, concuerdan en que la investigación empírica es crucial para develar la naturaleza íntima (esencia o microestructura) de las sustancias, organismos, etc., y, por lo tanto, esencial en la especificación de la extensión de los TCN. Pero, si bien es correcto reconocer a Kripke y a Putnam como los adalides de la revolución anti-nominalista, es necesario mencionar que existen diferencias entre sus análisis que los conducen a derroteros y posturas distintas. La semántica de Kripke, basada en las nociones de designador rígido y mundos posibles, parece comprometerlo con una noción de necesidad metafísica bastante robusta. Este rasgo de su propuesta es una de las razones (además de los pocos reparos que tiene para hablar de esencias) por las que se le suele considerar como un *esencialista metafísico*. Putnam, quien resiste la terminología esencialista, adopta una noción de nece-

⁶ Aquí glosamos la versión clásica, funcionalista, de la filosofía de la mente Putnam. Tenemos claro que, desde *Representación y realidad*, Putnam ha dejado de abogar por esta versión cruda de funcionalismo.

sidad física que parece ser suficiente para asegurar sus objetivos *realistas*. No obstante tal reticencia, aludiremos a varias ideas de Putnam en la subsecuente caracterización de esencialismo semántico. Esta insistencia —por no decir negligencia— en asociar a Putnam con el esencialismo, se sustenta en el hecho de que el esencialismo semántico al cual nos referimos, está íntimamente ligado a lo que se denomina ‘Teoría Causal de la Referencia’, con la cual se tiende a asociar —por las afinidades programáticas arriba señaladas—, tanto a Kripke y Putnam (Hansberg, 1990; Beebe, et. al., 2010 ; Soames, 2002). Básicamente, el esencialismo semántico al que nos referiremos funda su crítica al nominalismo en los análisis hechos por ambos autores sobre la referencia. En este sentido, y advertidos por las diferencias de fondo que zanján las propuestas de Kripke y Putnam, en adelante, señalaremos, cuando sea necesario, qué ideas pertenecen a cada autor.

2. METAFÍSICA Y SEMÁNTICA DE LAS CLASES NATURALES

Antes de exponer los detalles de la *semántica esencialista* de clases naturales (en adelante abreviada como SE), es importante distinguir entre dos asuntos. Una cuestión es la de determinar *qué es una clase natural*, y otra distinta es la de hablar de los rasgos semánticos que nos llevan a tratar ciertos términos como *términos de clase natural* (TCN). Aunque veremos que las cuestiones están íntimamente ligadas y se traslapan, es importante tener presente sobre qué se discute en cada caso. En cuanto a la SE, se hará evidente que, aunque es una teoría semántica sobre *términos de clase natural*, implica suposiciones *esencialistas* respecto de lo que *cuenta como una clase natural*.⁷ En nuestro caso, entonces, las consideraciones semánticas nos llevarán, quizás inadvertidamente, a la cuestión de *qué es una clase natural*. Siempre que se considere importante se hará notar la transición.

3. SEMÁNTICA ESENCIALISTA: SE

La siguiente caracterización de la SE se basa principalmente en observaciones de Kripke (1981/1995) y, en menor medida, en las ideas de Putnam (1975/1984), lo que significa que, a lo que llamamos SE, es básicamente la semántica kripkeana para clases naturales.

⁷ Hansberg describe sucintamente la continuidad entre ambos asuntos de la siguiente manera: “Para que una palabra sea una palabra de clase natural, es necesario cuando menos que la palabra sea usada como un término de clase natural y que su extensión sea efectivamente una clase natural” Hansberg (1990, p. 216).

Los términos de clase natural se introducen a partir del bautismo de ciertos conjuntos de objetos. Por ejemplo, tomamos sustancias como agua, o cosas como el oro o tigre, elegimos algunos ejemplos privilegiados o paradigmáticos y les asignamos los respectivos nombres de ‘agua’, ‘oro’ y ‘tigre’ (Kripke, 1981). Estipulamos que todo aquello que sea *del mismo tipo* que las muestras paradigmáticas pertenecerá a dicha clase. La relación *del mismo tipo* se elimina *a posteriori*, cuando la indagación empírica descubre la *composición interna* (rasgos esenciales, micro-estructura) de las muestras paradigmáticas. Se determina entonces que, por ejemplo, (el predicado) ‘ser agua’ se aplica a toda muestra (actual y posible) si y sólo si ésta comparte la *composición interna* del paradigma [H₂O] (Kripke, 1981; Putnam, 1975). El descubrimiento *a posteriori* de la composición interna hace posible que, eventualmente, reconozcamos que algunos objetos que considerábamos pertenecientes a la extensión de cierta clase natural, no pertenecen realmente a dicha clase, en virtud de carecer de dicha composición interna —y viceversa (Putnam, 1975). Una consecuencia de esta forma de explicar la naturaleza de los TCN es que no será suficiente, ni necesario, que las muestras satisfagan los rasgos superficiales a partir de los cuales seleccionamos las muestras paradigmáticas.⁸ Muchas cosas pueden satisfacer tales rasgos superficiales y carecer de la composición interna esencial de la clase natural (Kripke, 1981; Putnam, 1975). De nuevo, lo que resulta necesario y suficiente para que una muestra forme parte de la extensión de un término de clase natural es que comparta la misma composición interna del caso paradigmático.

Hansberg (1990) reconoce cuatro características lógico-semánticas de los términos de clase natural que vale la pena tener en cuenta:

1. Las definiciones de las palabras de clase natural, que proporcionen *condiciones necesarias y suficientes* de pertenencia, deberán ser *a posteriori*. Esto es, basadas en la indagación y descubrimiento de la *composición interna* de las muestras paradigmáticas —Hansberg habla de *estructuras explicativas* en lugar de *composición interna* (Kripke, 1981; Putnam, 1975).
2. La [extensión] referencia de estas palabras en parte se define *antes* de que conozcamos las condiciones necesarias y suficientes (composición interna) para su aplicación. La fijamos al apuntar y bautizar a las muestras paradigmáticas (Kripke, 1981).

⁸ Esta implicación se opone a la condición de pertenencia distintiva del convencionalismo.

3. Introducir estos términos a partir de muestras paradigmáticas les confiere un elemento *deíctico*: ‘agua’ refiere a la sustancia que guarda alguna relación de *similitud* con el agua que hay *aquí*. Todo aquello que haya de considerarse agua, en cualquier mundo posible, necesariamente debe satisfacer la relación *mismo líquido* con la muestra paradigmática (Kripke, 1981).
4. Los términos de clase natural funcionan como designadores rígidos en el sentido de Kripke: nombres o descripciones. Esto es, seleccionan al mismo objeto o conjuntos de objetos en todos los mundos posibles en que tales objetos o tipos de objetos existen... sobre la base metafísica de satisfacer condiciones necesarias y suficientes de pertenencia; a saber, tener en común *la misma composición interna* (Kripke, 1981).⁹

II.IV. ESENCIALISMO, NOMINALISMO Y SEMÁNTICA ESENCIALISTA

1. ALGUNAS DIFERENCIAS FUNDAMENTALES ENTRE EL ESENCIALISMO Y EL NOMINALISMO

Para la SE, lo que Locke llamaba ‘esencia real’ no es sólo una misteriosa suposición acerca de la constitución interna de las cosas, que no juega ningún papel en la determinación de la extensión de un término de clase natural. Putnam considera que diferentes clases de cosas tienen una constitución interna (esencia o propiedades esenciales, para Kripke) que es posible conocer. Kripke, por su parte, considera también que hay clases de cosas con propiedades esenciales, las cuales pueden ser físicas —en cuyo caso podrían descubrirse empíricamente— o no físicas. Se conozca o no dicha constitución interna o propiedades esenciales reales, son éstas las que, de un modo fundamental, ayudan a determinar la extensión de un término de clase natural.

Resumiendo, un rasgo crucial en el que difieren la semántica de clases naturales de Putnam y Kripke de la propuesta lockeana es en su carácter *a posteriori*. Son principalmente los hallazgos empíricos, y no las ideas (propiedades) que las personas tienen de las cosas, los que ayudan a determinar la referencia y extensión de los TCN; además, *si en efecto refieren a clases naturales*, los mismos hallazgos empíricos permitirán determinar la extensión de la clase natural

⁹ Que las descripciones sean designadores rígidos es discutible ya que los enunciados que las incluyen no son *de re*. La descripción *el tercer planeta más cercano al Sol*, no designa al mismo objeto en todos los mundos posibles en los que existe *el tercer planeta más cercano al Sol*. Aunque en este mundo la Tierra es *el tercer planeta más cercano al Sol*, puede ser el caso que haya un mundo posible en el que la Tierra no tenga este privilegio.

misma.¹⁰ De esta forma, se pretende superar el convencionalismo implícito en el nominalismo lockeano, y recuperar la esperanza de aspirar a alguna forma de realismo basado en una concepción *esencialista* de los TCN (1-4 del apartado anterior).

2. ESENCIALISMO

Hasta este punto hemos supuesto alguna especie de relación entre la SE y el esencialismo, sin explicar la naturaleza de tal relación. Hay que tener claro que se trata de posiciones filosóficas distintas. SE es una teoría *semántica* con implicaciones ontológicas. Esto es, propone una teoría del significado de los TCN según la cual el significado/referencia de dichos términos aludiría a un conjunto de *propiedades distintivas* de los miembros de dicha clase. Por otra parte, el esencialismo es una doctrina *metafísica* sobre la realidad con implicaciones ontológicas. El esencialismo supone que tanto los individuos como las clases de cosas tienen *propiedades esenciales* que los hacen ser lo que son. Esta distinción nos obliga, ahora, a desarrollar una tipificación del *esencialismo*, de suerte que resulte claro cómo se relacionan ambas teorías.

El primer paso para esclarecer esta relación, entonces, es dar alguna especie de definición del esencialismo. Aunque existen diferentes formulaciones podríamos partir de la siguiente distinción. Hay dos formas de esencialismo: el **esencialismo respecto de individuos** y el **esencialismo respecto de clases naturales**. El primer tipo se entiende como la tesis de que un individuo tiene propiedades esenciales y accidentales, las primeras son aquellas sin las cuales el individuo no pudo haber existido; las propiedades accidentales son propiedades, presentes en el individuo, sin las que el individuo pudo haber existido. El esencialismo respecto de clases naturales afirma, en el mismo sentido, que existen clases de objetos con propiedades esenciales y accidentales, y que sin las primeras no podría existir dicho conjunto de objetos o clase. El esencialismo de clases naturales reconoce que ciertas clases de cosas tienen ciertas propiedades, en virtud de las cuales la clase es lo que es.

3. DE LA NATURALIDAD O QUÉ SON LAS CLASES NATURALES

Pero la definición anterior es insuficiente para representar apropiadamente el esencialismo respecto de las clases naturales. El problema es que clases artificiales como lápiz o silla podrían en-

¹⁰ Nótese que la SE puede ofrecer respuestas interesantes a la cuestión de *cuáles son los rasgos semánticos de los términos de clase natural* o a *qué términos deberíamos tratar como tales*. Pero, a partir de lo dicho hasta este punto, no podemos decir nada sobre el asunto de *qué son las clases naturales*.

cajar también con esta definición. Si algo que parece lápiz no sirve para realizar trazos, entonces no podríamos decir que es un lápiz, aunque lo parezca. Luego, en ausencia de los rasgos esenciales de ‘ser lápiz’, no se puede decir que exista por lo menos un lápiz. Este se conoce como el problema de la *naturalidad* de las clases naturales (LaPorte, 2004) y, en la medida que se logre una formulación consistente de dicha noción, se puede responder la cuestión de *qué son las clases naturales*.

3.1. La concepción natural de la naturalidad. ¿Cómo debería entenderse la *naturalidad* de las clases naturales para el esencialismo? Pueden intentarse varias rutas para llegar a formular un esencialismo interesante respecto de las clases naturales (si no ha de diluirse una forma más general y trivial de esencialismo). Comencemos con lo que redundantemente se podría llamar la concepción *natural* de la *naturalidad*. Esta concepción sostendría que las clases naturales, a diferencia de las artificiales, se encuentran en la naturaleza. Veamos qué problemas se presentan a esta noción de naturalidad.

Muchas cosas que consideramos clases naturales son producidas por el hombre. Minerales como el cuarzo y los diamantes se obtienen en laboratorios; lo mismo que sustancias químicas sintéticas (como el tecnecio) y hasta nuevas variedades de plantas (semillas). Esta noción de naturalidad y de lo que son las clases naturales puede ser lógicamente insostenible *si* se es también esencialista.

Ya que el esencialismo sobre clases naturales sostiene que existen clases de objetos con propiedades esenciales (sin las cuales no podría existir dicha clase de objetos): ¿qué diríamos de dos diamantes, uno extraído de minas en África y otro producido en un laboratorio cuya composición química es la misma? Si los diamantes se consideran una clase natural, entonces (por el esencialismo) han de tener propiedades esenciales. Luego, según la concepción de naturalidad que ahora nos ocupa, si uno de los dos diamantes ha de considerarse como una muestra de la clase natural *diamante*, ha de ser el extraído de las minas de África (porque se encuentra en la naturaleza); y, también, debe serlo en virtud de sus propiedades esenciales (si se quiere, de su composición química), según el esencialismo. Hemos de concluir entonces que, ya que el diamante de laboratorio no pertenece a la clase natural ‘diamante’ (por no encontrarse en la naturaleza), (a) no tiene las propiedades esenciales de los diamantes (lo que es falso); y que (b), ya que no tiene las propiedades esenciales para ser parte de la clase natural diamante, por extensión, nada que tenga

sus mismas propiedades esenciales forma parte de dicha clase, lo cual implica que el diamante extraído de las minas de África tampoco es un diamante (lo que nos conduce a una contradicción). Con esto se reduce al absurdo la noción *natural* de *naturalidad* y de las clases naturales.

3.2. Hansberg y LaPorte: la concepción del valor explicativo. Otra noción de naturalidad se basa en *algún* sentido en el *valor explicativo* de las clases naturales. Hansberg (1990, p. 216) y LaPorte (2004, p. 19), por ejemplo, presentan una noción de *naturalidad* y de lo que son las clases naturales, en estos términos. El problema con la propuesta de estos autores es que, como intentaré evidenciar enseguida, *el valor explicativo* que pudieran tener las clases naturales respecto de cierto tipo de cuestionamientos, no es suficiente (por sí mismo) para establecer un criterio de naturalidad y de lo que son las clases naturales. Me inclino a pensar que las dificultades en la noción de naturalidad de Hansberg y LaPorte se deben a cierta reticencia argumentativa.

Según la concepción de estos autores, lo que ayuda a distinguir entre una clase artificial y una natural es que las últimas representan una respuesta *máximamente explicativa*¹¹ a preguntas del tipo ‘¿qué son estos?’. Se explica mucho al responder a una pregunta de este tipo con un término de clase natural; por ejemplo ‘oso polar’. Ser un oso polar explica por qué cría oseznos de la *manera* en que lo hace; por qué tiene el pelaje de tal color y tan denso; o por qué nada grandes distancias a través de agua gélida en busca de macizos de hielo. Existen también respuestas teóricamente satisfactorias sobre por qué los osos polares crían a sus cachorros como lo hacen; por qué tienen pelaje blanco y denso; y por qué nadan grandes distancias. La clase ‘oso polar’ es *máximamente explicativa* respecto de qué son y cómo se comportan los miembros de esa clase. Algo muy distinto sucede cuando se pregunta sobre objetos como lápices, sillas y la clase de *los-objetos-nombrados-en martes*, ‘¿qué son estos?’. No tenemos teorías sobre estos objetos que expliquen sus propiedades y por qué las tienen. Las clases artificiales no son, por lo tanto, *máximamente explicativas*.

Por lo menos una observación puede hacerse en torno a esta noción de naturalidad basada en el valor explicativo de las clases naturales. Resulta que, según esta concepción, se puede decir también de funciones y operaciones matemáticas que son clases naturales. Responder ‘multipli-

¹¹ El concepto *máximamente explicativo* es más bien utilizado por Hansberg —quien lo toma de Platts—, y no por LaPorte. En cualquier caso, LaPorte sostiene esencialmente lo mismo que Hansberg, al afirmar que ‘una clase natural es una clase con valor explicativo’ (*Ibid.*).

caciones' a una pregunta del tipo en cuestión —por ejemplo, un niño que pregunta ‘¿qué son estos?’ respecto de las multiplicaciones que ve en una hoja—, alude a un *tipo* de teoría bastante satisfactoria en torno a las multiplicaciones. La teoría, lo mismo que en el caso del oso polar, puede explicitarse hablando de la forma en que han de interpretarse y hasta llamarse a los números en el contexto de la operación; su comportamiento en relación con propiedades aritméticas (conmutación, distribución y asociación), etc. Muchas explicaciones teóricamente interesantes y explicativas existen detrás de ser una multiplicación, lo cual nos habría de autorizar para decir —según la concepción del valor explicativo de la naturalidad— que las operaciones matemáticas conforman clases máximamente explicativas y, por lo tanto, clases naturales. Si el esencialismo respecto de clases naturales no acredita una conclusión que incluya a clases *abstractas* dentro del conjunto de las clases naturales, es evidente entonces que el criterio de naturalidad propuesto por Hansberg y LaPorte no es suficiente para decidir *qué es una clase natural*.

El problema con esta concepción de naturalidad tiene que ver con lo que se entienda con las nociones de *valor explicativo* y *teóricamente interesante*. Seguramente habría que realizar algunas estipulaciones respecto de lo que se considera *explicativamente* y *teóricamente* interesante, de suerte que apunten hacia cierto tipo de *explicaciones* y *teorías*; a saber, explicaciones y teorías *empíricas* acerca de *propiedades físicas*. Esta restricción podría ayudar a excluir las clases abstractas del dominio de las clases naturales. Esta omisión para especificar lo que se habrá de entender con las nociones de *valor explicativo* y *teórico* de las clases naturales, es la reticencia argumentativa que se señaló unos párrafos arriba respecto de la propuesta de Hansberg y LaPorte.

3.3. Platts y las clases explicativas. Mark Platts (1997, p.268-70) —a quien tanto Hansberg como LaPorte aluden al discutir el tema de la naturalidad¹²—, defiende la misma concepción de *naturalidad* —de la importancia o *valor explicativo*—, con la diferencia de que parece evitar las omisiones de la propuesta anterior.

La tipificación de la *naturalidad* o de *clase natural* de Platts parte de algunas observaciones de Putnam sobre la importancia que tienen las *propiedades físicas relevantes* en la determi-

¹² Ambos autores reconocen la deuda que su concepción de naturalidad tiene con la propuesta de Platts. En virtud de este reconocimiento, es muy probable que tanto LaPorte como Hansberg hayan tenido presente la restricción *empírica* —presente en Platts— que completaría satisfactoriamente sus criterios de *naturalidad*. Que no la hayan incluido, quizás, tuvo que ver con exigencias de agilidad expositiva en sus respectivos trabajos. Es muy probable, entonces, que la omisión haya sido deliberada en ambos casos, y que estarían de acuerdo en seguir la ruta que se sugiere en este apartado.

nación de lo que ha de contar, precisamente, como clase natural. En este sentido, se plantea la pregunta, ‘¿por qué son relevantes las propiedades físicas relevantes?’ (Platts, *op. cit.* p. 268-9). Platts responde en los siguientes términos: *las propiedades físicas relevantes* son *relevantes* —para decidir si una clase es natural— porque es sobre ellas que tratan las leyes y teorías *empíricas* interesantes (correctas) que *explican* la naturaleza de los miembros de dicha clase. Además, de estas leyes y teorías *empíricas* que versan sobre las *propiedades físicas relevantes* de los miembros de la clase, obtienen su *máximo valor explicativo* las respuestas a preguntas del tipo ‘¿qué son estos?’, que implican el uso de un término que refiere a dicha clase. Cumple agregar que Platts indica puntualmente sobre qué versan tales teorías y leyes explicativamente valiosas; a saber, *propiedades físicas relevantes* (de los miembros de la clase natural). Esto nos lleva a concluir que tales teorías y leyes explicativamente valiosas, ya que versan sobre *propiedades físicas* (de ciertas clases de objetos), son *empíricas*. Platts llama *clases explicativas* (*explanatory kinds*) o clases naturales, a las clases de objetos cuya naturaleza de sus miembros se *explica* de esta forma: a partir de teorías y leyes *empíricas* o *a posteriori* sobre las *propiedades físicas* de tales clases de objetos.

Ahora podemos examinar esta noción de *naturalidad* para saber si puede integrarse a las tesis esencialistas, de modo que sea posible construir un concepto de clase natural que excluya clases artificiales y abstractas. La noción de naturalidad que propone Platts, puede articularse con la tesis esencialista, si identificamos las *propiedades esenciales* —sin las cuales no podría existir cada clase natural— con las *propiedades físicas relevantes* sobre las que tratan las leyes y teorías que explican la naturaleza de los miembros de una clase natural.¹³

Tendríamos entonces, tomando en cuenta las observaciones de Platts, que una *clase natural* (clase explicativa) desde el punto de vista *esencialista* puede tipificarse de la siguiente manera:

1. Un conjunto de objetos que tienen en común una esencia o propiedades esenciales, de suerte que, de no existir tal esencia o propiedades esenciales, no podría existir dicha clase de objetos.

¹³ Esta identificación es la que, de hecho, intenta establecer Kripke entre *propiedades físicas relevantes* y *esencia* o *propiedades esenciales*.

2. Existen (o pueden existir) leyes y teorías *empíricas* sobre tales objetos —explican su naturaleza... *propiedades físicas/esenciales*—, en virtud de las cuales una respuesta a cuestionamientos del tipo ‘¿qué es esto?’, con el término que refiere a dicha clase de objetos, es *máximamente explicativa* respecto de los miembros de la clase.

(Que la respuesta sea *máximamente explicativa* obedece a que está basada en leyes que son sobre las *propiedades físicas/esenciales* que tienen en común los miembros de la clase)¹⁴.

Así expuesta, esta noción de naturalidad, unida a la tesis esencialista, nos ayuda a resolver los problemas interpuestos a las concepciones anteriores. Por un lado, excluye a las clases artificiales porque no hay leyes ni teorías sobre las propiedades físicas de tales clases de objetos que ayuden a explicar la naturaleza de sus miembros (no satisfacen 2). Las definiciones de los términos de clases artificiales se basan en lo que sabemos sobre estas clases de objetos (sus propiedades esenciales), y todo esto lo conocemos *a priori*: no son *empíricas* o *a posteriori* nuestras teorías sobre tales clases, si es que las tenemos. Por esta razón, responder a una cuestión del tipo ‘¿qué son estos?’ con términos como ‘sillas’ o ‘lápices’ no resulta ser una respuesta *máximamente explicativa*. Las propiedades esenciales de esta clase de objetos, si las tienen, no son naturales. También excluye a las clases abstractas como las operaciones matemáticas ya que, aunque pueden ser *explicativamente valiosas* —en virtud de las leyes y teorías que explican su naturaleza—, sus propiedades —explicadas por medio de dichas teorías y leyes— no son *físicas*, ni se establecen o descubren *empíricamente (a posteriori)*; de lo cual resulta que tampoco tienen propiedades esenciales naturales: no son clases naturales (tampoco satisfacen 2).

Recapitulando, si entendemos la noción de *naturalidad* en los términos que Platts sugiere, podemos llegar a la siguiente definición aparentemente consistente de esencialismo. El esencialismo respecto de las clases naturales sostiene que existen conjuntos de cosas (clases) con propiedades esenciales, sin las cuales no podrían existir dichos conjuntos. Además, lo que determina la naturalidad de tales clases es que existen (o pueden existir) leyes y teorías empíricas sobre las propiedades físicas relevantes/esenciales de los miembros de dichas clases, en virtud de las cuales se logran respuestas *máximamente explicativas* a preguntas como ‘¿qué son estos?’ con la simple emisión del término que refiere a los miembros de dicha clase.

¹⁴ Platts, (1997, pp. 268-9).

Podría increparse la caracterización de naturalidad o de clase natural de Platts señalando que ésta supone que son ciertas propiedades *epistemológicas* las que, finalmente, ayudan a determinar qué conjuntos de objetos son clases naturales. El señalamiento indica un problema porque hablar de propiedades epistemológicas tiene más que ver con nuestras formas de describir y conocer las cosas, que con las cosas mismas. De ser así, la noción de clase natural que pretende ser útil para realizar distinciones ontológicas relevantes, termina refiriendo a una distinción más bien criteriológica. Sobre esto, Platts menciona que, el que podamos desarrollar y, de hecho, existan teorías y leyes empíricas (*a posteriori*) sobre *tigres* (clases naturales putativas) y no sobre lápices, refleja algo acerca de la naturaleza de los *tigres* y los *lápices*, no sobre nuestras capacidades epistemológicas.¹⁵ La noción de naturalidad de Platts apunta más hacia ciertas propiedades físicas relevantes *en* ciertas clases de cosas, que a nuestras capacidades para descubrirlas o conocerlas. La noción de naturalidad descansa, finalmente, en propiedades que están en las cosas mismas.

4. ESENCIALISMOS DE CLASES NATURALES, REALISMO DE CLASES NATURALES Y SEMÁNTICA ESENCIALISTA (SE)

¿Cómo termina relacionándose el esencialismo con SE? El esencialismo es una tesis metafísica que sostiene que existen clases de cosas con esencias o propiedades esenciales, sin las cuales no pueden existir ni la clase, ni los miembros que la componen. Guiados por Platts, identificamos estas esencias o propiedades esenciales con las propiedades físicas relevantes descubiertas *a posteriori* —estructuras explicativas— y, al hacerlo, ganamos una fórmula esencialista capaz de distinguir clases naturales de clases artificiales y abstractas. La Semántica Esencialista (SE), basada en las ideas de Kripke y Putnam, se articula muy bien con esta formulación.¹⁶ Las estructuras explicativas que distinguen a las clases naturales de otras clases de objetos, son aquello que, según SE, *determina* la extensión de una palabra de clase natural.

Pero distintos matices de esencialismo de clases naturales, se derivan de la idea anterior: que las estructuras explicativas ayudan a identificar clases naturales y a definir la extensión y, parcialmente, el significado de un TCN. Por ejemplo, para Ellis (2001), la estructura explicativa o

¹⁵ *Op. cit.* p. 270.

¹⁶ Es evidente que Hansberg (1990), Platts (1997) y LaPorte (2004) desarrollan nociones de clase natural, no sólo consistentes, sino hechas a la medida de SE.

esencia —propiedades esenciales o propiedades físicas relevantes descubiertas *a posteriori*— se entiende como una condición *necesaria y suficiente* que todo miembro, —pasado, presente y futuro— de una clase natural *debe* satisfacer para que pueda decirse verdaderamente de él que forma parte de la extensión de un término de clase natural. Este es un tipo de esencialismo relativamente fuerte, ya que, al considerar la estructura explicativa como condición de pertenencia *necesaria y suficiente*, supone que las clases naturales son categóricamente distinguibles y que no tienen límites ontológicos vagos.

Otro tipo de esencialismo afín al anterior, pero más robusto aún, uno al que llamaremos *esencialismo metafísico de clases naturales*, está asociado a las ideas de Kripke. También se deriva de ciertos rasgos —kripkeanos— de la SE y es compatible con las observaciones de Platts sobre la estructura o esencia de las clases naturales. El esencialismo metafísico de clases naturales también parece comprometerse con la tesis de que la estructura explicativa¹⁷ o esencia estipula condiciones necesarias y suficientes para fijar la referencia de un TCN y, por lo tanto, también para determinar de forma precisa la pertenencia a una clase natural. Además, insiste en que las condiciones necesarias y suficientes que estipula la estructura explicativa o esencia, para fijar la extensión de un TCN y la pertenencia a una clase natural, son operativas en *todos los mundos posibles*, no sólo en el actual. Según esta variante de esencialismo, en cualquier mundo posible en el que exista agua, debe *necesariamente* haber una clase de líquido compuesto por moléculas de H₂O. Evidentemente, el *esencialismo metafísico de clases naturales* que suscribe Kripke está ligado con su noción de *necesidad metafísica* (*supra* I, II.III, 1)¹⁸, la cual consigna que el referente de un TCN es necesariamente idéntico a su esencia —estructura explicativa— en todos los mundos posibles en los que exista.

Vale la pena hacer notar que la Semántica Esencialista (SE) caracterizada antes —I, II.III, 3— que es básicamente la semántica kripkeana, no es apta para derivar la postura de Putnam en relación a las clases naturales, a la cual llamaremos, no esencialismo, sino *realismo de las clases naturales*. Esto se debe a que nuestra caracterización de SE apela a ideas —kripkeanas— como la de que los TCN se comportan como designadores rígidos que fijan la referencia en todos los mundos posibles. Como se había indicado con anterioridad (I, II.III, 1) Putnam no se toma al pie

¹⁷ Hay que tener claro que, en el caso de Kripke, puede ser que la estructura explicativa o esencia de una clase natural incluya propiedades no físicas, como *qualia*.

¹⁸ Las referencias a otras parte de este mismo trabajo indican: PARTE, CAPÍTULO, número de apartado (si es el caso).

de la letra estas ideas. Más bien, considera que los TCN fijan la referencia sólo para el mundo actual, y aquello a lo cual refieren es la estructura interna de las cosas o sustancias que nombran, la cual se descubre por medio de indagación empírica. Esto, obviamente, se deriva de su más modesta noción de *necesidad física* (I, II.III, 1), según la cual el referente de un TCN es idéntico, en este mundo, a su composición interna o estructura explicativa —para utilizar la terminología de Platts. Luego, lo que sostiene el *realismo de clases naturales* de Putnam, a diferencia del *esencialismo metafísico* de Kripke, es que las clases naturales, en efecto, son idénticas *en este mundo* a una cierta composición interna o estructura explicativa, nada más. Además, también hay que recordar que Putnam no habla de la composición interna en términos de esencias, y que tampoco se refiere a ella como estableciendo condiciones necesarias y suficientes, ya sea para determinar la extensión de un TCN o para determinar la pertenencia a una clase natural.

Estas distinciones en torno a la forma en la que SE se compagina con ciertas formas de esencialismo de clases naturales, y cómo no es consistente con otras concepciones de clase natural, son importantes porque, más adelante, cuando definamos nuestra postura en torno a la semántica de los TCN y la metafísica de las clases naturales, retomaremos sólo algunos aspectos de la SE y reformaremos otros y, en consecuencia, terminaremos suscribiendo una metafísica de las clases naturales más apegada al *realismo de clases naturales* de Putnam que al *esencialismo metafísico* de Kripke. Algunos aspectos que retomaremos y que vale la pena identificar desde ahora son de tipo terminológico. En adelante y para el resto del trabajo, con la intención de descargar de peso metafísico la concepción de clase natural que adoptaremos, hablaremos de ‘la estructura explicativa de las clases naturales’ en lugar de ‘la esencia de las clases naturales’; y de ‘propiedades físicas relevantes’, en lugar de ‘propiedades esenciales’.

Finalmente, hay que tener presente que este capítulo tiene como finalidad, en primer lugar, ofrecer un escenario general de la historia de la noción de clase natural y, en segundo lugar, analizar las distintas posturas y realizar las distinciones necesarias para, eventualmente, construir nuestra propia semántica de los TCN y metafísica de las clases naturales.

5. *ESENCIALISMO: TEORÍAS DESCRIPTIVAS DE LA REFERENCIA Y SE*

Aunque se haya establecido la consistencia y afinidad entre SE y ciertas formas de esencialismo, y también la inconsistencia de SE con otras posturas sobre las clases naturales —*realismo de cla-*

ses naturales—, siguen sin explorarse otras posibilidades. Por ejemplo, ¿no es concebible que una teoría descriptiva de la referencia como la que sustenta el convencionalismo de Locke fuese también compatible con alguna forma de esencialismo de las clases naturales como los descritos en el apartado anterior? Después de todo, Locke no deja de reconocer que ciertas clases de cosas tengan esencias reales, en virtud de las cuales tienen las propiedades superficiales que reconocemos en ellas y que utilizamos (como descripciones) para determinar la referencia y extensión de los términos de clase natural. Luego, de manera indirecta, el convencionalismo lockeano supone también el esencialismo respecto de las clases naturales. ¿Por qué entonces se relaciona SE con el esencialismo y no con el convencionalismo de Locke?

Mackie (2006) muestra que la posibilidad de aceptar cabalmente alguna forma de esencialismo respecto de las clases naturales —como las que se describieron en el apartado anterior— depende por completo de elegir una semántica con las características de SE, y no una teoría descriptiva de la referencia. La idea fundamental es que la SE, es compatible con el esencialismo,¹⁹ lo cual no se puede decir de las semánticas de clase natural que derivan de las teorías descriptivas/convencionalistas. En este sentido, si consideramos que el esencialismo —cualquiera de los tipos mencionados antes— sobre clases naturales es intrínsecamente plausible, pero que su aceptación es problemática mientras dominen las teorías descriptivas de la referencia; entonces adoptar la SE ha de considerarse como indispensable para abrir el camino al esencialismo, en la medida en que ayuda a remover las viejas teorías descriptivas.

Mackie²⁰ piensa que puede dar un diagnóstico de porqué existe un problema entre el esencialismo y las teorías descriptivas a las cuales pretende remplazar la SE. Para evidenciar el problema comienza citando algunas características de las teorías descriptivas:

- (a) Las descripciones relevantes con las que se asocia un término de clase natural se establecen *a priori* por todos o al menos algunos (expertos) usuarios competentes del término (convencionalismo). A lo que Mackie llama *a priori* es al hecho de determinar, sin realizar indagaciones científicas ulteriores, qué propiedades se consideran como relevantes o características de cierta clase de cosas... a la cual se hace referencia por medio de un término de clase.

¹⁹ Véase apartado anterior.

²⁰ Mackie (2006; 10.3).

- (b) Las descripciones se expresan en términos de condiciones necesarias y suficientes para la pertenencia a la clase.
- (c) La relación que rige entre una muestra (o miembro) de la clase y las descripciones relevantes es simplemente la de satisfacer o cuadrar. En particular, no están involucradas relaciones causales entre usuarios y las muestras (miembros).

Luego Mackie propone lo siguiente. Abréviase el conjunto de las propiedades relevantes con D, y considérese que un esencialista postula alguna propiedad M —microestructural— como metafísicamente necesaria para pertenecer a alguna clase... aunque no es (o se determina) *a priori* que los miembros de la clase tienen M. Si imaginamos que existen objetos que satisfacen D y carecen de M, para el descriptivista, en virtud de (a-c), tales objetos pertenecerían a la clase. Por lo tanto, el descriptivista no acepta como metafísicamente necesario, para pertenecer a la clase, tener M; contrario a lo que el esencialismo afirma. De esto se sigue que un teórico descriptivo no puede aceptar que M represente una condición necesaria y suficiente para pertenecer a la clase... en contra de lo que sostiene el esencialismo.

De esta forma, se puede mostrar que existe inconsistencia entre los requerimientos de una teoría descriptiva y los del esencialismo. El esencialismo introduce condiciones necesarias *a posteriori* para la pertenencia a alguna clase; mientras el descriptivismo insiste en condiciones necesarias y suficientes estipuladas *a priori*. Esto hace posible que existan objetos que satisfacen las descripciones relevantes reconocidas por los descriptivistas, y fallen en satisfacer las condiciones estipuladas por el esencialismo.

III. IMPLICACIONES DEL ESENCIALISMO PARA EL ESTUDIO DE LAS EMOCIONES

El esencialismo sobre clases naturales que se discutió en el capítulo anterior, constituye una posición filosófica que, si bien no se halla exenta de controversia, sigue contando adeptos; de hecho, constituye uno de los últimos capítulos en la historia de la tradición del concepto de *clase natural*. Es importante reconocer, aunque sea de manera rústica, los avances que representa esta concepción, así como las causas que la hicieron posible.

Como ya se ha apuntado antes, el esencialismo sobre clases naturales debe su renacimiento principalmente a las teorías causales de la referencia, principalmente a la de Kripke,

de la cual se deriva principalmente la Semántica Esencialista (SE) que se ha expuesto y discutido en el capítulo anterior. La existencia de dichas teorías representó mucho para la filosofía del lenguaje y de la ciencia del último cuarto del siglo pasado. Por un lado, emergieron como una alternativa sólida respecto de las teorías descriptivas de la referencia, capaz de exhibir más perspicuamente sus problemas, limitaciones e implicaciones (nominalismo/convencionalismo). Por otro lado, aunque no al margen de lo anterior, las teorías causales de la referencia supusieron la posibilidad de una suerte de redención de nuestra intuición natural de *progreso científico*, frente a los problemas que planteaba para dicha intuición el presunto fenómeno de la *incommensurabilidad teórico-científica* (Khun, 1962). Si la referencia de nuestros términos de clase natural —los términos sobre los que versan las leyes en las teorías científicas— no se determina por medio de descripciones, sino a partir de su estructura explicativa, la cual fija la referencia de tales términos de una vez por todas, una vez que es descubierta —empíricamente—, entonces no debería preocuparnos de que las teorías vayan y vengán. El referente de nuestros términos de clase natural no cambia, y ellos pueden llegar a ser aquello que tienen en común las diferentes teorías científicas en la medida en que se suceden.²¹

El esencialismo relacionado con la SE, por su parte, ha ejercido una poderosa influencia, la cual se ha hecho sentir en diversos campos de interés filosófico. Los efectos que nos interesa explorar de manera más detallada son los que se resienten en la filosofía de la mente: en particular, en sus consecuencias para el estudio de la naturaleza de las emociones. En otras palabras, ¿tiene el esencialismo basado en la SE consecuencias para el estudio y nuestra comprensión de la mente en general (de sus capacidades, proceso y/o estados), y de las emociones en particular? ¿Tienen la mente y los estados mentales (incluidas las emociones) propiedades esenciales (conocidas *a posteriori*) que constituyan condiciones necesarias y suficientes para conformar una clase natural de fenómenos? ¿Son los estados mentales clases naturales?

Nuestra respuesta anticipada a estas cuestiones es que, la SE y la metafísica esencialista de las clases naturales que conlleva —más la asunción de otros supuestos en torno a la naturaleza de los estados mentales—, parece condenar a los estados mentales y, por lo tanto,

²¹ La imagen es simple y como tal está expuesta a muchos problemas. Pero como ficción es útil para ejemplificar las ventajas que representó en su momento SE ante la amenaza de la *incommensurabilidad*.

a las emociones a su expatriación de la nación de las clases naturales. Para mostrar esto, a continuación, expondremos y analizaremos los argumentos que Hansberg (1990) ofrece para mostrar que los estados mentales no son clases naturales, toda vez que asume una variedad de esencialismo de clases naturales a la Kripke —*esencialismo metafísico de las clases naturales*— derivado de una semántica esencialista (SE). La enseñanza que intentaremos extraer de las siguientes reflexiones estribará en la necesidad de revisar de forma crítica algunas asunciones entorno a la naturaleza de los estados mentales —asociadas con la tesis de la realización múltiple— y de reformular la semántica de TCN y, por lo tanto, la metafísica de las clases naturales.

III.I. LOS ESTADOS MENTALES EN LA MIRA DEL ESENCIALISMO

En un artículo que ha sido varias veces citado en este trabajo, Olbeth Hansberg (1990) lleva a sus últimas consecuencias, en el terreno de la filosofía de la mente, el esencialismo de clases naturales basado en la SE.²² En particular, nos referimos al *esencialismo metafísico de clases naturales* que adscribimos a Kripke (I, II.III, 1). La respuesta de Hansberg a la pregunta, “¿son los estados mentales (lo que incluye a las emociones) clases naturales?” es, basada en sus argumentos, un rotundo “no”. La generalidad con que se plantea la cuestión, de los argumentos utilizados y, por lo tanto, del alcance de la resolución negativa, hacen que cualquier intento particular de conferir a algún estado mental el rango de clase natural sea inútil: ni siquiera los más perspicuos miembros de la población mental (p. ej. las creencias), podrían aspirar a ser considerados clases naturales en virtud de los resultados de los análisis de Hansberg.

El valor del artículo de Hansberg —a pesar de haber sido publicado hace ya más de dos décadas— radica en hacer explícitos todos los supuestos e implicaciones que escoltan al esencia-

²² El uso que se hace del artículo de Hansberg (‘¿Por qué los estados mentales no son clases naturales?’, 1990) a lo largo de este capítulo, exige algunas aclaraciones. A partir de una conversación personal, nos resulta ahora bastante improbable que Hansberg suscriba aún (tal y como aparecen formuladas en el escrito en cuestión) todas las tesis y argumentos que aquí se citan y analizan. Aun así, nuestro interés por el texto referido se funda en los siguientes hechos: en él se expone de manera compendiosa y conspicua la Semántica Esencialista (SE). Además, y más importante para nuestros propósitos, evidencia el impacto y consecuencias que SE y su metafísica de clases naturales tiene para la cuestión de *la naturaleza de los estados mentales*. Igual de importante resulta que, al exhibir tales consecuencias, utiliza algunas asunciones características de la filosofía de la mente de finales del siglo pasado: a saber, la Tesis de la Realización Múltiple y la concepción ortodoxa de la reducción (la que se remonta a Ernest Nagel). En resumen, ‘¿Por qué los estados mentales no son clases naturales?’ reúne en un crisol argumental una confederación de tesis que, desde nuestra perspectiva, es necesario abatir, ya que en su conjunto han encarnado desde finales del siglo pasado —y en alguna medida lo siguen haciendo— el principal obstáculo para conferir a los estados mentales —incluidas las emociones— el estatus de *clases naturales*.

lismo basado en la SE, al aplicarlo al caso del estatus de los estados mentales en tanto que clases naturales. Por esta razón, es de suma relevancia comentar su contenido para los propósitos de este trabajo —a saber, elucidar el estatus de las emociones en tanto que clases naturales.

I. LOS ARGUMENTOS DE HANSBERG

No se puede negar que, de ser verdadero el esencialismo basado en la SE, junto con otras premisas que Hansberg agrega, resultaría completamente cierto que los estados mentales no conforman una clase natural o clases naturales. El propósito de este apartado es exponer los puntos de vista de Hansberg (1990) con la finalidad de mostrar que, a pesar de que sus argumentos cierran una cuña devastadora contra la tesis de que los estados mentales (incluidas las emociones) son clases naturales, este veredicto puede revertirse si se modifican algunos supuestos de SE y se cuestionan otros en relación a la naturaleza de los estados mentales.

En ‘¿Por qué los estados mentales no son clases naturales?’, Hansberg intenta demostrar que no pueden existir leyes psicofísicas. Si las leyes son enunciados cuyos sujetos y predicados son *necesariamente*²³ clases naturales, y los estados mentales no son clases naturales; entonces no puede haber leyes de ningún tipo (ni psicofísicas, ni psicológicas) que incluyan términos de estados mentales. Para llegar a esta conclusión es menester mostrar, antes que nada, que los estados mentales no son clases naturales. Por su puesto, si se es *esencialista* y se suscribe también SE, la mejor (para no decir que la única) manera de probarlo consiste en evidenciar que los estados mentales carecen de una *esencia* o *estructura explicativa*. Ahora bien, en el momento en el que Hansberg desarrolla su planteamiento existía un procedimiento filosófico estándar, útil para mostrar que los estados mentales no pueden tener *esencias* o *estructuras explicativas*: podríamos hablar de este procedimiento como *la prueba de la reducción*. Bajo cierta concepción (ortodoxa) de la reducción²⁴, un término de estado mental debe ser co-extensivo con *un* término de estado físico; en esto consistiría, compendiosamente, la reducción de lo mental a lo físico. Entendida así,

²³ Aquí la necesidad debe entenderse en un sentido más débil que el de la necesidad lógica. Quizás sería apropiado hablar de una necesidad metafísica (si somos en alguna medida realistas respecto de las leyes, y no instrumentalistas), en el sentido de que cualquier ley o teoría que se considere ontológicamente respetable debe tener como sujetos términos de clase natural.

²⁴ No referimos aquí al modelo de Nagel.

la reducción hace más que establecer una relación de co-extensividad²⁵ entre un término de la teoría reducida (psicología) y un término de la teoría reductora (neurofisiología, p. ej.); proporciona, también, la *esencia* o *estructura explicativa* de un estado mental, con lo cual se establecería que *por lo menos algunos* o *algún* estado mental conforma una clase natural. Naturalmente, para Hansberg, de igual modo que para muchos filósofos de la mente de ayer y de hoy, algo obsta para que los estados mentales acrediten la *prueba de la reducción*: a saber, la *Realización Múltiple*. La *realización múltiple de lo mental* (que un mismo tipo de estado mental se implemente en diferentes estructuras o bases físicas) impide establecer una relación *uno a uno* entre los términos de estados mentales y los términos de estados físicos; lo cual constituye el hecho mismo de que no es posible establecer la relación de co-extensividad biunívoca y, por lo tanto, tampoco la reducción.

Así es como la *prueba de la reducción*, coludida con la *realización múltiple* y con alguna forma de *esencialismo* (en este caso una vinculado con SE), parecen ser suficientes para mostrar que los estados mentales no son clases naturales y, en virtud de este hallazgo, para mostrar también que no existen leyes psicofísicas. Vale la pena señalar que esta estrategia —sólo esquemáticamente expuesta hasta aquí— cumple la doble función de mostrar que los estados mentales no son clases naturales y que, por la misma razón, no pueden existir leyes psicofísicas. El orden lógico en el que se llega a estas dos conclusiones es el mismo, aunque es básicamente un solo paso el que garantiza ambas. Cuando se advierte que los estados mentales no pasan la *prueba de la reducción*, se obtiene el paquete completo. Como a este trabajo no le interesa —por el momento— el paquete completo de conclusiones, nos hemos de concentrar en los argumentos de Hansberg en la medida en que, aparentemente, deciden que los estados mentales no son clases naturales.

Se pueden identificar por lo menos tres argumentos interrelacionados que intentan mostrar que los estados mentales no son clases naturales y que, por lo tanto no pueden existir leyes psicofísicas. El primero se basa principalmente en la SE y la *tesis de la realización múltiple*; el segundo se apoya en consideraciones respecto de la semántica de los términos de estados mentales;

²⁵ La relación de reducción puede entenderse, por lo menos, en dos sentidos más: uno más débil, *co-variación*; y otro más fuerte, *identidad*. No es un problema en este momento la falta de compromiso respecto de cualquiera de estas interpretaciones de la relación de reducción. Por tal razón se ha optado por la más neutral.

y el tercero se propone mostrar que para los términos de estados mentales no aplica “la división del trabajo lingüístico”, fenómeno característico relacionado con los términos de clase natural.

Enseguida expongo y analizo los argumentos por separado para hacer un balance de la situación en la que dejan la cuestión del estatus de los estados mentales (incluidas las emociones) en tanto que clases naturales. La finalidad de esta tarea consiste en hacer explícitos los supuestos que niegan a los estados mentales el estatus de clases naturales. En esta dirección, veremos que para los tres argumentos es fundamental (1) cierta construcción de lo que se ha de considerar como estructura explicativa —esencia— de las clases naturales y (2) la tesis de realización múltiple. Y aquí estará nuestra oportunidad: quizás la forma en la que los argumentos de Hansberg (1990) exigen que construyamos la noción esencialista de *estructura explicativa* no es la única, ni la más adecuada... y quizás tampoco deberíamos estar muy seguros de lo que el argumento de la realización múltiple parece probar.

1.1. Primer argumento: realización múltiple. La primera ruta que considera Hansberg para mostrar que los estados mentales no son clases naturales emplaza el esencialismo basado en la SE, y la tesis de la realización múltiple (TRM).

Plantea la pregunta de la siguiente forma: “¿los términos de estados mentales denotan clases naturales?”. Siguiendo los requerimientos que estipula la SE para los términos de clase natural, será necesario que los términos de estados mentales se definan a partir de sus propiedades esenciales o, usando la terminología de Hansberg y Platts, sus *estructuras explicativas*; además, se exige que tales *estructuras explicativas* se descubran *a posteriori*. Dados estos requerimientos, una posibilidad que considera Hansberg es la de determinar *a posteriori* la estructura explicativa de los estados mentales en términos, por ejemplo, **neurofisiológicos**. Esta propuesta es formulada de la siguiente manera:

“Para que sea verdad que un estado mental ψ tenga una esencia física real ϕ debe sostenerse que una criatura no puede ejemplificar ψ sin ejemplificar ϕ ; y *viceversa*. Es decir, ‘ ψ ’ y ‘ ϕ ’ deberán ser necesariamente co-extensivos.” (Hansberg, 1990, p. 218).

Evidentemente, piensa Hansberg, la tentativa falla en virtud del fenómeno de la realización múltiple (RM).²⁶ Los estados mentales pueden, según TRM, no sólo de manera contrafáctica, sino también de hecho, realizarse físicamente de diferentes maneras. Aquí Hansberg parece tener en mente una variante de la formulación *fuerte* de TRM, según la cual, un mismo *tipo* de estado mental, p. ej., dolor, puede ser realizada por diferentes estados o bases físicas —estructuras explicativas— en diferentes individuos de un mismo *tipo* o especie.²⁷ Por ejemplo, el dolor en Juan puede realizarse en cierto tipo de tejido cerebral, mientras que en Pedro en otro muy distinto. Esta modalidad *fuerte* de TRM es suficiente para hacer imposible que un tipo de estado mental sea idéntico —necesariamente tenga la misma estructura explicativa física— en todos los mundos posibles o siquiera un mismo mundo posible. La estructura común explicativa para clases de estados mentales no puede ser física; y aunque mucho se puede descubrir *a posteriori* respecto de los correlatos físicos de instancias de estados mentales particulares, no se conseguirá nunca (*per imposible*) descubrir una esencia o estructura explicativa común para ninguna clase de estado mental. Es imposible que un estado mental ψ sea co-extensivo con un estado físico ϕ .

Así, esgrimiendo las tesis de la SE y la realización múltiple, Hansberg intenta mostrar que (i) no es posible la reducción de estados mentales a procesos físicos; (ii) que los estados mentales no conforman *clases naturales* —no tienen una estructura explicativa física *común* que se descubra *a posteriori*—; y, por lo tanto, (iii) que no pueden existir leyes psicofísicas, ya que éstas exigen que deben ser términos que denoten clases naturales los que figuren en ellas. Finalmente, es importante notar que, tomando en cuenta las consideraciones contrafácticas exclusivamente, Hansberg parece haber mostrado *a priori* que los estados mentales no son clases naturales y que, por lo tanto, los enunciados en los que aparecen no pueden ser las leyes (psicofísicas).

1.1.1. Supuestos e implicaciones. En cuanto a (i), vale la pena notar que parece suponer la concepción tradicional de reducción: la que se remonta a Ernest Nagel y que entiende la relación reductiva como una relación interteórica que se funda en la *conexión* biunívoca de los términos

²⁶ RM es la abreviación para el fenómeno de la Realización Múltiple; TRM es la abreviación para la tesis *general* que afirma que el fenómeno de la Realización Múltiple es verdadero respecto de las propiedades, estados o sucesos mentales.

²⁷ *Op. cit.*, pp. 218-9. La formulación *fuerte*, en su versión *extrema* de TRM, fue desarrollada por Fodor y, según esta, la realización múltiple de un tipo de propiedad mental es posible incluso en un mismo individuo en diferentes momentos. La formulación *débil* de TRM es la más vetusta y se remonta a Putnam. Según ésta, un *tipo* de estado mental puede realizarse en distintas bases físicas en diferentes *tipos* de organismos.

de dos teorías (predicados o aquellos que refieren a las entidades cuya existencia es asumida por cada teoría): la teoría reductora T_B y la teoría reducida T_R . La conexión de los términos de T_B y T_R , se obtiene por medio de principios conectivos que tienen forma de bicondicionales (leyes puente) y que, en su interpretación estándar, especifican una relación de co-extensividad.²⁸ Obviamente, este modelo reductivo entra en conflicto con RM, ya que exige relacionar uno a uno los términos de ambas teorías. Pero, por RM, un término mental (de T_R) habría de relacionarse con muchos términos físicos (de T_B), dando lugar a principios conectores o leyes (puente) que carecerían del carácter *estricto* que distingue las leyes científicas. Por lo tanto, en parte por esta noción particular de reducción y por RM, no podría hablarse ni de clases naturales, ni de leyes psicofísicas (iii).

Por otro lado, vale la pena decir algo acerca de RM y lo que implica para este argumento. Se piensa que la realización múltiple revela, antes que nada, una característica definitiva de los estados mentales mismos —que diferentes sustratos físicos pueden realizar el mismo tipo de estado mental—, y de manera derivativa nos dice algo sobre el estatus de los términos de estados mentales —que no son términos de clase natural. La realización múltiple nos dice que no puede existir esencia o estructura explicativa común para una clase cualquiera de estado mental; lo cual, por sí sólo, evidencia que los estados mentales no son clases naturales (ii). Ahora bien, si como lo hace Hansberg, consideramos también SE, obtenemos que los términos de estados mentales no son de clase natural, en virtud de que no existe una estructura explicativa física que se pueda descubrir *a posteriori* sobre ellos y que ayude a determinar su extensión o referencia. Ya que esto es así, se deduce también, que los términos de estados mentales no son de clase natural (ii) y, por lo tanto, que ningún enunciado en el que figuren puede ser una ley (iii). Hansberg presenta, entonces, un argumento que ejemplifica la continuidad entre las cuestiones; ‘¿cómo sabemos que una clase de objetos constituye una clase natural?’ y ‘¿cómo sabemos si un término es de clase natural?’

1.1.2. Esquema del argumento. Este argumento es la ariete con el que se abre paso Hansberg para atacar la tesis de que los estados mentales son clases naturales. Las conclusiones que obtiene son de suma importancia para este trabajo, en virtud de que implican que las emociones (en tanto que

²⁸ Véase la última cita de Hansberg en el texto y nota 22 de este trabajo.

estados mentales) no pueden ser clases naturales. Aunque no puede cuestionarse la validez del mismo, puede reflexionarse en torno a las premisas que lo sustentan. Hansberg acepta que:

1. Cualquier término que aparezca en una ley²⁹ (psicofísica) debe ser un término de clase natural.
2. Un término de clase natural se conoce porque su extensión se determina (se define) a partir de una estructura explicativa *común* —a todos los miembros de la clase— que se descubre *a posteriori*.
3. Realización múltiple (TRM en una modalidad *fuerte*): los diferentes *tipos* de estados mentales pueden realizarse en diferentes bases físicas (en diferentes individuos de una misma especie).

Como consecuencia de (2) y (3), resulta que:

4. Los (términos de) estados mentales (no refieren a) no tienen estructuras explicativas que determinen su extensión y se descubran *a posteriori*.

Finalmente, de la consideración de (2), (3) y (4) podemos concluir que:

5. Los términos mentalistas *no son* [por (2) y (4)] *ni pueden ser* [por (3)] términos de clase natural.

Y de (1) y (5) se obtiene:

6. Los enunciados en los que figuran términos mentalistas *no son, ni pueden ser* leyes (psicofísicas).

(1) es necesaria para obtener una de las conclusiones que Hansberg quiere establecer; a saber, que no pueden existir leyes psicofísicas dado que los términos de estados mentales no son términos de clase natural. Hansberg argumenta a favor de (1) en el último apartado de su artículo (Hansberg, 1990, y ya que es importante para establecer una conclusión que va más allá de nuestros intereses actuales, no se cuestionará. (2) y (3), en cambio, sí son relevantes porque son *suficientes* para mostrar que los estados mentales *no son clases naturales*: ya que, como lo afirma (4) no

²⁹ Aquí, y para el resto de este capítulo, cuando se habla de leyes, nos referimos a leyes estrictas: es decir, leyes sin excepciones, como las de la física.

tienen, ni pueden tener estructuras explicativas que se descubran *a posteriori*. No cuestionaré por ahora (3) la tesis de la realización múltiple: en este momento pretenderé no tener escrúpulos respecto de lo que las consideraciones contrafácticas/*a priori* puedan probar. Optaré por el momento por emprender una revisión de la premisa (2). Quizás esta ruta, por sí misma, nos conceda una esperanza para pensar que los estados mentales pueden ser clases naturales.

1.1.3. Sobre la construcción ‘fuerte’ de la noción de clase natural en (2). ¿Qué puede decirse de (2)? Esta premisa expresa las aspiraciones esencialistas implicadas en SE. (2) establece que los términos de clase natural se definen a partir de sus propiedades esenciales o *estructuras explicativas*; y que tales *estructuras explicativas* se descubren *a posteriori*. Sobra decir que esta es una premisa fundamental para negar a los estados mentales el estatus de clases naturales. (2) instituye un criterio de pertenencia a la clase de las clases naturales que, considerando (3), los estados mentales (como clases) no pueden satisfacer. Pero ¿cómo es que los estados mentales no pueden satisfacer (2)? Se saben muchas cosas *a posteriori* o por investigación empírica respecto de los estados mentales y sus correlatos físicos. Se puede hablar de la existencia de un cuerpo de conocimientos importante respecto de las funciones del cerebro que está en proceso de integración y consolidación. Incluso puede decirse que este cuerpo de conocimiento en desarrollo nos ha llevado a un mejor entendimiento —en comparación con lo que sabíamos hasta hace aproximadamente un siglo— de la relación entre el cerebro y nuestros estados mentales. No obstante, por (2) principalmente, no estaríamos autorizados a decir que los estados mentales son clases naturales. El punto es que, aunque existe un conocimiento *a posteriori* importante de los correlatos físicos de los estados mentales —con lo que se satisface en parte (2)—, éste no puede dar como resultado —en virtud de (3)— el descubrimiento de *estructuras explicativas* distintivas para cada clase de estado mental. Pero, ¿qué son estas *estructuras explicativas* que los estados mentales no pueden tener? Según el esencialismo metafísico de clases naturales al que ataca Hansberg, las estructuras explicativas *son*, como ya se ha dicho, *conjuntos de propiedades físicas relevantes/esenciales* comunes a todos los miembros de una clase natural. En este sentido podemos preguntar, ¿no existen en realidad *conjuntos de propiedades físicas relevantes/esenciales* comunes a todas las instancias de *miedo* —por trabajar con un ejemplo de estado mental? *No existen en realidad* conjuntos de propiedades físicas relevantes/esenciales *comunes* para las diferentes clases de estados mentales: no si se piensa que todos los casos particulares de un tipo de estado mental —reales y posibles— *deben instanciar el mismo conjunto de propiedades esenciales*. Esto es, si

se entienden tales propiedades como *condiciones claramente determinables o especificables, estables, necesarias y suficientes para formar parte de o constituir una clase natural*.³⁰ Evidentemente, nos encontramos frente a una construcción bastante **fuerte** de la noción de *estructura explicativa*: una que supone que las propiedades físicas relevantes en las cosas forman conjuntos *con límites bien definidos*; que tales conjuntos —al ser instanciados— deben mostrar *variaciones mínimas o nulas* entre sí; y que la existencia de tales conjuntos de propiedades es *necesaria y suficiente* para pertenecer a, o constituir clases naturales.³¹

Finalmente, el caso en contra de los estados mentales como clases naturales en el argumento de Hansberg, en parte, se reduce al hecho —si acaso lo es— de que no pueden existir conjuntos de propiedades físicas relevantes que sean instanciados en todos los casos particulares de una clase cualquiera de estado mental: no existe una estructura explicativa *común*, tal y como se entiende implícitamente en la premisa (2).

Pero —y a esto se reduce también la última línea de defensa de los estados mentales, y en particular las emociones, como clases naturales—, *¿por qué todos los miembros de una clase natural putativa —reales y posibles— deben instanciar las mismas propiedades físicas relevantes/esenciales?, ¿por qué habríamos de entender las estructuras explicativas o lo esencial en las clases naturales en un sentido tan fuerte?, ¿no podríamos conceder que una clase natural está constituida por miembros que instancian de forma imperfecta, vaga e indeterminada ciertas propiedades físicas relevantes/esenciales: no siempre todas las que se consideran relevantes... incluso, en algunas ocasiones, muy pocas de tales propiedades?*

El argumento de Hansberg —en particular la premisa (2)— que nos ha traído hasta estos últimos cuestionamientos parece suponer que la metafísica esencialista de clases naturales de la semántica de Kripke es suficiente para derivar respuestas interesantes para las dos primeras cuestiones y una negativa para la última. Y de nuevo, si esto es verdad, entonces Hansberg con razón habría afirmado que los estados mentales no son clases naturales. Pero el esencialismo metafísico

³⁰ De nuevo, para el caso de los estados mentales, (3) hace imposible que exista un conjunto de propiedades físicas importantes/esenciales cuya posesión sea necesaria y suficiente para formar parte de un tipo de estado mental.

³¹ Esta noción **fuerte** de las estructuras explicativas o de lo que es *esencial* en las clases naturales implica una metafísica de la realidad que, no solamente entiende que la naturaleza siempre *corta fino* en sus articulaciones ontológicas (existen límites bastante bien definidos entre las cosas), sino que lo hace de una vez y para siempre: en tanto que no es tolerante con la variación en las propiedades que se suponen esenciales, parece sugerir que las esencias o estructuras explicativas son *estáticas*. Esto, como se verá después, representa un grave problema para esta concepción fuerte de las estructuras explicativas o propiedades esenciales.

de Kripke no es la única posibilidad de metafísica de las clases naturales a la que podríamos apelar para responder a las cuestiones formuladas arriba. Aunque no en este momento, veremos que la metafísica esencialista en cuestión tiene sus propios problemas, y que existe una forma alternativa de entender la semántica de TCN y lo que son las clases naturales, que no nos compromete con una construcción de la noción de estructura explicativa tan estricta. La alternativa en cuestión entenderá la idea de *propiedades relevantes/esenciales* —estructura explicativa— de una forma más débil; de suerte que, incluso aceptando (3) la tesis de la realización múltiple de lo mental, podremos alojar la esperanza de que los estados mentales son clases naturales. La propuesta a la que aludo está más emparentada con el *realismo de clases naturales* de Putnam. Nos referimos a la semántica de TCN y la metafísica de clases naturales de Richard Boyd (1988, 1991, 1993, 1999), quien defiende una concepción de las clases naturales en términos de conjuntos de propiedades homeostáticas que son instanciadas de forma imperfecta y vaga por los miembros (reales y posibles) de una clase natural.

Dejo por ahora el primer argumento de Hansberg en contra de los estados mentales como clases naturales, subrayando que existe la posibilidad de revertir sus crípticas consecuencias si se supera la metafísica esencialista y se propone una semántica ligeramente distinta de la SE: moverse de la noción *fuerte* (Kripke) de estructura explicativa, hacia una más *liberal* (Boyd). El asunto queda por el momento aplazado.

1.2. Segundo argumento: la naturaleza semántica de los términos de estados mentales. El segundo argumento de Hansberg contra la tesis de que los estados mentales son clases naturales se basa en un análisis de la semántica de los términos mentales. Veremos también, que es la misma noción *fuerte* de estructura explicativa, aunada a una versión *fuerte* de la tesis de la realización múltiple, la que garantiza la conclusión de que los estados mentales no son clases naturales.

Según Hansberg, un análisis semántico de los *términos de estados mentales* revela que éstos no hacen referencia alguna, ni explícita, ni implícita, a estructuras explicativas o microestructuras subyacentes o a esencias físicas para cada tipo de estado mental. La aplicación correcta de un término de estado mental o la atribución correcta de un estado mental a otra persona o uno mismo, no se determina en función de que esa persona o uno mismo esté en cierto estado físico —no se hace referencia ni siquiera implícita a una esencia entendida en términos de microestructura explicativa. La correcta atribución o uso de términos mentales se determina a partir de

critérios como la posibilidad de hacer inteligible y racional la acción de otras personas. La semántica de términos mentales conecta entonces estos términos con ciertos requerimientos para la interpretación de la conducta (verbal y no verbal), y no con sucesos físicos. Más aún, por la realización variable de estados mentales, también sabemos *a priori* que las condiciones de aplicación de los términos mentales *no pueden* determinarse haciendo referencia a hechos físicos/*a posteriori*. Tenemos entonces, que lo que sabemos respecto de la semántica —las formas en que se determina la correcta aplicación— de los términos mentales lo sabemos *a priori*, y no *a posteriori* como es el caso de los términos de clase natural. Por esta razón, nuevamente, los términos de estados mentales no pueden ser términos de clase natural —como lo estipula (2)— y, en consecuencia, no serán leyes los enunciados que los contienen.

Hansberg concluye que su argumentación muestra *a priori* que las leyes psicofísicas son *imposibles* y no sólo improbables o de hecho falsas; y que los términos de estados mentales no funcionan como términos de clase natural (Hansberg, *op. cit.* 219-220).

1.2.1. Supuestos e implicaciones. Este argumento se basa en un supuesto *hecho* semántico respecto de los términos mentales. Estos tienen ‘sus propios principios distintivos de aplicación’ (*op. cit.* 219), los cuales los *conectan* con las conductas de los agentes a quienes se aplican tales términos, con la finalidad de hacer dicha conducta racional e inteligible. Dichos *principios* establecen *a priori* la correcta aplicación de los términos, y no es ningún hecho *a posteriori* —p. ej., el descubrimiento de una *estructura explicativa conectada* con estos términos— lo que determina su correcta aplicación. Sobre este argumento, es necesario señalar que el hecho *semántico* en el cual se sostiene, no es suficiente para probar que los estados mentales no sean clases naturales. Todo lo que prueba es que los ‘términos de estados mentales’ no se comportan como ‘términos de clase natural’ en la forma en la que estipula SE. Esto mismo lo reconoce Hansberg (*op. cit.* 221). Por esta razón, apuntala su argumento cancelando la posibilidad de que este *hecho o realidad semántica* de los términos de estados mentales pueda cambiar. ¿Cómo lo hace?... apelando a RM. Si la realización múltiple es verdadera de los estados mentales, entonces los términos de estados mentales (de nuevo) *nunca* podrán definir sus criterios de aplicación *a posteriori*, mediante el descubrimiento de estructuras explicativas comunes a cada tipo de estado mental. RM asegura que el *hecho semántico* nunca cambiará; o por lo menos que no se alterará de modo que

haga posible conectar los términos de estados mentales con estructuras explicativas. Al asegurar lo último, RM ya ha impedido que los estados mentales puedan ser clases naturales.

Todo esto se hará más evidente en el siguiente párrafo. Pero antes, vale la pena notar cómo, de nuevo, gran parte del peso del argumento recae en la tesis de la realización múltiple (otro tanto cae, de forma soterrada, en la construcción ‘fuerte’ de la noción de clase natural). En tanto que *hecho semántico*, que sea verdadero que los principios de aplicación correcta de los términos de estados mentales los conecten con conductas (para hacerlas racionales e inteligibles), no deja de *ser un hecho contingente*; además, por sí solo no prueba que los estados mentales no sean clases naturales. Lo que le confiere peso a este *hecho* es la realización múltiple; la cual, en sí misma, parece ser suficiente (junto con SE) para mostrar que los estados mentales no son clases naturales, además de sugerir que este argumento se basa en algo más que un *hecho semántico meramente contingente*.

1.2.2. Esquema del argumento. De nuevo, será útil develar la estructura del argumento de para detectar algún resquicio que justifique la idea de que los estados mentales son clases naturales. Comenzaré entonces por describir más detalladamente el escenario argumentativo que Hansberg ha construido:

1. Cualquier término que aparezca en una ley (psicofísica) debe ser un término de clase natural.
7. Es un *hecho semántico* que los criterios de aplicación de los términos mentales se determinan *a priori* —para responder a ciertos requerimientos que ayuden a hacer inteligibles y racionales las conductas de otras personas.
8. Cualquier forma de determinar las condiciones de aplicación de términos de estados mentales *que no sea a posteriori* y *no refiera a sucesos físicos* es insuficiente para establecerlos como términos de clase natural.
3. Realización múltiple (TRM en una modalidad *fuerte*): los diferentes *tipos* de estados mentales pueden realizarse en diferentes bases físicas (en diferentes individuos de una misma especie).

De (7) y (8) resulta que:

9. Los términos mentalistas *no son* términos de clase natural.

La tesis de la realización múltiple (3) afirma algo sumamente relevante respecto de la naturaleza de los referentes de los términos mentalistas. Según (3), no parece posible que se pueda descubrir una *estructura explicativa común* para alguna clase de estado mental: por lo tanto, nos lleva a pensar que es imposible determinar la extensión y condiciones de aplicación de los conceptos mentalistas en términos *físicos/a posteriori*: La consideración de (7), (8) y (3), entonces, justifica una afirmación modal más fuerte que (9); a saber:

10. Es *imposible* que los términos mentalistas sean términos de clase natural.

Finalmente, de (1) y (10) se puede obtener:

11. Es *imposible* que los enunciados en los que figuran términos mentalistas sean leyes (psicofísicas).

1.2.3. La construcción 'fuerte' de la noción de clase natural... otra vez. El estatus de (1) y (3) no será discutido por el momento: en el caso de (1), porque no parece relevante para nuestros propósitos; y en el caso de (3), como ya se había dicho, su análisis queda aplazado. Las aceptamos como verdaderas por el momento. En cuanto a (7), ha sido establecida (si se quiere) sólidamente. Constituye el hecho/premisa fundamental para el argumento que ahora revisamos. (8) es una implicación de la semántica esencialista de Kripke (SE) de los términos de clase natural (*supra* I, II,III, 1] y es imprescindible detenerse para analizar lo que supone.

(8) enuncia una condición esencialista que los términos de estados mentales no pueden satisfacer; no sólo por el hecho señalado en (7) —verdadero pero contingente— de que los términos mentales determinan sus condiciones de aplicación *a priori*, también y principalmente por (3) la realización múltiple. (7) y (3) afirman en conjunto, que los términos mentalistas *de hecho* determinan sus condiciones de aplicación *a priori*, y que es *imposible* que lo puedan hacer *a posteriori* —basándonos, por ejemplo, en algunas propiedades físicas relevantes. Pero, ¿qué es exactamente lo que se considera *imposible* respecto de los términos de estados mentales? *Imposible* aquí es relativo a un marco definicional. Lo que es *imposible* es encontrar una estructura explicativa *común* para cualquier tipo de estado mental. Pero, como en el argumento anterior, estructura

explicativa refiere a una forma específica de entender o definir lo que es esencial y común en los miembros de una clase natural.

De nuevo, para obtener (10) de (7) y (3), la estructura explicativa o lo esencial debe entenderse en un sentido fuerte. Es decir, la estructura explicativa o esencia se concibe como algo bastante bien delimitado, estable y, quizás, estático. Es por lo tanto, una cuestión de todo o nada: existe o no existe; se tiene o no se tiene. El *esencialismo metafísico de clases naturales* (Kripke) derivado de SE no acepta la indeterminación o límites vagos en la definición de las estructuras explicativas. Podríamos de nuevo preguntar, si esta es la única forma de entender la metafísica de las clases naturales. Para no redundar en las opiniones que se ofrecieron sobre este asunto en el argumento anterior, sólo señalaré que no es necesario *entender* lo esencial o las estructuras explicativas de las clases naturales en este sentido fuerte; que hacerlo implica una serie de problemas serios para el esencialismo; y que, siguiendo a Boyd, podemos explorar otras vías metafísicas menos estrictas para entender dichas nociones.

Finalmente, este argumento, lo mismo que el primero, puede juzgarse como devastador para la idea de que los estados mentales son clases naturales sólo si aceptamos la asunción metafísica que entiende la esencia, propiedades esenciales o estructuras explicativas como nociones perfectamente estables, bien delimitadas —en términos de sus propiedades— y estáticas —que no cambian con el tiempo .

1.3. Tercer argumento: la división del trabajo lingüístico. Otro rasgo en el que difieren los términos de clase natural de los términos de estados mentales es que, mientras que para los primeros es el caso que tiene lugar lo que Putnam llama la “división del trabajo lingüístico”, para los segundos no.³² Respecto de los términos de estados mentales no existe, según Hansberg, una comunidad de expertos a la cual acudimos para que nos ayude a decidir si aplicamos correctamente algún término mental —la realización múltiple impide que existan expertos en este sentido. Luego, los términos mentales que aparecen en nuestras descripciones psicológicas no podrán ser sustituidos por formas o estructuras explicativas descubiertas *a posteriori*.

Algunos corolarios que Hansberg deriva de este argumento son:

³² Nótese que Hansberg, aquí, alude a un rasgo de los TCN que figura en la semántica de Putnam.

(i) “Nuestras descripciones psicológicas del sentido común definen a los estados mentales *de una manera definitiva*” (*op. cit.* p. 220). Todos somos expertos cuando se trata del uso de términos de estados mentales.

(ii) “Nuestras clasificaciones de estados psicológicos no se afectan por el hecho de que descubramos realizaciones físicas distintas, ni tampoco parece que podamos descubrir una esencia física desconocida cuya revelación pudiera requerir una revisión de nuestro esquema psicológico ingenuo” (*Ibid.*).

(iii) Por su parte, los términos de clase natural *sí* parecen ser sensibles a lo que se conoce como “cambio conceptual” en virtud de hallazgos científicos los cuales son generados y custodiados por los expertos (*ibid.*).

1.3.1. Esquema del argumento. La estructura del argumento es, aproximadamente, la siguiente:

12. Existe la *división del trabajo lingüístico*: respecto de algunos términos, existe un grupo de expertos que ayudan a determinar si una muestra forma parte de su extensión a partir del conocimiento de una estructura explicativa.

13. Los términos de clase natural son susceptibles a la división del trabajo lingüístico.

(3) entraña que no es posible descubrir una estructura explicativa común para determinar la referencia de los términos mentales. Esto implica que:

14. No hay, ni puede haber, especialistas que ayuden a determinar la correcta aplicación de un término mental a partir del conocimiento de una estructura explicativa.

De (12) y (14) se infiere que:

15. Los términos mentales no pueden ser susceptibles a la división del trabajo lingüístico.

Por (13) y (15) obtenemos que:

16. Los términos mentales no son términos de clase natural.

1.3.2. De nuevo... Realización múltiple y una noción ‘fuerte’ de clase natural. De (12) puede decirse que es un hecho lingüístico establecido. (13) tiene mucho sentido a la luz de SE, toda vez

que SE establece que los términos de clase natural determinan su referencia a partir de estructuras explicativas y, en la medida en que tales estructuras explicativas son conocidas en sus detalles por expertos, es sensato pensar que los términos de clase natural son susceptibles al fenómeno descrito en (12) —podríamos, por ejemplo, considerar el *realismo de clases naturales* de Putnam—*división del trabajo lingüístico*. Para abordar y analizar (14) vale la pena antes señalar que el argumento es más significativo de lo que aparenta. No sólo hay que atender a lo que prueba —que los términos mentales no se comportan como términos de clase natural—, sino a cómo lo prueba. Lo que distingue en este argumento los términos mentales de los de clase natural es, en primera instancia, que los primeros no son susceptibles a la división del trabajo lingüístico mientras los últimos sí los son; en última instancia —y más importante—, esto se debe a que —de nuevo por (3)— resulta *imposible* determinar una estructura explicativa para los estados mentales que pueda ser descubierta *a posteriori* y conocida por los expertos de la comunidad lingüística. Este hecho, si lo es, subyace al hecho lingüístico probado por Hansberg: a saber, que los términos mentales no son términos de clase natural. Ahora, si reflexionamos sobre (14) —indispensable para probar (15) y, por transitividad, (16)—, notaremos que, de nuevo, depende de (3) la realización múltiple y de la suposición de SE que insiste en entender las estructuras explicativas en un sentido *fuerte*: la estructura explicativa se entiende como algo bien delimitado, estable e inmutable. Construida de esta forma la noción de estructura explicativa, y tomando en cuenta (3), es lógico pensar que los estados mentales no pueden tener estructuras explicativas comunes; luego, ante tal imposibilidad, como señala (14), no puede haber expertos que ayuden a determinar la aplicación de un término mental: “todos somos expertos”, como señala Hansberg.

III.II. CONCLUSIONES: UN PAR DE SUPUESTOS ESENCIALES EN EL ESENCIALISMO DE CLASES NATURALES QUE SUPONE Y CRITICA HANSBERG

Como se apuntó al inicio de este capítulo (I, III) los argumentos de Hansberg que hemos revisado son valiosos en la medida en que ayudan a hacer explícitos todos los supuestos e implicaciones detrás del *esencialismo metafísico de clases naturales* basado en la SE: más provechosos resultan para nuestros propósitos en la medida en que su finalidad es proyectar tales supuestos al tema de los estados mentales en tanto que clases naturales: lo que implica que los resultados se aplican por extensión a las emociones. Después del análisis precedente, podemos permitirnos pensar en

un balance final. ¿En qué estatus dejan a los estados mentales —incluidas las emociones? ¿Muestran definitivamente que no son clases naturales?

Lo que las reflexiones preliminares nos autorizan a decir es que los argumentos en cuestión en contra de que los estados mentales son clases naturales descansan principalmente en una tesis explícita y una implícita. La primera es *(a)* una formulación *fuerte* de la tesis de la realización múltiple, según la cual un mismo *tipo* de estado mental puede ser instanciado, realizado o implementado en diferentes bases físicas; en diferentes *tipos* estructurales (pulpos, marcianos, robots y seres humanos) y en diferentes individuos de un mismo *tipo* estructural (Juan, Pedro, Ana, ...). La tesis implícita alude a *(b)* una construcción demasiado *fuerte* de la noción de ‘estructura explicativa’ o ‘esencia’ al interior del esencialismo metafísico de las clases naturales. En conjunto, son suficientes para mostrar que los estados mentales no son clases naturales.³³ En este sentido, la revisión del resto de las premisas en los argumentos de Hansberg nos exige considerarlas como superfluas. Por sí solas no pueden probar que los estados mentales no son clases naturales; precisan de las tesis que recién hemos identificado para llegar a tal conclusión [*(a)* y *(b)*].

Dispuestas así las nociones, podemos responder que los argumentos de Hansberg muestran definitivamente que los estados mentales no son clases naturales siempre y cuando *(a)* y *(b)* sean verdaderas. Pero, como se sugirió también al inicio de este capítulo, tal vez la forma en la que los argumentos de Hansberg exigen que interpretemos la noción de *estructura explicativa* —en el marco del *esencialismo metafísico de las clases naturales*—, no es la más adecuada; y quizás tampoco deberíamos ser acrílicos respecto de lo que el argumento o tesis de la realización múltiple parece probar. Si es posible objetar o poner en entredicho la necesidad o verdad o corrección de estas dos tesis entonces podríamos razonablemente decir que los argumentos de Hansberg no decide de forma definitiva el estatus de los estados mentales en tanto que clases naturales. Lo que implica que no deciden, tampoco, que las emociones no son clases naturales.

Lo que se intentará en el próximo capítulo es oponer algunas consideraciones a las tesis de la realización múltiple y proponer una alternativa a la construcción *fuerte* de la noción de *estructura explicativa* implicada en el esencialismo metafísico de clases naturales. En la medida en que se logre debilita la fuerza intuitiva de la Tesis de la Realización Múltiple y proponer una semántica de TCN y una metafísica de las clases naturales menos fuerte, podremos abrigar la

³³ Y no sólo que los términos mentales no son términos de clase natural.

esperanza más general de que los estados mentales son clase naturales, y también la más particular de que las emociones lo son también.

IV. REALIZACIÓN MÚLTIPLE Y CONSTRUCCIÓN FUERTE DE LA NOCIÓN DE ESTRUCTURA EXPLICATIVA: CRÍTICA Y REVISIÓN

IV.I. REALIZACIÓN MÚLTIPLE: ¿QUÉ PRUEBA?

Los argumentos de Hansberg estudiados en el capítulo anterior dependían fundamentalmente de la tesis de la Realización Múltiple. En este apartado plantearemos algunas objeciones respecto de la forma en la que habitualmente suele interpretarse tal tesis o fenómeno de la realización múltiple y de lo que supone probar.

I. ¿UNA TESIS A PRIORI?

El primer rasgo inquietante en la Tesis de la Realización Múltiple (TRM para abreviar) consiste en que a veces es tratada, no como un hecho empírico —o no solamente empírico—, sino como algo que puede mostrarse *a priori* (Kripke, 1981). Así, en su primer argumento, cuando intenta mostrar que no es posible encontrar una estructura explicativa común para los estados mentales, Hansberg señala que:

“... no es siquiera necesario que observemos casos reales de realización alternativa; bastaría con que existiera la posibilidad contrafáctica de realizaciones variables.” (Hansberg 1990, p. 218).

Más adelante en la misma página continúa:

“Supongamos que el correlato físico del dolor fuera el estado cerebral *B*. Definimos al dolor como el estado cerebral *B*, pero esta definición no puede fijar la extensión del término ‘dolor’ porque *podría* haber una sensación que caracterizáramos como dolor, esto es, que es dolorosa y que, sin embargo, no fuera un estado cerebral *B*.” (*Ibid.*)

En primer lugar, debemos notar que Hansberg no muestra mucho interés por la Realización Múltiple (RM) entendida como un hecho empírico. Esto se explica porque desea mostrar *a priori* que es imposible que los estados mentales sean clases naturales. En caso de entender la TRM como una tesis empírica, que los estados mentales no son clases naturales resulta ser un

hecho quizás verdadero, pero contingente. Hansberg no quiere conceder cuartel a los estados mentales y aspira a la afirmación modal más fuerte.

Hansberg, entonces, parece suponer que TRM es una tesis que se puede establecer *a priori*, que se deriva del análisis del concepto ‘dolor’. Básicamente, se trata de una variante de los argumentos modales —que se remontan a Descartes y llegan hasta Kripke— que pretenden, a partir de una distinción conceptual, derivar una distinción ontológica³⁴. Pero, un problema que se puede imputar a las estrategias modales consiste en que, probablemente, las conclusiones que se obtienen mediante el análisis conceptual que involucran, terminarán reflejando *sólo* la forma en la que funcionan nuestros conceptos y las implicaciones anidadas en ellos. Obviamente, esto no garantiza ninguna conclusión categórica sobre la naturaleza del referente. De algún modo, apelar a contrafácticos en la forma en la que Hansberg lo hace no es más que apoyar su argumento en la *posibilidad de concebir* casos en los que el dolor se realiza en distintas bases físicas. En otras palabras, el concepto ‘dolor’ abre un espacio de maniobras conceptuales que permite llegar a concluir que el dolor debe distinguirse de estados físicos-cerebrales. No se puede establecer de este modo una tesis que pretende afirmar un hecho (rasgo) *empírico* respecto de un tipo de fenómenos —estados mentales— como lo pretende la tesis de la Realización Múltiple. Casi cualquier cosa podría probarse así; incluso tesis que difícilmente aceptaría un seguidor de SE.

Piénsese en lo siguiente. Tenemos una cantidad n de muestras de un líquido que parece agua. Después de algunos análisis de laboratorio, encontramos que las muestras n_1 y n_2 tienen una compleja constitución química (XYZ) que no es H_2O . Teniendo presente la SE, tendríamos que decir que n_1 y n_2 no son muestras de agua: sin importar que satisfagan todas las propiedades superficiales que asociamos con el agua. La cuestión es, ¿por qué no tenemos escrúpulos para descalificar como muestras de agua a n_1 y n_2 ; y sí los tenemos para descalificar como muestras de dolor a *posibles* casos que no realizan —para usar el ejemplo citado de Hansberg— el estado cerebral B ? ¿Por qué no se presenta como problemático desdeñar los rasgos superficiales en el caso de las muestras de los líquidos que parecen agua (n_1 y n_2); y sí lo es en el caso de los rasgos *fenomenológicos* de las personas que reportan dolor? Quizás Hansberg nos debe una explicación de por qué considera ‘inaceptable’ (*ibid.*) decir de las personas que reportan dolor y no realizan B , que no sienten dolor; mientras que, por el contrario, no se opondría a decir de las muestras n_1 y

³⁴ Una *distinción real*, según Descartes.

n_2 que no son agua. Si no se ofrece una justificación para esta asimetría al aplicar los principios de la semántica esencialista (SE), no habría nada que nos impidiera concluir, considerando el caso de las muestras de líquidos que parecen agua y siguiendo la misma estrategia de Hansberg, que la realización múltiple también es verdadera respecto del agua; y esta, es una conclusión que nadie que suscriba SE admitiría³⁵. La apelación a contrafácticos puede establecer *a priori* la realización múltiple respecto de los estados mentales pero a un precio muy alto; a saber, que de clases de cosas como el agua, por ejemplo, también es verdadera TRM.

Una línea de respuesta, señalada por Shapiro (2000, p. 643), consistiría en señalar que, mientras los estados mentales constituyen una clase que se define por referencia a sus propósitos, capacidades o contribución para algún fin, el agua es una clase que se define por su estructura o composición física. En este sentido, los propósitos, las capacidades y las contribuciones para ciertos fines pueden, en principio, ser realizadas por distintas clases de cosas. Luego, TRM pretende tener aplicación sólo a las clases que se definen de esta manera —*carburadores, ratoneras, computadoras, mentes*—, y no a clases que se definen estructuralmente como el agua. El agua no puede considerarse un candidato a la RM porque estructuras micro-físicas (H_2O) no pueden ser múltiplemente realizadas.

Pero varios hechos se infradescriben en esta réplica. Las clases que se definen por referencia a las estructuras micro-físicas son relativamente nuevas. Hasta hace poco más de un siglo, ni el agua, ni la sal, ni el oro tenían definiciones de esta índole. De hecho, sus definiciones eran primordialmente disposicionales. Las definiciones por referencia a la estructura apuntan a un conjunto de clases —sustancias químicas— de las cuales la reducción interteórica tuvo éxito recientemente. Antes de tal reducción, deberíamos reconocer, estarían dentro del alcance de la Realización Múltiple, ya que sus primigenias (aunque no olvidadas) propiedades disposicionales —en

³⁵ La justificación estándar para la asimetría apela a la naturaleza de los estados fenoménicos, como el de dolor. Se suele pensar que tener tales estados ayuda a captar de forma directa y clara la naturaleza y rasgos esenciales de su referente. Así, muchos filósofos suponen que tener un dolor implica conocer los rasgos esenciales del dolor, a saber, cierta fenomenología. En esta dirección, se concluye que tales rasgos esenciales (fenomenología) son innegociables y, por tal razón, ante la posibilidad (conceptual) de disociar el dolor (su fenomenología) de los estados físicos con los que se pretende identificar, se les confiere preeminencia a dichos rasgos. En otras palabras, nuestros conceptos nos dicen que no hay contradicción en afirmar que los estados mentales no implican propiedades físicas, pero sí la hay al negar que tengan una fenomenología característica. En contra de esta forma de entender las implicaciones de los conceptos fenoménicos, se ha conformado lo que se conoce como 'la estrategia de los conceptos fenoménicos'. Según esta estrategia, defendida notablemente por autores como Loar (1990), Tye (2000) y Papineau (2002), los conceptos fenoménicos no ostentan la inmediatez y perspicuidad que habitualmente se les concede, de lo que concluyen que la fenomenología no debe considerarse un rasgo esencial o innegociable de muchos estados mentales.

términos de las que eran definidas— también son susceptibles de ser múltiplemente realizadas: congelarse a 0° Celsius y hervir a 100° Celsius son propiedades que bien pueden ser instanciadas por diferentes sustancias. Agua, definida disposicionalmente, es entonces candidata para RM y la ‘estrategia contrafactual’ *a priori* igual puede probarlo. Se podría replicar que, dado lo que *sabemos ahora* sobre la composición química del agua y otras sustancias, ya no tiene caso especular sobre tiempos en los que para tales sustancias teníamos definiciones disposicionales. Pero con esta réplica se termina por aceptar que es una cuestión *empírica* y no *a priori* la que determina si una clase de objetos o fenómenos es susceptible o (ya) no lo es de RM. Contar con una definición basada en la composición física de una clase de cosas o fenómenos es contar con hallazgos y datos *empíricos* suficientes para dicha definición. Tener una definición por referencia a capacidades y propósitos para una clase de cosas o fenómenos apunta a la carencia de hallazgos y datos empíricos suficientes para una definición estructural de dicha clase. Finalmente, aunque de forma negativa, establecer si una clase es *susceptible*³⁶ de RM se determina *empíricamente*; y no *a priori* como lo pretenden establecer las posibilidades contrafácticas que avanza Hansberg.

Estas consideraciones ponen en tensión la postura de Hansberg que intenta establecer *a priori*, por medio de contrafácticos, que TRM es verdadera respecto de los estados mentales. Por un lado, la *mera* estrategia contrafactual conduce a aceptar que de una sustancia como el *agua* también es verdadera RM, toda vez que es posible pensar que haya una sustancia con las propiedades superficiales del agua pero con otra composición física (realizador). Por otro lado, la distinción entre definiciones por referencia a propósitos y capacidad, y las que se basan en la composición física, que tiene como finalidad negar que ‘agua’ sea susceptible de RM, implica el reconocimiento de que la realización múltiple se establece *empíricamente*, y no *a priori*.

Por otro lado, hemos de considerar lo que evidencian los contrafácticos en el caso de RM: esto es, ¿nos dicen algo sobre la naturaleza de los estados mentales o sobre lo que sabemos o entendemos de éstos? Considérese la distinción que sugiere Nagel (1986/1998, cap.II) entre *posibilidad epistémica* y *posibilidad metafísica*. La primera refiere a lo que podemos concebir acerca de las cosas en virtud de lo que conocemos o entendemos de ellas (nuestros conceptos de ellas). La última refiere a lo que pueden o no pueden ser las cosas en virtud de su naturaleza. Encuentro

³⁶ Nótese que para una clase de cosas o fenómenos el ser *susceptible a RM* ha de entenderse como un hecho *contingente* y, quizás, *transitorio* respecto de dicha clase, la cual puede estar en espera de mejores teorías e instrumentos para reconocer su constitución o estructura física reales.

muy poco probable que las posibilidades que se exploran mediante estrategias contrafactuales sean de índole metafísica; esto implicaría, de algún modo, la poco plausible tesis de que podemos conocer *a priori* rasgos fundamentales de las cosas (de lo que metafísicamente pueden o no pueden ser). Un poco de mesura epistemológica nos debería llevar a reconocer que no debe ser este el caso. Resultará más bien que las elucubraciones contrafácticas deben revelar rasgos de nuestros conceptos de las cosas... de lo que conocemos y entendemos acerca de ellas. Los contrafácticos de realización múltiple de estados mentales no evidencian más que las implicaciones anidadas en nuestro —escaso— conocimiento de los estados mentales (conceptos mentales). Pero esto no las vuelve posibilidades metafísicas: en otras palabras, no dice nada acerca de lo que *pueden o no pueden ser* de hecho los estados mentales. Si esto es así, Hansberg incurre en un error al pensar que los contrafácticos pueden establecer una tesis empírica (TRM) sobre los estados mentales.³⁷

Finalmente, no parece estar justificada la pretensión de establecer *a priori* un hecho sobre el mundo físico como la realización múltiple. En tal caso, debemos reconocer que no está firmemente sustentada la afirmación modal de Hansberg que reza que es *imposible* que los estados mentales tengan estructuras explicativas comunes y sean, por lo tanto, clases naturales. Naturalmente, esto no establece que sí lo sean. Todavía la tesis RM puede ser establecida por medio de casos reales de realización múltiple y, en este sentido, está garantizada la afirmación, menos fuerte y fáctica (no modal), de que los estados mentales no *son* clases naturales, en virtud de que no es empíricamente plausible —por el fenómeno (hecho) de la realización múltiple— que existan estructuras explicativas comunes para ellos. En el siguiente apartado, se pretenderá objetar incluso esta afirmación factual menos fuerte avanzando algunos cuestionamientos a la forma (acrítica) en la que es habitualmente interpretada la tesis de la RM.

2. UN DILEMA PARA LA REALIZACIÓN MÚLTIPLE

Lawrence Shapiro (2000) considera que la Tesis de la Realización Múltiple (TRM) funciona bien como argumento en contra del reduccionismo psico-físico en la medida en que no se le somete a un análisis minucioso. La primera cuestión que uno debería plantear respecto de TRM tendría que inquirir por las condiciones necesarias y suficientes para un caso de RM: ‘*N* y *M* son distintas

³⁷ Lawrence Shapiro (2000) también intenta evidenciar el escaso o nulo apoyo que aporta para establecer la tesis empírica de la RM, el argumento aparentemente *a priori* de Putnam basado en el isomorfismo funcional.

realizaciones de *T* cuando y sólo cuando _____' (Shapiro, 2000, p. 636). Shapiro persevera en esta cuestión y entiende que, si nos detenemos un poco, notaremos que no es tan simple decidir si un par de realizadores *aparentemente* distintos implementan de formas diversas una clase de objeto o fenómeno.

Podríamos estar inclinados a pensar que un par de sacacorchos estructuralmente similares, pero distintos en composición —uno es de aluminio y otro de acero—, realizan de forma diversa la clase *sacacorchos*. Quizás tendríamos más reservas respecto de dos sacacorchos similares en estructura y composición que sólo difieren en el color. Los ejemplos muestran que la TRM no tiene límites de aplicación claros; por lo tanto exigen un análisis que nos guíe hacia el reconocimiento de tales límites y nos faculte para dictaminar qué casos hacen verdadera la TRM.

En su análisis, Shapiro comienza por señalar que las clases de cosas o fenómenos que son susceptibles de RM —que pudieran hacer verdadera TRM— son aquellas que se definen por referencia a sus capacidades, propósitos y/o contribución a un fin: *sacacorchos*, *ratoneras*, *computadoras*, *carburadores* y *mentes* son las clases que habitualmente ejemplifican este tipo de capacidades. Enseguida Shapiro destaca un punto fundamental; a saber, que *sólo algunas de las propiedades de los realizadores de las clases que son susceptibles a la realización múltiple son relevantes —y otras no— para el propósito, actividad o capacidad en función de la cual se definen dichas clases* (op. cit., p. 643-44).

Dispuestas así las nociones, respecto del ejemplo de los dos sacacorchos similares en constitución y estructura, pero diferentes en color, debemos responder que no han de considerarse como realizaciones distintas de la clase *sacacorchos* porque el color no es una propiedad importante para la función o propósito para el que son hechos los sacacorchos. El color no es una propiedad que contribuya a su propósito de remover tapas de corcho.

¿Qué hay respecto de dos... uno hecho de aluminio y otro de acero? ¿Representan un caso de RM? Para responder debemos preguntarnos lo siguiente: ¿el sacacorchos de acero y el de aluminio *difieren en las propiedades causales relevantes* para realizar la función características de los *sacacorchos*? La respuesta debe ser 'No'. Tienen las mismas propiedades relevantes para realizar la función en cuestión (rigidez, si se quiere). No son, por lo tanto realizadores distintos de la clase *sacacorchos* porque, en relación con las propiedades causales relevantes que permiten que

funcionen como tales, son idénticos. El hecho de que uno sea de acero y otro de aluminio no constituye una mejor razón para llamarlos realizadores múltiples de una clase que el hecho de que sean de diferentes colores.³⁸

Teniendo presente estas consideraciones Shapiro plantea el siguiente dilema para la TRM. Considérese un caso posible de RM: un caso en el que dos clases de objetos supuestamente hacen lo mismo pero de formas muy diferentes. En tal caso, cada clase de realizador difiere de hecho en sus propiedades relevantes o no lo hace. Si no lo hace, como en el caso del ojo de pulpo y el ojo de mamífero —ambos ejemplifican lo que se denomina *cámara ocular*³⁹—, entonces no se trata de realización múltiple. Si en efecto difieren en sus propiedades relevantes, como lo puede ser en el caso del ojo del mamífero —cámara ocular— y el de un artrópodo —ojo compuesto—, entonces, lo que tendríamos realmente serían *dos clases diferentes*: no una misma clase y, en este caso, tampoco tenemos un caso de RM —de *una sola* clase con distintos realizadores (*op. cit.*, p. 646-7).

No obstante, podría replicarse, tendemos a agrupar diferentes *clases* de realizadores bajo una misma clase funcional —ojo, ratonera— a partir de similitudes, ya que sobre estas se pueden construir leyes o proponer generalizaciones en ciencias de nivel superior —abstrayendo de propiedades micro-estructurales. Pero Shapiro responde que, cuando dos *clases* de realizadores difieren significativamente en sus propiedades causalmente relevantes, las leyes y generalizaciones que se pueden derivar sólo de las similitudes funcionales pueden ser bastante *anodinas* (*'dumbingly dull'* *op. cit.*, p. 649).

Piénsese de la siguiente forma. Las ratoneras están hechas para atrapar ratones (vivos o muertos); los ojos de los mamíferos y los ojos de los artrópodos están hechos para ver. Si las *clases* de realizadores —p. ej., ojo de pulpo y ojo de mamífero— comparten la mayoría de las propiedades relevantes entonces no son diferentes realizadores. Si no comparten las propiedades

³⁸ Un caso de realización múltiple se obtiene si consideramos un sacacorchos simple que utiliza el principio de la palanca y otro que utiliza el sistema de cremallera y piñón. En este caso debemos reconocer que ejemplifican distintos mecanismos, que exigen diferentes formas de manipulación y que obedecen a diferentes principios y/o leyes. Sus propiedades relevantes —para ejecutar el propósito característico de la clase— difieren y, por lo tanto, pueden considerarse como realizadores distintos de la clase (*op.cit.*, p. 644).

³⁹ En el análisis de Shapiro, el ojo del mamífero y el del pulpo son estructuralmente muy similares. Las propiedades causales relevantes para desarrollar la función de la visión son las mismas. En ambos casos encontramos lentes que generan una imagen invertida en la retina que luego es transformada en señales eléctricas. La evolución convergente diseñó, esencialmente, la misma clase de estructura funcional (ojo) utilizando los mismos mecanismos causales de realización.

causales relevantes —p. ej., cámara ocular y ojo compuesto—, entonces no habrá, o de haberlas serán muy pocas, las leyes (no *anodinas*) que sean válidas para las dos clases de realizadores. Así, el primer cuerno del dilema nos conduciría a reconocer que tenemos sólo una clase funcional que no es múltiplemente realizada. El segundo socava la razón principal —formulación de leyes y generalizaciones— que tenemos para agrupar clases de realizadores realmente diferentes bajo una *misma* clase funcional.

Finalmente, atacar el reduccionismo psico-físico y defender la autonomía de las ciencias especiales por medio de la TRM,⁴⁰ nos conduce a enfrentar el dilema que hace explícito Shapiro. Un cuerno implica que, quizás, muchas *clases* de realizaciones de una *clase* de objeto o fenómeno que en principio consideraríamos como distintas, resultan no serlo, en virtud de que no difieren realmente en sus propiedades causalmente relevantes: como en el caso de los sacacorchos de colores y de distintas composiciones; y el de los ojos de los mamíferos y los pulpos. El otro cuerno del dilema tiene la grave consecuencia de que, en caso de aceptar la realización múltiple sobre una *clase funcional* —*carburador, ratonera, mente, ojo*—, el género de leyes y generalizaciones que se pueden obtener sobre ellas es bastante pedestre: pudiendo ser tan distintas en sus propiedades causales relevantes las distintas clases de realizaciones, existirán muy pocas leyes y generalizaciones comunes a tales realizaciones; en este caso tendríamos que conformarnos sólo con unas cuantas y muy poco específicas.

3. IDENTIDAD FUNCIONAL Y REALIZACIÓN MÚLTIPLE

Mark Couch reconoce que, para establecer un caso de RM, es necesario (i) mostrar que los supuestos realizadores físicos son, en efecto, diferentes; y (ii) que las propiedades funcionales que supuestamente son múltiplemente realizadas son las mismas (Couch, 2004, p. 202). Las objeciones de Shapiro descritas en el apartado anterior se dirigen principalmente a (i): sugieren que, en muchos casos, no son en realidad distintos los realizadores físicos de una clase funcional, toda vez que comparten las mismas propiedades causales relevantes. Naturalmente, otra forma de cuestionar la TRM consiste en mostrar que (ii) no es el caso: esto es, objetar la premisa de que las propiedades funcionales realizadas por dos estructuras físicas distintas sean las mismas. Para tal efecto Couch considera el caso del ojo humano y el ojo del pulpo. Estos ejemplifican un caso de evolución convergente; sin un ancestro común, a partir de presiones adaptativas desarrollaron

⁴⁰ Ambos objetivos evidentes en el trabajo de Hansberg (1990).

mecanismos de visión *similares*. Pero sólo *similares* en un nivel muy abstracto. De hecho, insiste Couch, un análisis anatómico muestra diferencias estructurales cruciales. Aunque de hecho ambas clases de ojos se describen en un sentido *general* como cámaras oculares, con cristalino (lentes) y retina, difieren en los siguientes aspectos: mientras el ojo del pulpo tiene tan sólo un tipo de pigmento visual en sus foto-receptores, el ojo humano tiene cuatro, lo que hace al primero insensible al color; mientras la retina humana enfoca la luz por medio de un mecanismo que flexiona el cristalino (lente) para cambiar su forma, la retina en el pulpo utiliza un mecanismo que empuja el cristalino hacia adelante y hacia atrás para enfocar la luz. Así, los componentes de cada ojo son diferentes y *funcionan* de formas distintas, dando lugar a diferencias en los estímulos que procesan y las respuestas a que dan lugar. El ojo del pulpo es incapaz de ver colores pero logra percibir el espectro de luz polarizada; además, la estructura de la retina del pulpo sugiere que el procesamiento visual en el cerebro se realiza de manera distinta en cada especie: en el pulpo el procesamiento es menos complejo por las estructuras y pasos que parecen implicados.⁴¹ Finalmente, las diferencias estructurales apuntan a divergencias funcionales; esto es, involucran respuestas a diferentes tipos de estímulos, velocidades de reacción distintas, etc. La conclusión de Couch es que, a pesar de la *similitud* en las funciones de las estructuras oculares, es plausible sostener que las funciones de cada clase de ojo son de hecho *distintas*. La TRM exige, más que *similitud* de funciones, *identidad* de funciones; en este sentido “que los ojos tengan funciones *similares* no significa que tales funciones *sean idénticas*.” (*ibid.* Las cursivas son agregadas). Así dispuestas las nociones, tendríamos que pensarlo un par de veces antes de afirmar que *una misma propiedad o mecanismo funcional* —ojo en el caso presente— es múltiplemente realizada. Queda como una cuestión —*empírica*— abierta determinar en cada caso si en realidad se trata de *una misma función*.

Couch pretende, con estas observaciones, ofrecer un ejemplo de cómo cuestionar la TRM atacando (ii) la idea de que existe una *sola* propiedad funcional realizada por sistemas o mecanismos físicos distintos. Aunque las reflexiones que suscitan los planteamientos de Couch son de suma relevancia, pocas veces se persiste en ellas. Es habitual hablar holgadamente de *una misma propiedad psicológica* realizada en distintas bases físicas —en individuos o especies— sin reparar en la cuestión de si, en efecto, se trata de la misma *propiedad o mecanismo funcional*. Por lo regular, las definiciones de tales propiedades o mecanismos alcanzan registros de abstracción

⁴¹ Bechtel y Mundale (1999, p. 197).

bastante elevados y se basan en distinciones psicológicas ordinarias. En contextos definicionales tan amplios, resultará bastante propicio que funciones *similares* se tomen por *idénticas* y que, por lo tanto, se concluya que una *misma* propiedad funcional es susceptible de RM.

En otros términos, los filósofos tienden a señalar diferencias de tipos neuronales que subyacen a un *mismo* tipo de estado psicológico: por un lado confían en que la indagación *empírica* mostrará de forma *detallada* que habrá diferentes clases de realizadores físicos —para un estado psicológico—; por otro lado, no parecen encontrar ningún problema en la individuación de los estados psicológicos a partir de intuiciones psicológicas bastante *generales*. No es habitual reparar en una *problemática de la individuación* de los estados psicológicos; no en contextos en los que se alude a la tesis de la realización múltiple para cuestionar pretensiones reduccionistas. Pero, el punto sobre el que hace presión Couch, es el de que muy probablemente la individuación de las propiedades funcionales —léase estados psicológicos— es también una cuestión sobre la que las especificaciones *detalladas* y los hallazgos *empíricos* son muy importantes. Sobre este particular, Couch refiere a la cautela con la que los psicólogos hablan, por ejemplo, de estados psicológicos compartidos por seres humanos y pulpos. Al sugerir tales propiedades comunes, los psicólogos tienen muy claro que son posibles diferencias cruciales —entre los estados psicológicos aparentemente idénticos de un ser humano y un pulpo— que pueden hacerse evidentes en la medida en la que las definiciones basadas en los rasgos relevantes funcionales —del estado psicológico— sean más *detalladas* y específicas. Más aún, “al detectar diferencias neuronales importantes entre especies, los psicólogos se inclinarán por considerar propiedades funcionales distintas... reconocen que a un nivel *más detallado*, las diferencias neuronales conllevan diferencias funcionales.” (*op. cit.* p. 203).

Recapitulando, las observaciones de Couch nos demandan observar con recelo cualquier afirmación *fácil* de TRM. Por las razones esgrimidas en este apartado, es probable que muchos casos en los que se presume RM —incluyendo los estados mentales—, la indagación empírica termine por persuadirnos de que, en realidad, se trata de propiedades y/o mecanismos funcionales distintos que se realizan por medios similares.

4. INTEGRANDO LAS OBSERVACIONES A LA LUZ DE LA NEUROFISIOLOGÍA

En conjunto, las consideraciones de Shapiro y Couch revelan que la Tesis de la Realización Múltiple funciona bien como trinchera defensiva en contra del reduccionismo, sólo si es entendida en términos meramente *intuitivos* y bastante *generales*. Si en general cualquier diferencia estructural es suficiente para hablar de realizaciones distintas; y si en general, de dos propiedades o mecanismos funcionales que se implementan de formas similares, decimos que realizan la *misma función*, entonces resultará bastante natural dar el siguiente paso y sostener que se trata de un caso de realización múltiple. Pero TRM comienza a perder lustre cuando, como lo hacen Shapiro y Couch, se reflexiona sobre la *supuesta diversidad* de los realizadores o la *identidad* de las propiedades o mecanismos funcionales en cuestión. Ambos autores, aunque desde flancos diferentes, aportan razones para pensar que TRM requiere muchas más aclaraciones y especificaciones de las que sus prosélitos se exigen al esgrimirla. Más aún, en un sentido positivo, sugieren que el destino de TRM respecto de los casos de los que se dice verdadera se debe decidir, indefectiblemente, de manera *empírica* y no *a priori*. No deberíamos simple e irreflexivamente *suponer* que dadas ciertas diferencias aparentes dos realizadores son *distintos*. Como indica Shapiro, determinar la diversidad requiere de *datos empíricos* acerca de los realizadores y de un *análisis* de los mismos que nos faculte para tomar alguna resolución.⁴² Por otro lado, tampoco podemos simplemente *suponer* que dos propiedades o mecanismos funcionales son *idénticos* por el mero hecho de que se realizan de manera *similar*. También aquí, como señala Couch, los datos empíricos y el análisis de los mismos jugarán un papel fundamental para determinar si se trata de la *misma* propiedad o mecanismo funcional. Cualquier presunto caso de RM tiene que establecer, en primer lugar, la *identidad de funciones* y, en segundo lugar, la *diversidad de realizadores*. Hemos visto que las intuiciones en las que se suelen basar tales *identidades* y *multiplicidades* se tambalean si se les presiona un poco. El problema serio para TRM se hará patente en caso de que se descubra —después de analizar la evidencia empírica disponible, si la hay— que no existen muy buenas razones para hablar de *realizadores distintos* o de *una misma función*; y esto es relevante para cualquier presunto caso de RM porque es algo que, como lo sugieren los casos que examinan Shapiro y Couch, en efecto sucede.

⁴² Como lo ha evidenciado Shapiro, en este caso la resolución puede tener un costo muy alto en términos explicativos.

En resumen, los argumentos de Shapiro y Couch reivindican un poco de esperanza. Después de todo, la moneda parece seguir en el aire. TRM, en contra de lo que se piensa, no ha logrado mostrar que los estados mentales son físicamente irreducibles. No en tanto que no satisfaga los requerimientos (i) y (ii). Respecto del caso que nos interesa, los estados mentales y en particular las emociones, aún tiene que establecerse en cada caso que los supuestos realizadores diversos son en efecto diferentes: que tienen propiedades causales distintas y que, por lo tanto, realizan de formas diversas la misma función psicológica. En el otro extremo, tiene que establecerse también que los distintos realizadores (en caso de serlo), implementan la *misma* función psicológica, y no sólo una función *similar*.

Estas cuestiones, desde el ángulo que se las vea, siguen abiertas. La neurofisiología las aborda directa y *empíricamente*, y sus avances más recientes parecen sugerir una imagen de la relación entre estados psicológicos y estados cerebrales bastante distinta a la que propone la TRM. La neurofisiología, como disciplina *empírica*, habría de suspender sus protocolos de investigación o considerar desesperadamente ingenuas sus pretensiones de conocer los mecanismos cerebrales que subyacen a los mecanismos psicológicos si se atuviera a lo que TRM pretende probar. Pero este no aparenta ser el caso. Por el contrario, puede decirse que la neurofisiología ha logrado mucho al suponer que existe *continuidad* en los mecanismos neuronales que subyacen a las funciones psicológicas, no sólo entre diferentes individuos, sino también entre especies. En este sentido, Bickle (2006; 2.3) apunta que la neurofisiología plantea retos y cuestionamientos importantes para TRM: ¿cómo podría ser posible nuestro conocimiento del sistema visual humano, a partir del estudio del cerebro de otros primates, si TRM afirma algo verdadero respecto de los estados y capacidades psicológicas y sus *radicalmente diversos* correlatos físicos?; ¿cómo es que la tomografía por emisión de positrones y la resonancia magnética funcional revelan actividad metabólica intensa en regiones cerebrales comunes, no sólo entre individuos de la misma especie, sino entre individuos de especies diferentes, al realizar determinadas tareas?

Bickle también observa que el éxito de la neurofisiología implica que las capacidades y estados psicológicos no son, después de todo, tan variables en su realización. En este sentido, incluso el fenómeno de la plasticidad neuronal tiene que entenderse como un proceso mucho más sistemático y restringido de lo que habitualmente se piensa. En casos de daño cerebral, en efecto, en virtud de la plasticidad neuronal, otras regiones cerebrales se habilitan para realizar la ‘misma’

tarea que realizaban las regiones dañadas. Pero no cualquier tejido cerebral puede suplantar las funciones de las regiones dañadas; por lo regular, son mecanismos neuronales sanos que antes ya contribuían para el desempeño de la función los que, finalmente, terminan realizando ahora por completo la (*supuestamente*) misma función (*ibid*). Para esta descripción de la plasticidad cerebral, antes que afirmar llanamente un caso de RM, quizás deberíamos plantear dudas del tipo que propone Shapiro: ¿se trata realmente de realizadores distintos? Esto es, ¿las regiones cerebrales que terminan haciéndose cargo de las funciones que realizaba la región dañada tienen propiedades causales diferentes?

Otro aspecto que también evidencia Bickle en torno a la plasticidad cerebral tiene que ver con las graves deficiencias funcionales que sobrevienen al daño. Las personas que lo sufren pueden seguir hablando, teniendo representaciones espaciales y moviendo sus extremidades, pero la forma en la que desempeñan estas tareas se ve seriamente deteriorada, cuantitativa y cualitativamente (*ibid*). En este escenario, las dudas de Couch vuelven a tener sentido. Aunque no son los mismos mecanismos neuronales los que realizan la función después del daño cerebral, son estructuralmente *similares*. Pero, ¿la *similitud* de los realizadores prueba que se trata de la *misma* función? Por otro lado, dado el deterioro funcional que sobreviene al daño cerebral, ¿deberíamos seguir hablando de que son las *mismas* funciones psicológicas las que realizan las estructuras neuronales alternativas?⁴³

Nótese como TRM, otra vez, se ve despojada de su potencia ante algunas pocas exigencias de detalle. Una descripción sólo un poco más detallada de las que se ofrecen sobre la plasticidad cerebral, aunada a la batería de cuestiones que Shapiro y Couch plantean acerca de la diversidad de los realizadores físicos o la identidad de las funciones realizadas, son suficientes para poner en suspenso la verdad de TRM respecto de la mente y sus estados y capacidades en general.

De nuevo, llegamos a un punto en el que se reconoce como indispensable dar cuenta de cuestionamientos de este tipo siempre que se presuma un caso de realización múltiple. Y toda vez

⁴³ Bechtel y Mundale (1999, p. 184) señalan que algunos casos de daño cerebral apuntan a límites infranqueables para la plasticidad cerebral. Existen regiones cuyo daño implica la pérdida definitiva de ciertas funciones psicológicas. Este es el caso de la llamada área de Broca: lesiones en esta zona conllevan la pérdida de la capacidad de hablar articuladamente. Este hecho, unánimemente reconocido en neurofisiología, implica el rechazo de RM respecto de la capacidad del lenguaje en humanos.

que se pretenda avanzar en este sentido, de nuevo, parece que será necesario tomar para su análisis la evidencia *empírica* que se tenga a la mano. Las preguntas no carecen de justificación, y no parece que sea posible responderlas apelando a distinciones psicológicas ordinarias y consideraciones *a priori*.

Lo que ha logrado la neurofisiología recientemente, sin duda, apunta hacia algunas clases de correlaciones entre tipos de estados psicológicos y cerebrales, no sólo interesantes, sino insoslayables. La vaga alusión a ellas no pretende nada más que reconocer que aún es necesario un arduo trabajo de análisis filosófico sobre los datos que aportan evidencias de correlaciones *aparentes* entre estados psicológicos y cerebrales; que ninguna interpretación general y poco detallada de tales correlaciones puede probar nada a favor o en contra de la posibilidad masiva o parcial de realización múltiple de lo mental. Si a esto sumamos las dudas que avanzan Shaprio y Couch, parece que comenzamos a vislumbrar un escenario en el que la Realización Múltiple no resulta ser una propiedad *evidentemente distintiva* de lo mental. No es la poderosa, incuestionable y *a priorísticamente* válida premisa que muchos filósofos anti-reduccionistas o defensores de la autonomía de la psicología, consideran que es.

5. CODA: UNA HIPÓTESIS SOBRE LA APARENTE PLAUSIBILIDAD DE TRM

¿Por qué TRM resulta tan atractiva para los filósofos? Bechtel y Mundale (1999) sugieren que la plausibilidad de TRM se funda principalmente en establecer distintos criterios de individuación para los realizadores físicos, por un lado, y las funciones psicológicas, por otro. Es decir, mientras que los criterios para especificar la identidad de las *funciones psicológicas* son *gruesos* o *demasiado amplios* (coarse-grained), los criterios para especificar la diversidad de realizadores físicos son bastante *finos* (fine-grained). Cuando se trata de RM, las funciones o estados psicológicos regularmente son definidos en términos bastante *amplios*, de suerte que un gran número de instancias (reales o posibles) de estados psicológicos —implementadas en organismos de diferentes especies, en alienígenas y en computadoras— cuadrarán con tal definición. Registros de especificidad mucho más altos se observan en la individuación de realizadores físicos. Cuando se trata de este expediente, los defensores de TRM se inclinan a interpretar la más mínima diferencia estructural como evidencia de multiplicidad de realizadores (*op. cit.* p. 202). Bechtel y Mundale señalan que los registros de *especificidad* para individuar pueden aumentarse —en el caso de los estados psicológicos— o disminuirse —en el caso de los estados neuronales—; lo importante es

reconocer la posibilidad de que, en un escenario en el que las exigencias de *especificidad* en los criterios de individuación son, aproximadamente, las mismas, las identidades entre tipos de estados psicológicos y tipos de estados neuronales pueden ser más plausibles.

Según Bechtel y Mundale, tales identidades se descubren poco a poco en el campo de la neurofisiología. Los autores sostienen que la noción de ‘estado cerebral’ que utilizan los filósofos es una *ficción* que nada tiene que ver con las prácticas de los neurofisiólogos (*op. cit.* p.177). Los neurofisiólogos utilizan una noción más bien *amplia* para individuar los estados neuronales; tan *amplia* como la que utilizan los filósofos para individuar los estados psicológicos. Esta paridad en la *amplitud* de los criterios de individuación ha hecho posible, en primer lugar, encontrar que diferentes especies con ancestros comunes, por *homología* —sin importar lo distintos que luzcan sus cerebros— comparten las mismas estructuras cerebrales. En segundo lugar, y más importante, que las mismas estructuras cerebrales, en diferentes especies, realizan las mismas funciones.⁴⁴

Por todo lo anterior, según observan Bechtel y Mundale, los defensores de TRM incurren en una forma de negligencia argumentativa al no percatarse de que sus demandas de *especificidad* para individuar estados psicológicos y estados neuronales son asimétricas. Quizás no exista una explicación satisfactoria para tal disparidad, y tampoco debe existir un argumento que aspire a ser válido, y que descansa en una disparidad tal.

Si las reflexiones precedentes son correctas, y justamente cuestionan la especificidad del fenómeno de la RM y la TRM, entonces, por extensión, deberíamos conferirle menor peso del que usualmente le conferimos a los argumentos en los que figuran dicha noción y tesis, y a lo que pretenden probar. Esto incluye, obviamente, los argumentos del tipo que avanza Hansberg, para los que resulta crucial la apelación a la RM y la TRM. Podemos seguir pensando, por lo tanto, en la posibilidad de abrir más aún esa brecha que nos permita repatriar a los estados mentales y, por lo tanto, a las emociones, al espacio de las clases naturales. El siguiente paso consiste en construir una semántica de TCN y una metafísica de las clases naturales apta y aceptable, más dócil y liberal, para recibirlos de nuevo.

⁴⁴ La caracterización de los estados cerebrales por homología —la cual podría entenderse como una construcción *amplia* de los criterios de individuación de los estados cerebrales—, es lo que permite que la neurofisiología pueda llegar a conocer mucho del cerebro y sus funciones psicológicas, al estudiar el cerebro de otras especies de mamíferos

V. CLASES NATURALES Y ‘ESTRUCTURA EXPLICATIVA’: DOS VERSIONES

V.I. LA CONSTRUCCIÓN DE UNA NOCIÓN FUERTE DE ‘ESTRUCTURA EXPLICATIVA’

La forma en la que se conciba la noción de ‘esencia’ o ‘estructura explicativa’ es crucial, como se hizo notar en su momento (I, III), para que procedan argumentos como los que Hansberg (1990) avanza en contra de la tesis de que los estados mentales son clases naturales. La forma en la que dicha noción se entiende en tales argumentos está anidada en lo que hemos denominado *esencialismo metafísico de clases naturales* (I, II.IV, 4)⁴⁵. En adelante, para agilizar la exposición nos referimos a esta postura, simple y llanamente, como ‘esencialismo’. Desde el punto de vista metafísico, el esencialismo sostiene que en el mundo existen *clases* de cosas, sustancias y organismos con propiedades esenciales sin las cuales no podrían existir dichas clases. La idea es que la realidad tiene una estructura fabril: produce en serie diferentes *clases* de entidades, cada una con propiedades distintivas o esenciales: clases naturales.⁴⁶ Esta caracterización de la estructura metafísica de la realidad, describe un *hecho* acerca de o un *rasgo* de la naturaleza que es independiente de hechos psicológicos o sociales relativos a los seres humanos. Las propiedades esenciales o esencias que identifican a cada clase natural son *físicas* y, por lo tanto, se descubren *a posteriori*. Además, es sobre estas *propiedades esenciales* (físicas) que se construyen leyes y teorías *explicativas* sobre las clases naturales; por esta razón Platts y Hansberg llaman a las propiedades esenciales o esencias, ‘estructuras explicativas’. En una versión moderada de este esencialismo (Ellis, 2001), la estructura explicativa es crucial para determinar la extensión de las condiciones necesarias y suficientes de pertenencia para este mundo. En una versión más fuerte, las propiedades esenciales dictan condiciones del mismo tipo pero *para todos los mundos posibles* (Kripke, 1981).

Cumple anotar que la *metafísica* del esencialismo, por lo menos tal y como la hemos descrito, no incluye o sugiere más *especificaciones* en torno a la naturaleza de las esencias o estructuras explicativas, que la que afirma que éstas refieren a propiedades físicas relevantes. Quedan sin responder las siguientes preguntas: ¿las propiedades físicas relevantes —que determinan las estructuras explicativas— son disposicionales, estructurales o ambas... o de algún otro tipo?, ¿los

⁴⁵ Boyd (1991) llama *doctrina esencialista de las clases naturales* a lo que nosotros llamamos aquí *esencialismo metafísico de clases naturales* y a la *semántica esencialista* (SE) que hemos asociado con dicha postura.

⁴⁶ En una fórmula esencialista radical como la de Ellis (2001), las líneas de producción serían procesos naturales que, en sí mismos, son también *clases naturales*.

conjuntos de propiedades que conforman dichas esencias son *estáticos* y *exhaustivos* o *dinámicos* e *indeterminados*: es decir, no cambian a través del tiempo y constituyen conjuntos bien delimitados de propiedades; o pueden evolucionar en el tiempo y consisten en conjuntos de propiedades más bien vagos e indeterminados?⁴⁷ Esta cuestión puede formularse figurativamente del siguiente modo: ¿la realidad corta fino y de una vez por siempre sus coyunturas, o más bien se permite trazos brumosos y susceptibles a la variación?

Respecto de la primera pregunta, para empezar, puede decirse que goza del favor de haber sido planteada y respondida por diferentes autores relevantes para la doctrina de las clases naturales. Quine (1969), por ejemplo, sugiere que las esencias interesantes desde el punto de vista científico han de ser de carácter estructural. Hacking (1991/2007) sugiere identificar las esencias en términos de propiedades disposicionales y Boyd (1979/1993) propone caracterizaciones en términos de propiedades relacionales complejas. Aunque la metafísica esencialista, en sí misma, no se compromete con ninguna postura en particular, nos inclinamos a pensar que la respuesta de Boyd —a la que denominaremos *realismo de clases naturales*— es la más plausible por razones que, esperamos, se harán explícitas más adelante.

⁴⁷ Por supuesto que, dados los adjetivos ‘estático’, ‘dinámico’, ‘exhaustivo’ e ‘indeterminado’, pueden concebirse otras posibilidades metafísicas respecto de la naturaleza de las clases naturales que produce la realidad: que el mundo sea *estático* e *indeterminado*; que el mundo sea, respecto de algunas clases naturales *estático*, y respecto del resto *dinámico*, pero siempre indeterminado, etc. Pero hay posibilidades que parecen metafísicamente poco razonables: aquellas en que la realidad se concibe, en su totalidad o parcialmente (para algunas clases naturales), como *dinámica* y *exhaustiva* a la vez. Estas posibilidades implicarían que todas o algunas clases naturales pueden cambiar sus esencias a través del tiempo. Pero, si esto sucede —dentro de esta posibilidad metafísica— dicho cambio se debería entender como un salto *esencial* de un conjunto bien delimitado de propiedades a un conjunto distinto de propiedades bien delimitado. Ninguna conjunto potencial de objetos en el mundo que podamos llamar clase natural, nos proporciona alguna motivación para defender estas posibilidades metafísicas. Por otro lado, más allá de que el mundo no ofrezca motivaciones para adscribirle la metafísica en cuestión, una *interpretación filosóficamente intuitiva* —donde lo intuitivo se reconoce como modelado por las ideas de Putnam y Kripke—, de este tipo de saltos en las esencias o estructuras explicativas, nos complicaría la posibilidad de hablar de clases naturales *dinámicas*. Para una *interpretación filosóficamente intuitiva*, los saltos en las esencias (de un conjunto delimitado de propiedades esenciales a otro conjunto distinto), más que atribuibles a una naturaleza *dinámica* de las clases naturales, representarían simplemente la extinción de *una* clase natural y el nacimiento de *otra* distinta. La *interpretación filosóficamente intuitiva* parece asumir un compromiso tácito y profundo con el carácter *exhaustivo* de las esencias de clases naturales, lo que hace difícil conceder que éstas sean de naturaleza *dinámica*. En esta dirección, parece que es el mismo compromiso con la *exhaustividad* de las esencias de las clases naturales, el que está detrás de la tesis que niega que las especies sean clases naturales: son *dinámicas* (cambian/evolucionan con el tiempo); sus propiedades relevantes no configuran nunca conjuntos bien delimitados. Veremos que el supuesto de que las esencias de las clases naturales son *exhaustivas*, es responsable no sólo de impedir que se considere que las clases naturales son *dinámicas*; sino que, al mismo tiempo, es responsable de construcciones excesivamente *fuertes* de la noción de esencia o estructura explicativa.

I. ALGO MÁS QUE ‘SEMÁNTICA ESENCIALISTA’ (SE).

Por otro lado, la articulación de una respuesta para la segunda cuestión —¿los conjuntos de propiedades que conforman las esencias o estructuras explicativas son *estáticos* y *exhaustivos* o *dinámicos* e *indeterminados*?— puede colegirse mediante un escrutinio de la semántica (SE) de la que deriva *esencialismo metafísico de las clases naturales*. La *Semántica Esencialista* (SE) expuesta en I, II.III, 3, no sólo está hecha a la medida de la metafísica esencialista, también sienta las bases para una *construcción fuerte* de la noción de ‘esencia’ o ‘estructura explicativa’. Esto tiene como consecuencia que los conjuntos de propiedades que constituyen la esencia o estructura explicativa de las clases naturales, se consideren *estáticos* y *exhaustivos*. Esta *construcción fuerte* de la noción de esencia está detrás de muchos argumentos que conducen al esencialismo a conclusiones desalentadoras: como ya se ha mencionado varias veces está detrás de los argumentos de Hansberg (1990) que niegan que los estados mentales sean clases naturales, y está, también, detrás de los argumentos que niegan que las especies sean clases naturales (Boyd, 1999).

Pero, ¿cómo es que SE ayuda a responder la cuestión sobre la naturaleza de los conjuntos de propiedades que constituyen las esencias de las clases naturales, toda vez que construye una noción fuerte *esencia* o *estructura explicativa*? Según SE, los términos de clase natural adquieren su significado y extensión por medio de un acto de bautismo u ostensión, en el que se le asigna un nombre ‘N’ a una supuesta clase natural utilizando una muestra paradigmática de dicha clase. El procedimiento implica un acto *deíctico*, por medio del cual se estipula que, toda muestra que se haya de considerar parte de la extensión de la clase natural y de la cual se pueda decir que es ‘N’, debe tener, en todos los mundos posibles en los que exista, la *misma* composición interna o esencia *E* que tiene la muestra paradigmática a la que se señala durante el bautismo. Así, el significado y extensión de un término de clase natural queda determinado por la composición interna o esencia *E* de la muestra paradigmática de la clase para todos los mundos posibles: ‘N’ es el nombre de todo aquello, en todos los mundos posibles, que tenga como composición interna o esencia *E*. Otro aspecto relevante de SE respecto de los términos de clase natural, es que sostiene que *E* se descubre y determina *a posteriori*, mediante indagación empírica. En este punto, SE no parece ofrecer ninguna pista que nos faculte para articular una respuesta a la cuestión sobre la naturaleza de los conjuntos de propiedades que constituyen las esencias o estructuras explicativas: ¿son *estáticos* y *exhaustivos* o *dinámicos* e *indeterminados*? Son algunas implicaciones ampliamente

reconocidas las que ayudan a entender cómo se construye una noción metafísica fuerte de esencia a partir de *SE*. En el capítulo III, con ayuda del un artículo de Hansberg (1990), se identificaron algunos rasgos distintivos de *SE* que ahora es oportuno emplazar.

En el contexto de *SE*, se tiende a interpretar la esencia de las clases naturales como un conjunto de propiedades físicas relevantes (esenciales), que proporcionan *condiciones necesarias y suficientes* para determinar la membrecía, así como el significado y extensión de los términos de clase natural. La tesis no es sólo semántica, sino metafísica. La idea de *condiciones necesarias y suficientes* supone que las clases naturales tienen como esencias conjuntos de propiedades esenciales *bien delimitados*: sólo si las esencias tienen límites claros, es posible pensar que proporcionen condiciones tan específicas (*necesarias y suficientes*) para determinar cuestiones semánticas como la del significado, extensión y aplicación correcta de términos de clase natural, así como cuestiones ontológicas como las que se plantean al decidir la membrecía de muestras respecto de una clase natural. De alguna forma, la semántica trae consigo una metafísica que sugiere que la naturaleza es isomórfica con nuestras exigencias *definicionales en términos de condiciones necesarias y suficientes*; que la naturaleza corta fino en sus coyunturas, tal como lo hacen nuestras mejores definiciones. Finalmente, la semántica esencialista termina siendo un poco más que semántica, e incluye una metafísica según la cual, las esencias de las clases naturales son *exhaustivas* —están constituidas por conjuntos bien delimitados de propiedades—, y gracias a ello pueden proporcionar condiciones necesarias y suficientes para determinar el significado y la extensión de los términos de clase natural.

Otro rasgo de *SE* crucial para terminar de construir una noción fuerte de esencia o estructura explicativa, consiste en entender los términos de clase natural como *designadores rígidos* en el sentido de Kripke. Los términos de clase natural, en tanto que *designadores rígidos* seleccionan un conjunto de propiedades físicas esenciales (esencia) en todos los mundos posibles: a saber, el conjunto de propiedades que constituyen la esencia de la muestra paradigmática que se utiliza en el acto de definición ostensiva de una clase natural. La selección de *ese* conjunto concreto de propiedades físicas relevantes (de la muestra paradigmática) es crucial para *SE*, ya que sirve para decidir, *en todos los mundos posibles*, la extensión y la correcta aplicación del nombre con el que se bautizó la clase natural. Lo que hace el término de clase natural en tanto que designador rígido, es petrificar la esencia o estructura explicativa de una clase natural, impedir la posi-

bilidad de variación a través del tiempo y el espacio: fáctico y contrafáctico. Finalmente, la propiedad de la *rigidez* en *SE*, concibe los conjuntos de propiedades físicas relevantes que constituyen las esencias o estructuras explicativas de las clases naturales como *ahistóricos* y *estáticos*, idénticos en todos los mundos posibles en los que existen dichas clases naturales. De nuevo, la semántica es más que eso y se torna metafísica: una metafísica según la cual, la naturaleza corta de una vez y para siempre —y en todos los mundos posibles también— sus coyunturas ontológicas.

2. IMPACTO DE LA NOCIÓN FUERTE DE ‘ESTRUCTURA EXPLICATIVA’: ESPECIES Y EMOCIONES.

Puede verse, entonces, de qué manera se construye una noción fuerte de esencia o estructura explicativa en el marco de la *semántica esencialista* (*SE*) que da lugar al *esencialismo metafísico de clases naturales*. Una noción que responde a la cuestión sobre la naturaleza de los conjuntos de propiedades que conforman las esencias o estructuras explicativas al concebirlos como *estáticos* y *exhaustivos*. Resulta notable percatarse de que es propiamente la *semántica esencialista*, y no la *metafísica esencialista*, la que aporta las bases metafísicas para la construcción fuerte de la noción de ‘esencia’ o ‘estructura explicativa’.⁴⁸ Como quiera que sea y cualesquiera que sean el origen y las motivaciones, lo que es ahora importante es reconocer que esta construcción fuerte de la noción de esencia o estructura explicativa subyace a los argumentos de Hansberg (I, III) y, en general, a una tradición crítica en contra de la noción de clase natural. En este sentido, el análisis de los tres argumentos de Hansberg ayudó a identificar la noción fuerte de esencia o estructura explicativa que en este apartado se ha inferido de *SE*. Y no sólo eso, también se estableció que era una premisa fundamental en la articulación argumentativa de Hansberg en contra de la tesis de que los estados mentales —incluidas las emociones— son clases naturales.

⁴⁸ La motivación para derivar de *SE* una noción de ‘esencia’ con estas características (metafísicas) tiene que ver con el hecho de que ayuda a reconocer que algunos enunciados que postulan identidades teoréticas (v. gr. ‘el agua es H₂O’) son metafísicamente necesarios y, al mismo tiempo, cognoscibles *a posteriori*. Estas tesis, en sí mismas, resultan ser antídotos muy eficaces contra el relativismo kuhniano y su noción de inconmensurabilidad. Como se verá más adelante, puede construirse una semántica y una metafísica a partir de la cual se configure una noción de ‘esencia’ o ‘estructura explicativa’ más tolerante y flexible; y que, del mismo modo, funcione en contra de la amenaza de la inconmensurabilidad kuhniana. En cuanto a *SE*, nos inclinamos a pensar, siguiendo a Boyd (1999; 1991; 1979/93), que se funda en una veneración por la precisión en las definiciones heredada de una tradición empirista que se remonta a Locke. Esta obsesión con la ‘precisión’ impone paradigmas de exactitud en las definiciones que se filtran en *SE* y se transforman en compromisos metafísicos que le atribuyen a la realidad la cualidad de ‘cortar fino, y de una vez y para siempre, sus coyunturas ontológicas’.

Tenemos, entonces, buenas razones para sostener que ha existido cierta tendencia —presente en los argumentos de Hansberg— hacia una construcción *fuerte* de la noción de ‘esencia’ o ‘estructura explicativa’ dentro del *esencialismo metafísico de clases naturales*: una noción que entiende las esencias de las clases naturales como conjuntos de propiedades *estáticos* y *exhaustivos*. En esta dirección, identificamos algunas tesis (metafísicas) dentro de la *Semántica Esencialista*, como las responsables detrás de dicha construcción.

Lo que ha derivado de esta combinación de eventos son actitudes restrictivas en torno a la naturaleza de las clases naturales, con algunas consecuencias filosóficas relevantes. Boyd (1999) señala que posturas como las de Ghiselin (1974) y Hull (1978), que niegan a las especies el estatus de clases naturales y las consideran ‘individuos’, descansan en algunos atavismos empiristas en torno a la definición. La revisión de tales ‘atavismos’, evidencia que detrás de ellos están las mismas tesis que, en *SE*, favorecen una construcción fuerte de la noción de ‘esencia’ o ‘estructura explicativa’. Según los autores mentados, las especies no son clases naturales porque:

1. No son definidas en términos de condiciones necesarias y suficientes (propiedades *intrínsecas*) como lo son las clases naturales respetables.
2. Están limitadas a circunstancias y periodos históricos particulares, mientras que las clases naturales no tienen estas restricciones y se consideran universales.
3. No caen bajo leyes universales sin excepciones como lo hacen las clases naturales auténticas.
4. Difieren de las clases naturales en que los criterios que unen a sus miembros son relaciones históricas y no un conjunto de propiedades compartidas.

(1) y (4) exhiben la imagen de las esencias como conjuntos de propiedades bien delimitados: *exhaustivos*. (2) y (3) muestran un compromiso con las esencias entendidas como invariables en el tiempo y sujetas eternamente a las mismas leyes: *estáticas*.

Otro tanto acontece en el campo de los estados mentales y, en particular, en torno a las emociones. Los argumentos de Hansberg sobre los estados mentales en general, lo mismo que los de otros autores que niegan en particular a las emociones el rango de clases naturales, descansan de un modo fundamental en una construcción fuerte de la noción de ‘estructura explicativa’. En el caso concreto de las emociones, Prinz (2004, cap. 4) denomina ‘tesis de la desunión’, a una

actitud común entre los teóricos de las emociones que consiste en negar que el complejo grupo de fenómenos que llamamos ‘emociones’ tenga un conjunto unificado de propiedades comunes. Es habitual, de los autores que suscriben de alguna u otra manera esta tesis, negar que las emociones sean clases naturales.⁴⁹ Pero aquí, de nuevo, está presente una noción fuerte de ‘estructura explicativa’ y de clase natural, que se ve defraudada ante el hecho de que, aquello que llamamos emociones, no satisface la condición de tener en común un conjunto *bien delimitado* y unificado de propiedades esenciales. Por lo menos, una imagen de las esencias como *estáticas* y *exhaustivas* subrepticamente motiva estas posiciones.

V.II. LA CONSTRUCCIÓN DE UNA NOCIÓN MÁS LIBERAL DE ‘ESTRUCTURA EXPLICATIVA’: LA PROPUESTA DE BOYD.

Las ideas de Richard Boyd en torno a la naturaleza de las clases naturales, han sido sugeridas por varios autores como antídotos frente a los deletéreos efectos de una construcción fuerte de la noción de ‘estructura explicativa’, en el marco de la discusión del estatus de las emociones en tanto que clases naturales. Prinz (2004), Charland (2002) y Griffiths (1997) han abordado dicha problemática a la luz de las ideas de Boyd, con diferentes resultados.⁵⁰ Griffiths resuelve que algunas emociones básicas son clases naturales que, en conjunto, conforman *una* clase natural. Llama programas afectivos a estas emociones y las define como respuestas (fisiológico/psicológicas) adaptativas, complejas, coordinadas y automáticas, diseñadas para enfrentar situaciones recurrentes con relevancia evolutiva. Otro grupo de estados afectivos, según Griffiths, lo constituyen las emociones cognitivas superiores (culpa, celos, amor, envidia, vergüenza, etc.), las cuales tienen una historia evolutiva distinta y utilizan mecanismos causales distintos. Debido a estas diferencias, Griffiths piensa que todo aquello a lo que llamamos emoción no constituye una clase natural. Más aún, sugiere moverse hacia alguna forma de *eliminativismo* de las emociones como cate-

⁴⁹ La ‘tesis de la desunión’ tiene al menos tres advocaciones. Desde el punto de vista cognitivista, significa que no existe un conjunto específico de actitudes proposicionales o de objetos formales comunes a todas las emociones; desde el punto de vista neurofisiológico, significa que no existe una estructura o circuito neuronal especializado en la producción de los episodios emocionales; desde una perspectiva mixta, significa que no hay rasgos cognitivos y fisiológicos comunes a todas las emociones. Respecto de la primera formulación, se pueden señalar como ejemplos las posturas de De Sousa (1987), Hansberg (1996) y Rorty (1980). Feldman Barrett (2006) ejemplifica la tesis de la diversidad neurofisiológica. Elster (1999) sirve como ejemplo para la tercera advocación de la tesis de la desunión.

⁵⁰ Incluso Feldman Barrett (2006), menciona las ideas de Boyd sobre las clases naturales. Pero no parece tomarlas muy en serio en su discusión en torno la naturaleza de las emociones. La cantidad de formas en las que son distintas las cosas a las que nos referimos con términos emocionales, según esta autora, hace muy complicado poder considerarlas, en conjunto, una clase natural. La ‘tesis de la desunión’ y, en esta dirección, un compromiso implícito con una construcción fuerte de la noción de ‘estructura explicativa’, están detrás de su negativa a considerar a las emociones clases naturales. No están bien representadas las ideas de Boyd en el trabajo de Feldman Barrett (2006).

goría psicológica general: la psicología *general* de las emociones se desplazará hacia una ciencia afectiva que estudiará diferentes tipos de fenómenos (afectivos). Charland y Prinz son más optimistas en relación a la cuestión de las emociones en tanto que clase natural. Ambos consideran que las posturas en contra de la idea de que las emociones son clases naturales no están del todo justificadas y ofrecen distintas propuestas para construir una teoría general de las emociones que más adelante tendremos oportunidad de revisar.

Corresponde ahora, exponer las ideas de Boyd para obtener por nuestra cuenta una resolución acerca de la manera en la que pueden proporcionar o no, una alternativa más tolerante y fructífera —más que la que aporta una construcción fuerte de la noción de ‘esencia’ o ‘estructura explicativa’ en el marco de la SE y el *esencialismo metafísico de clases naturales*— para abordar la problemática de las emociones en tanto que clases naturales.

1. HACIA UNA SEMÁNTICA REALISTA DE LOS TÉRMINOS DE CLASE NATURAL.

Ninguno de los autores arriba citados, menciona la semántica de Boyd al referirse a sus aportaciones en torno a la doctrina de las clases naturales. En la medida en que este trabajo pretende proporcionar un marco alternativo sólido para re-examinar el tema del estatus de las emociones en tanto que clase naturales, creemos indispensable exponer detalladamente las tesis semánticas de Boyd. Para tales efectos, comenzaremos por exponer la semántica de términos de clase natural que elabora en su artículo publicado originalmente en 1979, y revisado y ampliado en 1993: ‘Metaphor and theory change’.

Las ideas de Boyd en torno a la semántica de los términos de clase natural, son las ideas de un filósofo de la ciencia —un *realista científico* profundamente influido por Putnam, de hecho— que reflexiona en torno a la naturaleza y características distintivas de la misma.⁵¹ Para Boyd, las *teorías, métodos y prácticas* científicas tienden a lo que denomina *acomodamiento* (*accommodation*). El *acomodamiento* tiene lugar entre los tres aspectos científicos mencionados y las estructuras causales del mundo. El uso de términos de clases naturales —que figuran en las teorías— es uno de los vehículos que utiliza la ciencia para alcanzar su objetivo de *acomodar* el lenguaje (teorías, leyes, generalizaciones, inducciones, explicaciones) con las estructuras causales de la realidad. Boyd considera que las teorías causales de la referencia de Kripke y, principalmente la de Putnam, ejemplifican cómo tiene lugar el proceso de acomodamiento a partir de defini-

⁵¹ La influencia de Putnam en Boyd, tanto en su semántica de términos de clase natural como en la metafísica que concibe se hará evidente durante el desarrollo de este apartado.

ciones ostensivas (fijación de referencia) de términos de clase natural, que se fundan en *estructuras explicativas*: la fijación de referencia por medio del bautismo u ostensión determina la extensión de una clase por medio de la estructura interna que designa. La referencia de un término de clase natural, según las teorías causales de la referencia, se entiende como una *relación* (de acomodamiento) entre el lenguaje teórico-científico (término de clase natural) y la realidad (esencia).

Pero Boyd sólo le reconoce estos méritos a las teorías causales de la referencia. Boyd respeta el viraje *realista* y *esencialista* que representaron dichas teorías y la forma en que hicieron frente a la amenaza kuhniana del relativismo, pero desestima las inclinaciones metafísicas que principalmente exhibe la postura kripkeana. En este sentido, Boyd no considera que el procedimiento de fijación de referencia (extensión/significado) que tiene lugar mediante la ostensión y bautismo nos deba comprometer con una imagen de la definición en términos de *condiciones necesarias y suficientes*. Esta tesis, piensa Boyd, proviene más bien de una idealización —filosófica— de lo que logra o debería lograr la ciencia al estudiar diferentes clases de fenómenos, objetos, sustancias y organismos. En los ejemplos clásicos utilizados por Kripke y Putnam (Agua, Oro, Tigre), las esencias reales de clases naturales descubiertas *a posteriori*, son sometidas a un proceso de *idealización* en el que una clase es *definida adecuadamente* por un conjunto de condiciones necesarias y suficientes.⁵² El error de esta imagen es que no se percata de que los *conceptos* científicos están más comprometidos con requerimientos externos de acomodamiento con las estructuras causales relevantes de la realidad, que con exigencias y escrúpulos semántico-filosóficos. Luego, la historia que propone Boyd, a partir del acto de fijación de referencia que proponen Kripke y Putnam como el rasgo originario y definitivo del ‘referir’, es un tanto diferente.

2. REFERENCIA: APERTURA INDUCTIVA Y ACCESO EPISTÉMICO.

La definición ostensiva o bautismal (fijación de referencia) da lugar a lo que Boyd llama *apertura inductiva*: cuando se acuña un término de clase natural, se abre un programa de investigación (informal / formal) mediante el cual se pretenderá descubrir, *a posteriori*, la esencia o estructura explicativa de la clase que se ha definido ostensivamente.⁵³ Recuérdese que esta forma de *fijar la*

⁵² Es necesario apuntar que el proceso de idealización no fue sancionado, propiamente, por Putnam.

⁵³ Nótese cómo, en este particular, las ideas de Boyd son consistentes con la metafísica esencialista, según la cual las esencias o estructuras explicativas de las clases naturales, están constituidas por propiedades físicas importantes que han de ser descubiertas *a posteriori*.

referencia ostensivamente debe entenderse como un procedimiento que tiene como finalidad el *acomodamiento* entre componentes lingüísticos y estructuras causales *aún no descubiertas*.

Para Boyd, ‘referencia’ es un término epistemológico. Lo que esto significa es que el tipo de relaciones causales que son relevantes en el ‘referir’, son también fundamentales para la justificación de nuestras creencias. En otras palabras, entre más y mejor especificadas están las relaciones referenciales de un término de clase natural, mejor justificadas estarán nuestras creencias en torno al referente de dicho término. Boyd caracteriza el valor epistemológico de la referencia —por medio de términos de clase natural— del siguiente modo:

Un término *t* refiere a una clase *k*, sólo en el caso de que un conjunto de *mecanismos* haga posible que, todo lo que es predicado del término *t*, sea verdadero de la clase *k*. Los mecanismos que hacen esto posible incorporan procedimientos fiables para identificar a los miembros de *k*, y factores sociales como el reconocimiento de un grupo de expertos en *k*, de quienes se obtienen creencias verdaderas sobre *k* y, en cierto, modo, regulan el uso correcto de *t*. Cuando se obtienen relaciones de este tipo, puede decirse que son las propiedades de *k* las que regulan el uso de *t*; y se puede pensar que el uso del término *t* nos confiere *acceso epistémico* a *k* por medio de ciertos mecanismos de coordinación social. (Boyd, 1988/2007, pp. 170-171)

Los *mecanismos* a los que se alude en este pasaje, incluyen teorías, métodos y prácticas científicas, así como instrumentos de medición y detección de los miembros de *k* y/o sus propiedades. También se reconoce, como un elemento de estos mecanismos, lo que Boyd denomina *división del trabajo cognitivo*. Junto con la noción putnamiana de *división del trabajo lingüístico*, reconoce la existencia de diferentes grupos de expertos en diferentes niveles. En lo que difiere de la noción de Putnam, es en que el trabajo que se reparte, antes que lingüístico o incluso epistemológico, es propiamente cognitivo: se dividen las tareas sociales en un sentido más general, de suerte que cada una exige diferentes habilidades y esfuerzos cognitivos, para acceder a distintos aspectos de la realidad (natural y/o social).⁵⁴ Todos estos *mecanismos* ayudan a enriquecer nuestros

⁵⁴ Por esta razón Boyd incluye a los mecánicos y a los floricultores como expertos en sus ramas (Boyd, 1979/1993, p. 510).

conocimientos de las relaciones ‘referenciales’ de un término de clase natural con, precisamente, su referente —al mismo tiempo, se garantiza el *acomodamiento* entre las teorías científicas que incluyen dichos términos y la estructura causal de la realidad. Todas estas complejas relaciones dan lugar a lo que Boyd denomina *acceso epistémico*. El *acceso epistémico* es una propiedad de la ‘referencia’... un fenómeno que hace posible la existencia de los términos de clase natural —entre otros. Consiste en la capacidad que tiene la ‘referencia’ por medio de términos de clase natural, para proveernos de conocimiento sobre sus referentes, el mundo y su estructura causal —acomodamiento.

Existen dos formas de *acceso epistémico*. Por un lado, el que tiene un usuario del término “ADN” respecto de su referente (un compuesto molecular que porta el código genético en los mamíferos). Este tipo de *acceso epistémico* descansa en nuestra confianza hacia el testimonio de un grupo de expertos. Pero esta forma de *acceso epistémico* no es la más relevante. El que es importante es el *acceso epistémico* que el término “ADN” proporciona respecto del *ADN*, en virtud del rol que el término desempeña en la organización de un programa de investigación coordinado socialmente: *apertura inductiva*.

A modo de ejemplo, Boyd menciona tres modos en los que el uso del término “ADN” en una comunidad científica contribuye al descubrimiento de propiedades o información acerca del *ADN*.

1. Su uso permite el intercambio y reportes de resultados acerca del *ADN*... entre los científicos.
2. Su uso permite *articulación teórica* (debate, crítica, justificación y refinamiento) en torno al *ADN*. De este modo, la interpretación de la evidencia y la evaluación de las teorías acerca del *ADN* se constituyen como una empresa social.
3. Su uso, no sólo hace posible la conformación de una sociedad de cooperación experimental, observacional e interteórica sobre el *ADN*, también hace posible el razonamiento verbal sobre el *ADN*. (Boyd, 1979/1993, p. 506).

3. SOBRE LA NATURALEZA DE LAS DEFINICIONES DE LOS TÉRMINOS DE CLASE NATURAL.

Desde el punto de vista de Boyd, la introducción de un término de clase natural, por medio de una definición ostensiva, inaugura un *programa de investigación* —*apertura inductiva*— que

tiene como objetivo descubrir *empíricamente* la ‘estructura explicativa’ del referente del término acuñado. Se dice que la ‘referencia’, mediante términos de clase natural, es una noción epistemológica porque la relación de referencia se nutre y robustece mediante el uso de distintos *mecanismos* característicos de la indagación empírica, que ayudan a develar la esencia o estructura explicativa del referente y, al mismo tiempo, justifican nuestras creencias sobre el mismo. Este enriquecimiento de las relaciones de referencia entre un término de clase natural y su referente, favorecen el *acomodamiento* y el *acceso epistémico*: todo esto tiene lugar y es posible por medio de los términos de clase natural.

GRADUALISMO DEFINICIONAL. Pero el *acomodamiento* y *acceso epistémico* que es posible mediante el uso (referencial) de términos de clase natural, tiene lugar de manera aproximativa y gradual. Es bastante complicado que las esencias o estructuras explicativas de las clases definidas ostensivamente sean distinguidas inicialmente de forma precisa y adecuada. La *apertura inductiva* avanza de esta forma y evidencia que la fijación de referencia es un proceso gradual e impreciso (en todas sus etapas) que difícilmente aceptaría definiciones en términos de condiciones necesarias y suficientes, como cierta interpretación de SE lo sugiere. Boyd rechaza una imagen monolítica de la referencia, en la que se la entiende como una relación rígida entre una palabra y una clase perfectamente definida. Según su concepción *gradualista*, los nuevos hallazgos y descubrimientos *cambian* nuestro uso de ciertos términos genéricos, a la vez que *enriquecen* nuestro conocimiento de sus referentes. Esto es parte del fenómeno de la *referencia*. En la imagen monolítica, los *cambios* en nuestro uso de ciertos términos, colapsan la relación referencial “*término de clase natural – referente*”; se sustituye una relación arcaica por una nueva.

Lo que justifica la concepción *gradualista* de la referencia, en contra de una concepción *monolítica* —como la que se propone en SE mediante los designadores rígidos, a partir de supuestas esencias estáticas en todos los mundos posibles—, tiene que ver con la naturaleza de la empresa socio-cognitiva detrás de la fijación de referencia que describe Boyd. La finalidad de *acomodamiento* entre el lenguaje y la estructura causal de la realidad, da lugar a la *apertura inductiva* en el momento en que se acuña un término de clase natural ostensivamente. La *apertura inductiva* avanza *gradualmente* por medio del refinamiento referencial—*acomodamiento*—, el cual es posible mediante diversos *mecanismos* metodológicos, sociales y cognitivos. Todo esto hace posible el *acceso epistemológico*, el cual no solo nos abre paso hacia los referentes de los términos de clases naturales, sino que posibilita la creación de una sociedad de cooperación

epistémica, que tiene la tarea de *revisar*, criticar, justificar, *refinar* o *reformular*, si es necesario, las definiciones de nuestros términos de clase natural, sobre la base de la interpretación de la evidencia y/o nuevos hallazgos —de nuevo, *acomodamiento*. La *revisión*, el *refinamiento* o la *reformulación* definicional (relaciones referenciales) de los términos de clase natural, son constitutivos, entonces, de su semántica: lo cual implica que su naturaleza es *gradual*, y no *rígida* o *inamovible*.

Las ideas de Boyd acerca de la *gradualismo definicional y referencial*, constituyen una teoría sobre el *cambio conceptual*. Para Boyd, el carácter *gradual* y *aproximativo* de la referencia garantiza, más que el colapso (inconmensurabilidad: Khun), la continuidad conceptual. El gradualismo referencial nos ayuda a entender que nuestra comprensión de las ‘estructuras explicativas’ de los referentes de los términos de clase natural, puede modificarse y obligar a revisar o redefinir dichos términos, sin que ello signifique la disolución, desaparición o extinción del referente original. El referente continúa existiendo y no necesariamente cambia en su naturaleza, lo que cambia es la relación referencial que tenemos con él: en palabras de Boyd, tiene lugar un ajuste o acomodamiento entre nuestras teorías y las estructuras causales de la realidad.

Si la ‘referencia’ en general tiene el carácter epistemológico que le atribuye Boyd, entonces los términos que refieren a estados mentales podrían también detentar la dimensión *gradual* y *aproximativa* implicada: serían sensibles al refinamiento referencial que tiene lugar mediante el esfuerzo socio-cognitivo que subyace a la fijación de la referencia; *el acceso epistémico* los somete a procesos de *revisión* y/o *redefinición* característicos de términos generales/científicos y de clase natural. Esto se opone a la caracterización semántica de los términos de estados mentales que propone Hansberg en dos de los argumentos que fueron revisados anteriormente (*supra*, III). Según Hansberg, **(a)** los términos de clase natural no denotan, ni explícita, ni implícitamente, propiedades físicas relevantes (micro-estructuras) subyacentes a los estados mentales; los términos de estados mentales, más que referir, son útiles para interpretar y hacer inteligible nuestra conducta y la de los demás. No hay empresa socio-cognitiva que cambie esta situación. Relacionado con este hecho (aparentemente) *semántico*, **(b)** considera que la tesis de la *división del trabajo lingüístico* (el cuasi-equivalente de lo que Boyd llama *división del trabajo cognitivo*) no se aplica respecto de los términos de estados mentales. No existe una comunidad de expertos que nos oriente o decida si aplicamos correctamente algún término mental. “Nuestras descripciones psicológicas del sentido común definen a los estados mentales *de una manera definitiva*” (Hans-

berg 1990, p. 220). Todos somos expertos cuando se trata del uso de términos de estados mentales.

(a) y (b) son consistentes con una concepción *rígida* o *inamovible* de los términos de psicología popular de estados mentales, toda vez que niegan que la indagación empírica —el descubrimiento de propiedades físicas relevantes— o lo que determine un grupo de expertos acerca de los estados mentales, pueda modificar el significado o reglas del uso correcto de los términos de estados mentales. Esto quiere decir que el significado o las reglas sobre el uso de los términos de estados mentales, no sólo no se modificará —pase lo que pase con la psicología o la neurofisiología— sino que nunca se ha modificado. Pero una revisión superficial de los hechos lingüísticos impugna esta afirmación, junto con las tesis (a) y (b). Es innegable que nuestra imagen actual de las capacidades, disfunciones y estados mentales no es la misma que la que imperaba durante la Edad Media, por ejemplo. No sólo hemos dejado de explicar las capacidades —lenguaje, razonamiento, memoria— y disfunciones mentales —epilepsia, depresión, esquizofrenia— en términos de entidades trascendentes o fuerzas sobrenaturales, sino que, en algún sentido, en occidente estamos dispuestos a asentir a la afirmación de que tales capacidades o discapacidades dependen y se explican, de algún modo *especificable por los expertos*, por ciertas propiedades y acontecimientos físicos/cerebrales. Ya que, en efecto, nuestros conceptos sobre tales ‘aspectos’ de lo mental han cambiado en este sentido, entonces es falso sostener que dichos conceptos, junto con los términos que los refieren, son *inamovibles*. Cumple agregar, que ha sido el trabajo y hallazgos teóricos y empíricos del conjunto de investigadores (*expertos*) que han estudiado el fenómeno de lo mental en toda su complejidad —*apertura inductiva*—, el que se halla detrás de las *revisiones*, *reformulaciones* y *modificaciones* en nuestros conceptos y términos de *capacidades* y *disfunciones* mentales —*acceso epistémico*. En este sentido, los estados mentales concretos y los términos que los refieren, tampoco se sustraen a cierta *evolución semántica*, motivados por el avance de la neurofisiología y los hallazgos empíricos. Sabemos que un esquizoide puede tener *creencias* falsas acerca del entorno y que un niño autista difícilmente desarrollará *creencias* sobre la situación emocional de los demás al observar sus expresiones faciales; y este conocimiento tiene más que ver con hechos cerebrales/físicos relevantes, identificados por *expertos*,⁵⁵ que con lo que sabemos acerca de la correcta aplicación de los términos desnudos de la psicología popular, y su utilidad

⁵⁵ Se sabe que el autismo tiene como causa un desarrollo neuronal anormal, y que la esquizofrenia, entre una pléyade de factores que incluyen algunos de índole social, indica una producción excesiva del neurotransmisor *dopamina*.

para interpretar y hacer racional la conducta de los demás. Este conocimiento inyecta nuestros términos y nociones de los estados mentales de un componente fisicalista del que carecían hace algunos siglos. Los términos de la psicología popular han sido *revisados* y *reformulados*, a partir de los resultados empíricos que derivan del estudio del cerebro y la mente, de suerte que ahora exhiben este talante *naturalista* ausente en la mayor parte de la historia del léxico de la psicología popular.

Otra forma en la que los avances en la neurofisiología influyen en la evolución de nuestros términos psicológicos se puede apreciar por el impacto que ha tenido dicho avance en cuestiones jurídicas. La capacidad de tener nociones del bien, el mal, los derechos y las obligaciones, y de realizar inferencias a partir de tales nociones, es fundamental para atribuir a los agentes responsabilidades respecto de las acciones que realizan u omiten, y que involucran a terceros. En muchos casos, tales capacidades se reconocen limitadas por factores físicos como lesiones cerebrales. Una persona que ha sufrido pérdida de tejido neuronal en la amígdala, puede tener dificultades para identificar y aprender de situaciones potencialmente peligrosas: *ceguera emocional*. Luego, para una persona en estas circunstancias puede ser complicado identificar escenarios riesgosos y, por lo tanto, difícilmente podrá experimentar *miedo*. Sin esta respuesta afectiva, es muy probable que seamos incapaces de reconocer lo que puede ser *bueno* o *malo* para nuestra subsistencia y la de los demás. Elizabeth Phelps (2005), describe el caso de una madre estadounidense que, como resultado de un tumor cerebral, sufrió daño considerable del tejido de la amígdala izquierda, y a la que se le diagnosticó el cuadro clínico arriba descrito: *ceguera emocional*. Sin la alarma emocional que representa tener una amígdala normal, la mujer fue incapaz de reconocer el riesgo que representaba el que su hija de 4 años entrara a la cocina, en donde sufrió quemaduras graves de tercer grado. Un jurado local —informado de la historia clínica de la madre y de la evidencia empírica en la que los expertos se han basado para identificar a la amígdala como una estructura fundamental para el aprendizaje emocional y la activación de la experiencia del miedo— la encontró *no responsable* del accidente de su hija, juzgándola incapacitada para construir una *representación completa de la situación en la que ocurrió el incidente*. Este caso, que de ninguna manera es único y aislado, hace evidente que nuestras nociones ordinarias de estados mentales han sido alcanzadas por lo que un grupo de expertos han logrado establecer respecto de la manera en que éstos dependen de propiedades físicas (cerebrales) relevantes. Los hechos demuestran cómo el avance de la empresa cognitivo-social de indagación coordinada —que se inauguró en

tiempos inmemoriales con la acuñación de los términos de estados mentales—, influye en la forma en la que interpretamos los estados mentales. Ejemplifican los efectos que la *apertura inductiva* y el *acceso epistémico* tienen sobre nuestros términos de psicología popular: la forma en la que, a partir de la investigación de un grupo de *expertos*, estos términos se *revisan* y *redefinen* a la luz de la evidencia empírica. La relación de estos hechos pretende dar apoyo a la imagen que propone Boyd acerca de la naturaleza *gradual* y *aproximativa* de la referencia y de los términos de estados mentales —y de clase natural en general—; en contra de la visión *rígida* o *inamovible* de los mismos, la cual puede presentar resistencia ante la posibilidad de revisar el significado de dichos términos. Al mismo tiempo, si los hechos sirven de algo y los hemos interpretado de la forma correcta, entonces también ayudan a desacreditar las premisas (*a*) y (*b*), fundamentales en dos de los argumentos de Hansberg en contra de que los (términos de) estados mentales son (términos de) clases naturales.

IMPRECISIÓN. Un expediente relacionado con la *dinámica* de los términos de clase natural, es el de la naturaleza de su definición. La tradición semántica respecto de los términos de clase natural, sea nominalista (Locke) o esencialista (Kripke), tiende a asimilar las ambiciones empiristas en torno la definición: es decir, persiguen celosamente el rigor y la precisión en las definiciones, además de entenderlas en términos de condiciones *necesarias* y *suficientes*. En el caso de la *Semántica Esencialista*, esta imagen de la definición, se halla ligada a una metafísica esencialista que entiende las esencias o estructuras explicativas —a partir de las que se definen los términos de clase natural— como conjuntos de propiedades físicas bien delimitados: *exhaustivos*. Esta imagen de la definición es una pieza fundamental para una construcción *fuerte* de la noción de ‘esencia’ o ‘estructura explicativa’.

Boyd, vale la pena establecerlo, también rechaza la concepción alternativa de la definición que proviene de Wittgenstein. La concepción que se basa en la noción de *parecido de familia*. Aunque esta propuesta incluye la ventaja de ubicar la definición más allá del atavismo empirista de la precisión, concediendo la posibilidad de entenderla, en algunos casos, como esencialmente *vaga* e *imprecisa*, su desventaja radica en que conduce veladamente hacia el nominalismo (Boyd, 1999). Según la noción de *parecido de familia*, lo que pone algo dentro de alguna familia no es la naturaleza, sino el *vago* acuerdo de la gente en una comunidad lingüística (Boyd, 1991). Si es el convenio el que traza los brumosos contornos de las definiciones de algunos términos de clase,

entonces, cuando se trate de términos de clase natural, no habrá ‘esencia’ o ‘estructura explicativa’ que cumpla una función relevante en la definición de dichos términos. La noción de *parecido de familia* nos aleja del realismo de clases naturales y nos conduce, por otra ruta, al nominalismo lockeano.

Boyd propone que las definiciones de términos de clase natural sean susceptibles a la *vaguedad o imprecisión*. E insiste en que esta característica de las definiciones no tiene por qué comprometer nuestros compromisos esencialistas. Pueden colegirse dos justificaciones para esta concepción de la definición. Como en el caso de *SE*, una es *semántica* y la otra es *metafísica*, y ambas están relacionadas. La justificación *semántica* para la *imprecisión* de las definiciones, se basa en la tipificación de la noción de ‘referencia’ expuesta al inicio de este apartado. La *apertura inductiva* y el *acceso epistémico*, junto con los *mecanismos* en los que se apoyan, *gradualmente* erigen, enriquecen, revisan y modifican las relaciones referenciales entre un término de clase natural y su referente —*acomodamiento*. Este ciclo, en cualquiera de sus etapas, aunque puede reducir la vaguedad definicional, raramente goza de la precisión (definiciones en términos de condiciones *necesarias* y *suficientes*) que pretende *SE*. La relación referencial es vaga en sus inicios y, en muchos casos, ni siquiera en etapas más avanzadas pierde este carácter *impreciso*. Los términos que refieren a enfermedades, especies biológicas y propiedades mentales, son ejemplos paradigmáticos de esta *imprecisión definicional*.

En este punto, podemos preguntar por la o las razones de la imprecisión definicional y reconocer dos posibles causas. La primera la explicaría por referencia a nuestra *imperfecta* capacidad de conocer las esencias o estructuras explicativas *reales* de las clases naturales en cuestión, razón por la cual nos resulta imposible formular definiciones precisas de los términos que las refieren. Esta explicación supone que las esencias son *exhaustivas*⁵⁶ y que, puesto que nuestros *mecanismos* no son suficientes o adecuados para develarlas, nuestro ‘conocimiento’ de la realidad se funda en otros hechos —convenio, aceptación de paradigmas científicos por razones sociales, políticas, económicas—, que no tienen que ver con las ‘esencias’ o ‘estructuras explicativas’ de las cosas. Otra vez el fantasma nominalista puede aprovechar la oportunidad para introducirse en

⁵⁶ Si, por el contrario, se presupone que los conjuntos de propiedades que constituyen las esencias son indeterminados, entonces no habría reproche que hacer ante la indeterminación de las definiciones, y no se culparía a nuestras *imperfectas* capacidades cognitivas como las responsables de tal indeterminación, puesto que ésta obedecería a la naturaleza misma de los conjuntos de propiedades que constituyen las esencias. Luego, aludir a nuestra *imperfcción* cognitiva como causa de la indeterminación de las definiciones, sólo tiene sentido si se asume que los conjuntos de propiedades que constituyen las esencias son exhaustivos —se presupone una construcción fuerte de la noción de ‘esencia’.

sus advocaciones operacionistas o constructivistas: las esencias reales están, de hecho, fuera de nuestro alcance y fuera de nuestras teorías, por lo que el conocimiento no resulta más que una forma de construcción social. Hacking (1991) parece avanzar en esta dirección (véase Boyd, 1991).

4. UNA METAFÍSICA REALISTA DE LAS CLASES NATURALES: CLASES DE RACIMOS DE PROPIEDADES HOMEOSTÁTICAS (RPH).

Pero hay otra forma de explicar la imprecisión de la definición que nos resguarda de seguir el sendero nominalista. Esta vía, como la esencialista, involucra una forma de metafísica sobre la naturaleza de la realidad. Según Boyd, la imprecisión definicional de (algunos de) los términos de clase natural, obedece a que las clases a las que refieren tienen como esencias lo que Boyd denomina *Racimos de Propiedades Homeostáticas (RPH)*, y no conjuntos *bien delimitados* de propiedades. Las clases *RPH*, son por naturaleza *indeterminadas* en su composición esencial, y esto explica por qué son *imprecisas* las definiciones de los términos que las refieren.

Las clases *RPH* tienen esencias constituidas por racimos indeterminados de propiedades, lo que significa que dichas esencias no tienen límites bien definidos, sino vagos. Esto descarta la posibilidad de identificar las clases naturales *RHP* (sus esencias o estructuras explicativas) en términos de condiciones necesarias y suficientes: descarta la posibilidad de entender sus esencias como *exhaustivas*. Otro rasgo característico de las clases *RHP*, es que instancian *mecanismos homeostáticos*, los cuales favorecen la co-ocurrencia de (la mayoría de) las propiedades que constituyen un *racimo* de propiedades explicativas. La presencia de tales mecanismos hace posible la integración de las propiedades del racimo, de suerte que la presencia de algunas tiende a favorecer la presencia de las demás: los mecanismos que favorecen la aparición de propiedades como ser líquido, insípido e incoloro, promoverán la co-ocurrencia de la propiedad de ser inodoro. Como sea, es posible que, en casos particulares (muestras de miembros de una clase putativa), algunas de las propiedades identificadas en los racimos (estructuras explicativas), no sean instanciadas. Las esencias de las clases *RPH* deben entenderse, no simplemente en términos estructurales (Quine) o disposicionales (Hacking), sino como racimos de *propiedades relacionales* unidas por medio de mecanismos homeostáticos.⁵⁷

⁵⁷ En efecto, puede hablarse de estructuras físicas que configuran en alguna medida la esencia de una clase natural (composición atómica, química, genómica, etc.), pero estas se entienden como racimos de propiedades (en los casos de clases *RPH*) más bien *vagos* en sus límites, y unificados por mecanismos homeostáticos que mantienen unidas y

Las clases *RPH* son *indeterminadas*, del mismo modo que las definiciones de los términos que las refieren son *imprecisas*. La *indeterminación* se reconoce como un atributo *metafísico* de la realidad, que se refleja en la *imprecisión* de las definiciones de los términos generales (de clase natural) que figuran, incluso, en nuestras mejores teorías; y que se proyecta en el carácter *aproximativo e imperfecto* de muchas leyes, generalizaciones e inducciones que se derivan de aquellas. Finalmente, una de las causas últimas de la imprecisión definicional se halla en la realidad misma: en la naturaleza *indeterminada* de las estructuras explicativas (*RPH*), y no solo en nuestras imperfectas capacidades cognoscitivas.⁵⁸ Esto nos exige renunciar a una construcción fuerte de la noción de ‘estructura explicativa’, como la que se haya implicada en *SE*, según la cual las esencias a partir de las cuales se definen los términos de clase natural, son *exhaustivas*. Por otro lado, la *imprecisión* definicional derivada de la *indeterminación* ontológica misma de las clases *RPH*, nos aleja de la postura que aboga también por la imprecisión y vaguedad definicional, pero que se basa en la doctrina wittgensteiniana del *parecido de familia*. Esta doctrina parece implicar un aspecto estipulativo/*a priori*, en tanto que sostiene que lo que finalmente determina el significado de un término general es un *vago acuerdo en una comunidad lingüística* sobre las o la propiedad relevante que define al término. La definición, aunque imprecisa, en este caso se funda en lo que una comunidad lingüística *conviene* en reconocer vagamente como lo esencial de una clase de fenómeno al que se mienta por medio de un término general. Así, la *imprecisión definicional* depende más de un conjunto de hechos y prácticas lingüísticas, que de la naturaleza misma de las clases a las que refieren los términos generales. En este sentido se considera *a priori* la definición en términos de *parecido de familia*, en la medida en que son prácticas lingüísticas (*a priori*), y no la investigación empírica (*a posteriori*), las que determinan las propiedades relevantes que ayudarán a definir un término general. Pero la *imprecisión* definicional que deriva del reconocimiento de clases *RPH*, no incluye estos elementos nominalistas/*a priorísticos*, ya que se funda en un *hecho acerca del mundo*, en la naturaleza misma de las clases naturales, y reconoce que es por medio de indagación empírica/*a posteriori* (*apertura inductiva*), que se

en *correlaciones* causales las propiedades que constituyen tales racimos. Naturalmente, los racimos de propiedades relacionales (unidas por los mecanismos homeostáticos) incluyen propiedades disposicionales: comportamientos típicos de las muestras que comparten una misma esencia (racimo de propiedades homeostáticas).

⁵⁸ Seguro que nuestras capacidades cognoscientes condicionan en alguna medida lo que podemos conocer y cómo lo conocemos, pero también es seguro que, con todas sus limitaciones, nos aportan recursos valiosos para conocer *muchos* aspectos de la realidad (*acomodamiento*).

identifican, aproximadamente, los indeterminados límites de las estructuras explicativas a partir de las cuales se definen, de forma *imprecisa*, los términos de clase natural.

Ahora bien, ¿qué argumentos se pueden ofrecer para sostener el tipo de metafísica que implica la postulación de clases *RPH*? ¿Qué razones existen para pensar que la naturaleza (en muchos casos) no corta tan fino en sus coyunturas ontológicas y propicia la existencia de clases naturales *indeterminadas* y, quizás, *dinámicas*? Las razones se pueden encontrar precisamente en los casos que han representado dificultades para el *esencialismo metafísico de clases naturales*. Ya se ha visto cómo la construcción fuerte de la noción de ‘esencia’ o ‘estructura explicativa’ (I, III y I, v.I) ha conducido a muchos filósofos a la controversial conclusión de que las especies biológicas no son clases naturales, sino individuos. La propuesta realista *semántico-metafísica* de Boyd, proporciona un marco teórico alternativo, dentro del cual puede entenderse el problema de las especies en términos menos radicales.

Según la propuesta de Boyd, las especies biológicas son clases *RPH*, cuya esencia está constituida por un racimo de propiedades homeostáticamente relacionado, compartido imperfectamente por sus miembros: propiedades morfológicas, fisiológicas y conductuales. El papel definicional de los mecanismos homeostáticos puede verse en la función que tiene la cruce entre diferentes poblaciones en el concepto moderno de ‘especie’: a partir del intercambio de material genético entre diferentes poblaciones se presenta, en las poblaciones intermedias, cierta unidad homeostática de propiedades características de ambas especies (*indeterminación*); esto se refleja en las definiciones *imprecisas* de las especies que resultan de la cruce de poblaciones distintas. La *indeterminación necesaria* en los racimos de propiedades que definen las especies es una consecuencia de la teoría de la evolución (*variación*). Como Darwin observó: la especiación depende de la existencia de poblaciones que son intermedias entre las antecesoras y la que emerge. Luego, si tenemos presente el precepto metafísico de la *indeterminación ontológica de las esencias* implicado en la doctrina de las clases *RPH*, no resulta necesario sostener que las especies son individuos. Al contrario, parece razonable sostener que, por su carácter *indeterminado* y *dinámico* —evolutivo—, las especies ejemplifican un caso de clases naturales del tipo *RPH*. En la misma dirección, piensa Boyd, cualquier pretensión por refinar los criterios de clasificación (ir en pos de definiciones en términos de condiciones necesarias y suficientes) podría oscurecer el hecho central de las variaciones hereditarias fenotípicas de las que la evolución biológica depende, lo cual resultaría ser inapropiado y científicamente desorientador (Boyd, 1991; 1999).

Es necesario reconocer que el caso de las especies, interpretado en términos de las clases *RPH*, soporta sólo heurísticamente la metafísica realista de clases naturales que propone Boyd. Las características de las teorías y explicaciones con las cuales contamos sobre la naturaleza y comportamiento de las especies, hacen plausible la concepción metafísica implicada en la doctrina de las clases *RPH*: una metafísica según la cual existen clases naturales *indeterminadas* y *dinámicas*. Pero no sólo las especies aportan razones heurísticas que apoyan esta metafísica; varias disciplinas desarrollan teorías y corren inducciones a partir de la postulación de entidades, procesos, organismos y/o sustancias que tienen los rasgos distintivos de las clases *RPH*. Los variados e indeterminados fenómenos biológicos que se denotan con el término ‘enfermedad’ en medicina, y los complejos procesos y capacidades cognitivas que se estudian en la neuropsicología, representan ejemplos conspicuos de *clases* de entidades o fenómenos que postulan distintas disciplinas, y que aparentan tener el carácter *indeterminado* y *dinámico* distintivo de las clases *RPH*.

5. UNA CONCEPCIÓN ALTERNATIVA DE CLASE NATURAL (*RPH*).

La semántica y la metafísica realista propuestas por Boyd sostienen una concepción alternativa de lo que son las clases naturales. Las tesis que integran dicha concepción han sido enunciadas por Boyd en distintas ocasiones (Boyd, 1979/1993; 1988, 1999) y ahora es oportuno citarlas:

1. Existe una familia *F* de propiedades contingentemente⁵⁹ agrupadas en la naturaleza en el sentido de que co-ocurren en un número importante de casos. *F* constituye la estructura explicativa (esencia) de las clases *RPH*.
2. Las propiedades en *F* están unidas homeostáticamente: la presencia de una propiedad de la familia *F* tiende a favorecer (en condiciones normales) la presencia del resto. Existen mecanismos homeostáticos (causales) que tienden a mantener unidas las propiedades.
3. Existe un término de clase *t* que se aplica a todas las cosas (en el mundo actual) en las que la mayoría del racimo de propiedades de *F* ocurren.
4. *t* no tiene una definición analítica.⁶⁰ Todas o la mayoría de las propiedades homeostáticas del racimo, así como todos o la mayoría de los mecanismos subyacentes proveen la defi-

⁵⁹ Contingente y no metafísicamente agrupadas por la naturaleza... como lo estipularía la metafísica esencialista de clases naturales.

⁶⁰ En contra de cualquier forma de *nominalismo/convencionalismo/a priorismo*.

nición de la clase. Qué propiedades y mecanismos formarán parte de la definición de t —constituirán la estructura explicativa o esencia de la clase— es una cuestión *a posteriori*: el gradual y progresivo *acceso epistémico* que es posible mediante la *apertura inductiva* decide *empíricamente* lo que constituye la esencia de una clase *RPH*.

5. *Indeterminación ontológica*: la imperfección nomológica es posible o real. Una cosa puede desplegar algunas propiedades del racimo homeostático y otras pueden estar ausentes. Lo mismo en relación con los mecanismos homeostáticos subyacentes.
6. En tales casos la importancia relativa de las propiedades de F y de los mecanismos subyacentes presentes determinarán si se puede aplicar el término t : esto se decide *a posteriori* y no conceptualmente.
7. Es posible la *imprecisión definicional/extensional*. Habrá cosas que no tengan todas las propiedades de F (o en las cuales no operen todos los mecanismos subyacentes).
8. La importancia causal del racimo de propiedades homeostáticas F , y la relevancia de los mecanismos subyacentes es tal que la clase o propiedad denotada por t es una clase natural.
9. La sustitución de t por un término más preciso y extensionalmente menos vago conducirá a la pérdida de naturalidad (entendida en términos de relevancia causal y explicativa). El refinamiento desmesurado puede llevarnos a considerar importantes distinciones que no lo son para la explicación e inducción o a ignorar similitudes que sí lo son.
10. El racimo de propiedades homeostáticas que sirve para definir t no se individua de forma extensional. El racimo se individua como un proceso u objeto histórico. Ciertos cambios en las propiedades o mecanismos pueden presentarse a lo largo del tiempo sin que el objeto o proceso deje de ser la clase que es. Por lo tanto las propiedades que determinan t pueden variar en el tiempo mientras que t tiene la misma definición.

Las tesis (1-7) compendian las afirmaciones sustanciales de la semántica y la metafísica realista de Boyd que se han glosado en los apartados anteriores: el carácter *gradual* y epistemológico del fenómeno de la referencia, y la naturaleza ontológicamente *indeterminada* de muchas clases de cosas, organismos o sustancias, que deriva en la naturaleza *imprecisa* de las definiciones de los términos que refieren a dichas clases. (8) representa una novedad a medias. Es una novedad porque expresa sucintamente la noción de naturalidad que suscribe Boyd, la cual no había sido citada

o discutida hasta ahora. ‘A medias’ porque Boyd, en este particular, concuerda con Platts, Hansberg y LaPorte (I, II.IV, 3), al considerar que lo que confiere ‘naturalidad’ a distintas clases de objetos, organismos y/o sustancias es su valor explicativo: donde el *valor explicativo* se funda en las propiedades físicas relevantes —descubiertas *a posteriori*— que distinguen a los miembros de dichas clases, y a partir de las cuales se definen los términos generales (de clase) que figuran en las teorías y explicaciones que son fundamentales para correr generalizaciones e inducciones exitosas sobre fenómenos que involucran, precisamente, a los miembros de dichas clases (Boyd, 1991; 1999). (9) constituye una crítica metodológica hacia la concepción tradicional de las clases naturales (la que deriva de *SE*) que insiste en definiciones *precisas*, y que supone una metafísica *exhaustiva* de las mismas. Sobre este particular, Boyd considera que la búsqueda obsesiva del refinamiento definicional puede desviarnos de la empresa explicativa de las teorías científicas, y de la finalidad de *acomodamiento* que se busca entre aquellas y la realidad. La precisión definicional tiende más a satisfacer atavismos filosófico-semánticos, que a satisfacer las demandas explicativas y de *acomodamientos* que se exigen a las teorías científicas. Finalmente (10), concibe la posibilidad *ontológica-metafísica* de que las clases naturales tengan una naturaleza *histórico-dinámica*: que sus esencias o estructuras explicativas sufran modificaciones o alteraciones a través del tiempo.

Esta concepción alternativa de lo que son las *clases naturales*, que se funda en una noción más *tolerante* de ‘esencia’ o ‘estructura explicativa’ —por entenderla, en muchos casos, como metafísicamente *indeterminada* y *dinámica*— servirá como marco conceptual para abordar la cuestión del estatus de las emociones en tanto que clases naturales. En tanto que teoría filosófica-metafísica sobre la realidad, la metafísica realista de las clases *RPH* ganará fuerza en la medida en que ella misma responda a las demandas de *acomodamiento* que describe Boyd. Dicho *acomodamiento* tendrá lugar en la medida en que el análisis de distintas teorías en distintas ramas del conocimiento empírico, revele que las clases de entidades o fenómenos que postulan, constituyen clases *RPH*; o que, en otras palabras, se reconozca que son más inteligibles si se las interpreta como postulando la existencia de clases *RPH*. Así, una forma práctica de argumentar a favor de la metafísica indeterminada y dinámica de las clases *RPH*, consiste en analizar las propiedades que se atribuyen a las clases de fenómenos o entidades postulados en teorías de distintas disciplinas, y mostrar que configuran *Racimos de Propiedades Homeostáticas*. En este sentido, si bien no se han ofrecido argumentos contundentes a favor de la metafísica de clases naturales *RPH*

—*indeterminista y dinámica*— que propone Boyd, es posible que el análisis del expediente de las emociones en filosofía, psicología y neurofisiología, nos conduzca a reconocer a muchas de ellas, individualmente o en su conjunto, como clases *RPH*. Si este es el caso, habría razones para hablar de la verdad de la metafísica y la semántica *indeterminada y dinámica* que suscribe Boyd. Dispuestas así las nociones, tenemos que la verdad de la tesis de que las emociones son clases naturales —entendidas como clases *RPH*— aportará para el *acomodamiento* entre la teoría metafísica de clases naturales de Boyd y la estructura de la realidad. En lo que hace a este trabajo, consideramos que es posible que el caso de las emociones abone más de lo que suele pensarse a favor de dicho *acomodamiento*.

PARTE II

EMOCIONES Y CLASES NATURALES

I. INTRODUCCIÓN

El objetivo general de esta segunda parte consiste en articular un marco teórico y conceptual en el cual haya espacio para construir una Teoría General de las Emociones. Es decir, una teoría que mantenga unida, bajo la misma categoría ontológica y epistemológica, a las emociones. En términos un poco más precisos, nos proponemos argumentar a favor de que las emociones (básicas y no-básicas) pertenecen a una misma categoría o clase natural, en virtud de que comparten propiedades y mecanismos causales que hacen posible una teoría unificada, explicativa y válida, para todas o la mayoría de ellas.

Parte del aparato conceptual requerido se ha manufacturado en la primera parte de esta investigación, al revisar las distintas concepciones de clase natural, y al exponer las características y ventajas que tiene la propuesta de Boyd, que define ‘clase natural’ como un *Racimo de Propiedades Homeostáticas*. En esta segunda parte, también será necesario allanar el camino hacia nuestro objetivo. La discusión sobre el estatus de las emociones en tanto que clases naturales no carece de opiniones encontradas. Por tal razón, el propósito de los dos primeros capítulos consiste en bosquejar el escenario de la discusión actual en torno al estatus de las emociones en tanto que clases naturales: identificaremos los bandos y señalaremos cómo están distribuidas las lealtades. Esto es de utilidad por dos razones. Por un lado, ayudará a ubicarnos en el flanco que nos interesa defender y, por otro, favorece la familiarización con las objeciones que se lanzan en nuestra dirección. Así las cosas, en el primer capítulo comenzamos delineando el espacio conceptual de las posibles posiciones en torno al estatus de las emociones en tanto que clases naturales. Se reconocen al menos tres posturas y se discuten algunas de sus implicaciones. Las tres posturas en cuestión son: (i) Cada una de las emociones básicas es, individualmente, una clase natural. (ii) El conjunto de las emociones básicas constituye una clase natural. (iii) Las emociones (lo cual incluye a las emociones básicas y a las que no lo son) constituyen una sola clase natural. En el segundo capítulo, identificamos a los autores y glosamos brevemente las motivaciones que tienen para adscribir una u otra de las tres posturas. Por supuesto, esta investigación se moverá en la dirección de la postura (iii).

Antes de comenzar la defensa formal de (iii) o algo muy cercano a esta postura, analizamos detalladamente, en el capítulo IV, una versión de la postura (ii): la cual afirma que

las emociones básicas constituyen una clase natural. Nuestro análisis parte de la defensa que hace Griffiths (1997) —apoyándose en la noción de clase natural de *racimo de propiedades homeostáticas*— de la idea de que las emociones básicas son programas afectivos y que, como tales, evolucionaron al mismo tiempo y de los mismos principios, de suerte que todas comparten las mismas propiedades fisiológicas y los mismos mecanismos causales. En la misma dirección, se discuten algunas objeciones en contra de la idea, implícita en la postura y argumentos de Griffiths, de que las emociones básicas conforman una clase natural. En este punto, vale la pena señalar que, ya que suscribe abiertamente (ii), Griffiths muestra un abierto rechazo hacia (iii). De hecho, considera que debe haber una línea divisoria ontológica y, por lo tanto, epistemológica, entre las emociones básicas y lo que él llama emociones cognitivas superiores (vergüenza, orgullo, amor, etc.)¹. Lo que resulta relevante para nuestro objetivo de articular una teoría general de las emociones, es que la propuesta de Griffiths en relación a las emociones básicas como una clase natural, ofrece un marco conceptual y metodológico que, a pesar de las reservas del propio Griffiths, consideramos que puede proyectarse al resto de las emociones. La idea es que las emociones básicas comparten un racimo de propiedades a nivel fisiológico (cambios musculo-esqueléticos, expresivos, vocales y endocrinos) y a nivel de arquitectura cognitiva (modularidad). Nosotros propondremos que, además de propiedades fisiológicas y modularidad, otro tipo de propiedades serán pertinentes para individuar a las emociones.

En el capítulo v se exponen once objeciones, planteadas por distintos autores, en contra de la tesis (iii) que sostiene que las emociones —las básicas y las que no lo son— conforman una sola clase natural. En el mismo capítulo se ofrecen réplicas para cinco de ellas. Las seis objeciones que permanecen sin refutación en este capítulo, cinco de ellas avanzadas por Griffiths (1997) y una de ellas por LeDoux (1996), configuran un conjunto más difícil de rebatir. Las cinco objeciones de Griffiths, que pretenden abonar a la tesis de la desunión de las emociones identificando diferencias cruciales entre las Emociones Básicas (EB) y las que denomina, Emociones Cognitivas Superiores (ECS), afirman, *grosso modo*, lo siguiente:

¹ Lo que aquí significa división epistemológica es que deben existir una explicación y leyes (teoría) distintas para las emociones cognitivas superiores.

- (1) Las ECS requieren y disponen de recursos conceptuales innecesarios en las EB.
- (2) Las EB aparecen en el primer año, mientras que las ECS se presentan hasta los tres años, aproximadamente.
- (3) No se sabe que las ECS involucren patrones de cambios fisiológicos distintivos y/o expresiones características.
- (4) Las EB parecen exhibir modularidad, mientras que las ECS no.
- (5) La función adaptativa de las EB es de tipo *organismo-entorno*, mientras que la función adaptativa de las ECS es *resolver problemas de compromiso*.

La objeción de LeDoux sostiene que no hay evidencia suficiente para sustentar la idea de un sistema neuronal especializado en la producción de emociones. Nuestras respuestas a estas objeciones se articulan en los capítulos posteriores, mientras se exponen las características del modelo de teoría general de las emociones que proponemos y se recaba más evidencia para replicar.

En el capítulo VI se trabaja en la elaboración del marco teórico y conceptual para una teoría de las emociones. En primer lugar, reiteramos nuestra preferencia por la concepción de clase natural como Racimo de Propiedades Homeostáticas (RPH), en virtud de que proporciona un marco conceptual más flexible y liberal. Los racimos de propiedades que constituyen la estructura explicativa de una clase natural incluyen sólo propiedades físicas relevantes, las cuales están vinculadas de forma causal-homeostática. El carácter *indeterminado* y *dinámico* de las clases naturales RPH es idóneo para encuadrar fenómenos un tanto difusos en marcos teóricos respetables, como las especies y enfermedades en biología, y las capacidades cognitivas y las emociones en psicología. El siguiente expediente de este capítulo consiste en responder a la cuestión de qué entenderemos por ‘propiedades físicas’. Ya que consideramos incluir propiedades psicológico-funcionales y fenomenológicas como componentes imprescindibles de la estructura explicativa de las emociones, es importante ofrecer alguna justificación para su inclusión. Y es necesario, ya que, desde el punto de vista filosófico, el estatus ontológico de las propiedades psicológicas y fenomenológicas está lejos de considerarse resuelto. En este sentido, optamos por *asumir*, como lo hacen la mayoría de las aproximaciones neurofuncionales, que las propiedades en cuestión son físicas, toda vez que la investigación empírica ha mostrado que están íntimamente vinculadas al apropiado funcionamiento de ciertas estructuras, circuitos y sistemas neuronales. Aunque

no hay consenso en relación a la naturaleza del vínculo, principalmente entre las propiedades fenomenológicas y las propiedades neuronales con las que se las asocia, nos inclinamos a adoptar una postura fisicalista. Nuestras razones, como se verá, son más bien epistemológicas. Sencillamente, no sería aceptable excluir de la ecuación, en una explicación de las emociones, sus aspectos psicológicos y fenomenológicos, ya que hay buenas razones para pensar que están vinculadas *causalmente* con otras propiedades físicas —neuronales. En este capítulo también se dice algo acerca de la metodología que guía la construcción de una teoría de las emociones. Básicamente, consideramos que es fundamental sustentar cada concepto y tesis de la teoría con investigación y datos empíricos, y que la pluralidad de enfoques y aproximaciones en el estudio de las emociones es crucial en este momento. A esta estrategia metodológica la denominamos *apertura inductiva*. El capítulo concluye con una lista y caracterización tentativa de diez propiedades —de niveles explicativos distintos— que, según nuestra propuesta, constituyen la estructura explicativa de las emociones.

En el capítulo VII aportamos argumentamos y datos para sustentar la tesis (iii) que sostiene que las emociones —Básicas y Cognitivas Superiores— conforman un solo dominio ontológico y explicativo. La forma en la que intentamos apoyar esta tesis consiste en mostrar que las emociones —EB o ECS— instancian las diez propiedades que constituyen su estructura explicativa. En este sentido, mostramos que todas las emociones instancian propiedades neurofisiológicas y expresivo-conductuales; propiedades psicológicas —como ser representacionales y dar lugar a efectos psicológicos característicos; propiedades biológicas —como tener una función adaptativa; propiedades o rasgos de su arquitectura —como *cuasi-modularidad*, sistema de recalibración y duración breve; y propiedades fenomenológicas. Si tenemos éxito en este sentido, habríamos abonado bastante a la idea de que las emociones conforman una clase natural y, en consecuencia, un dominio ontológico y explicativo único. Al final de este capítulo ofrecemos una caracterización de lo que es una emoción. Según nuestra teoría, una emoción es una instancia de una clase natural RPH, cuya estructura explicativa está constituida por diez propiedades físicas —físicas, psicológicas, biológicas, de arquitectura y fenomenológicas— articuladas de forma homeostática causal.

Finalmente, en el capítulo VIII, se intenta mostrar que la teoría de las emociones que proponemos es viable, en la medida en que puede proyectarse sobre dos tipos concretos de emociones, independientemente de que sean básicas o cognitivas. En particular, mostramos cómo pueden lograrse descripciones adecuadas de emociones como el orgullo y el miedo, en términos de las diez propiedades que, según nuestra teoría, debe instanciar, aproximadamente, cada emoción. El último expediente que se aborda en este capítulo y en la investigación en general, tiene que ver con el impacto que tendría nuestra teoría respecto de la ontología emocional. Ya que en algún momento argumentaremos a favor de la posibilidad de que las emociones cognitivas superiores tengan homólogas básicas y de que un solo tipo de emoción pueda dividirse en varias subespecies, resultará que nuestro léxico emocional ordinario resultará insuficiente para reflejar la variedad real de la ontología emocional. Este resultado parecerá extraño, dada la aparentemente rica variedad de distinciones emocionales que son posibles gracias a los nutridos léxicos emocionales. Según nuestras reflexiones, tales distinciones fallan en reflejar distinciones psicológicas y ontológicas reales, en virtud de que todo lo que hacen es poner énfasis en distintos aspectos de las experiencias afectivas: la causa de la emoción, la fenomenología, los cambios fisiológicos y otros factores contextuales.

II. TRES TESIS SOBRE EL ESTATUS DE LAS EMOCIONES EN TANTO QUE CLASES NATURALES

Podemos comenzar distinguiendo tres tesis en torno al estatus de las emociones en tanto que clases naturales:

- i) Cada una de las emociones básicas es, individualmente, una clase natural.
- ii) El conjunto de las emociones básicas constituye una clase natural.
- iii) El conjunto de los fenómenos emocionales (lo cual incluye a las emociones básicas y a las que no lo son) constituye una clase natural.

Nótese que por lo menos dos de las afirmaciones suponen la difundida, aunque no poco polémica distinción entre emociones básicas y no-básicas.² Aunque se reconoce que es problemática, adoptamos la distinción para esclarecer el punto de vista que es relevante para esta investigación. Más adelante se discutirá el caso de las emociones básicas.

En cuanto a las tesis mismas, puede advertirse que difieren en su grado de generalidad y cada una implica, por lo tanto, una ontología y taxonomía peculiar: (i) sostiene que las emociones básicas (miedo, alegría, enojo, sorpresa, asco y tristeza) tienen, individualmente, un conjunto de características distintivas o estructura explicativa, la cual determina las leyes y regularidades distintivas de cada una. Por otro lado, (i) niega que el conjunto de las emociones básicas, y las emociones en general, compartan una esencia común, en virtud de la cual estarían sujetas a las mismas leyes y regularidades. (ii) afirma que el conjunto de las emociones básicas constituyen una clase natural en la medida en que comparten un conjunto de características distintivas o estructura explicativa, que determina las leyes y regularidades que las gobiernan. Por otro lado, (ii) implica que, respecto de las emociones no-básicas —en tanto que distintas de las emociones básicas—, no sabemos si tienen una estructura explicativa o están sujetas a leyes y regularidades típicas. Por lo tanto, no podemos decir si constituyen o no una clase natural. (iii) sugiere que las emociones, básicas y no básicas, conforman un conjunto de fenómenos que comparten características distintivas o una estructura explicativa, la cual determina las leyes y regularidades que las gobiernan. Finalmente, (iii) supone que, aun si existe una diferencia entre emociones básicas y no básicas, ésta no debería ser una diferencia tan radical como para evitar que pudieran compartir una esencia.

En cuanto a las consecuencias, podemos reconocer que (i) implica que subsumir bajo una misma categoría a las emociones básicas —como lo hacemos ordinariamente—

² Existen diferentes formas de describir las emociones no-básicas. Se puede hablar de ellas como emociones complejas (en el sentido de que están compuestas por emociones básicas), o como emociones cognitivas (en el sentido de que implicarían procesos psicológicos más complejos... como opuestos a procesos meramente fisiológicos) o como emociones culturales (en el sentido de que la posibilidad de su experimentación o existencia depende de factores sociales). La razón por la que rehusamos utilizar, por ahora, cualquiera de estas descripciones, es que no existe consenso en torno a cuál sería la más apropiada. La descripción que se elija para distinguirlas, muy probablemente dependerá del tipo de aproximación metodológica que se adopte al estudiarlas. Dejamos como una cuestión abierta (por ahora) la forma en que se distinguen las emociones básicas de las no-básicas. En cuanto a las emociones básicas, Paul Ekman (1992 y 1994) y James Russell (1994) han sostenido una acalorada contienda en torno su realidad o existencia.

constituiría un craso error. El escenario que plantea (i) suscita algunas inquietudes ontológico-terminológicas. Ya que las emociones básicas no componen un dominio natural, no deberíamos tener un término teórico genérico ('emoción') para aglutinarlas y designarlas. Además, los términos que denotan las emociones básicas que postulan los expertos son los mismos que figuran en nuestro léxico emocional. La cuestión es, ¿tienen el mismo referente? La tesis (ii) constituye una posición aparentemente más apegada a nuestras categorizaciones psicológicas ordinarias, según las cuales, todo aquello a lo que denominamos emociones básicas (miedo, enojo, tristeza, asco, alegría, etc.) caería bajo una y la misma categoría general. Pero (ii) también suscita la cuestión de si los términos que designan las emociones básicas tienen el mismo referente que los términos homónimos que figuran en nuestro léxico emocional vernáculo. (ii), además, abre la puerta a la cuestión de cómo habríamos de referirnos a las emociones no-básicas. En cuanto a (iii), en tanto que mantiene en el mismo grupo al conjunto de los fenómenos a los que llamamos habitualmente 'emoción', no da lugar a ninguna de las cuestiones que suscitan (i) y (ii). Las consecuencias de (i) y (ii), por lo tanto, apuntarían a una revisión severa de nuestra ontología y terminología emocional de sentido común: la verdad de cualquiera de estas tesis implicaría que nuestro discurso ordinario sobre emociones, en realidad, referiría a una amalgama de diferentes fenómenos, y el trabajo científico consistiría, entonces, en desfragmentar y distinguir los distintos tipos de fenómenos que constituyen dicha amalgama. En última instancia alguna forma de eliminativismo de mayor o menor escala resultaría inevitable (Griffiths, 1997, pp. 14-17).

En cuanto a la tesis (i), desconocemos si alguien la ha defendido o suscrito. Su mención tiene como finalidad agotar el espacio lógico de nuestra discusión, de suerte que se puedan definir e identificar con mayor precisión la o las tesis que nos interesa explorar con mayor detalle. En adelante, (i) será relegada, en virtud de que, en nuestro punto de vista, no refleja un compromiso lo suficientemente fuerte con la tesis de que las emociones son clases naturales —como sí lo hacen (ii) y (iii) . En cuanto a la tesis (ii), aunque discutible, parece contar con buenos fundamentos y aceptación. No obstante, será indispensable examinarla con más detalle para poder juzgar si acaso, en el sentido que lo sugiere la tesis, estamos justificados a conferir el carácter de clase natural a las emociones básicas. (iii) es la más polémica y ambiciosa de todas las afirmaciones. En torno a (iii), las voces expertas,

—que incluyen a filósofos, psicólogos, neurólogos, etc.— forman un robusto coro que sentencia: “aquello a lo que llamamos emociones (en conjunto) no conforman una clase natural”. Es un objetivo fundamental de este trabajo examinar las razones detrás de esta negación multitudinaria y explorar la posibilidad de revertir el veredicto. Las reflexiones precedentes en torno a la noción de ‘clase natural’ y a la supuesta imbatibilidad de la tesis de la realización múltiple, han tenido como finalidad servir a este propósito.

III. LA INTEGRACIÓN DE LOS BANDOS

Existe cierto consenso en relación a la idea de que las emociones no conforman una clase natural en ninguno o algunos de los sentidos posibles. El grupo de resistencia está conformado, hasta donde sabemos —principalmente por noticia de Charland, (*Op. cit.* p. 511)—, por Barrett (2006), Ben-Ze’ev (2000), de Sousa (1987)³, Elster (1999), Griffiths (1997), Hansberg (1990; 1996)⁴, Neu (2000), A. Rorty (1978) y Solomon (1995). Griffiths niega (i) y (iii), pero acepta (ii). Del resto, ninguno consideraría verdadera la disyunción de (i), (ii) o (iii). Lo que podría explicar esta diferencia, es que Griffiths es el único consciente del acalorado debate en torno a la noción de clase natural y, por lo tanto (en la línea de nuestro trabajo), reconoce que la cuestión del estatus de las emociones en tanto que constitutivas de una clase natural, exige una meticulosa discusión en torno a qué ha de contar como una clase natural. Por otro lado, el resto del grupo, no exhibe posiciones críticas en relación a la noción de clase natural, por lo que tiende a suscribir posturas como las que critica Boyd (1991; 1999)⁵, según las cuales una clase natural se especifica por medio de un conjunto *exhaustivo y estático* de propiedades (esencia), que sirven a su vez como condiciones *necesarias y suficientes* para delimitar su extensión. En otro sentido, Griffiths y Barrett también defienden un punto de vista valioso para nuestro enfoque: consideran que el estudio e investigación en torno a las emociones es una cuestión eminentemente *empírica*, por lo que

³ De Sousa constituye un caso peculiar, ya que reconoce con algo de ironía que los filósofos que afirman que las emociones no constituyen una clase natural, al final, lanzan a todo lo que llamamos ‘emoción’ al mismo saco.

⁴ La inclusión de Hansberg en esta lista se justifica de formas distintas (dependiendo de la fuente). Ya que en su artículo de 1990 argumenta en contra de que los estados mentales en general sean clases naturales, por extensión, se sigue que rehusaría considerar a las emociones como clases naturales. En el caso de su libro de 1996, la heterogeneidad de aspectos que reconoce como incluidos en lo que llamamos emociones, parecen conducirla a negar la posibilidad de que las emociones constituyan una clase natural (véase, principalmente, la Introducción).

⁵ *Supra*, Parte I, IV.II.

consideran insuficiente la metodología del *análisis conceptual* para elucidar la pregunta de ‘*qué son las emociones*’. El resto de los autores, en contraste con esta actitud escéptica respecto del análisis conceptual, sus méritos y límites, lo practican y utilizan como único recurso metodológico en sus investigaciones en torno a las emociones.

En términos positivos, Griffiths sostiene que existen seis emociones básicas o programas afectivos: enojo, miedo, asco/repugnancia, sorpresa, regocijo y tristeza. Estos programas afectivos son respuestas adaptativas modeladas evolutivamente. Activan distintas respuestas a nivel fisiológico que exhiben homeostasis causal. Dada esta descripción, Griffiths simpatiza con la concepción de clase natural que defiende Boyd (clases *RPH*) e incluye a la homología como uno de los criterios fundamentales para individuar a los programas afectivos. En el siguiente capítulo se exponen y evalúan con mayor detalle los argumentos que ofrece Griffiths para sustentar (ii).

Jaak Panksepp (1998; 2000), como indica Charland, también suscribe la idea de que existen emociones básicas (ii). A diferencia de Griffiths, Panksepp cuenta siete emociones básicas (búsqueda, ira, miedo, lujuria, cuidado, pánico y juego)⁶ y las llama ‘sistemas emocionales’. Panksepp también piensa que dicho repertorio afectivo lo compartimos —en virtud de una historia evolutiva parcialmente compartida— con el resto de los mamíferos, y que está codificado en estructuras y procesos químicos a nivel neuronal (Panksepp, 2000, p. 137). Panksepp, en aparente concordancia con la concepción de clase natural de Boyd, parece conceder que los sistemas emocionales constituyen racimos *indeterminados* de propiedades neuronales y neuroquímicas que especifican e individúan los diferentes sistemas emocionales (*ibid.* p. 144). Finalmente, en contra de lo que Griffiths sostiene, Panksepp no parece pensar que haya una discontinuidad ontológica entre las emociones básicas y las emociones adquiridas o sociales⁷. Considera que las emociones adquiridas o sociales pueden derivarse de los sistemas afectivos básicos con la ayuda del aprendizaje, mecanismos de asociación corticales y del lenguaje (*ibid.* p. 138).

⁶ Por razones que no alcanzo a entender, Charland (2002, p. 515) propone una lista de sistemas emocionales que no concuerda con la que proporciona Panksepp (2000, p. 144, cuadro 9.1). Charland no menciona el miedo y el pánico y, en su lugar, habla de la alegría y la aflicción.

⁷ Griffiths habla de ‘Emociones Cognitivas Superiores’ y considera que constituyen un dominio psicológico aparte y distinto del de las emociones básicas o Programas Afectivos (Griffiths, 1997, cap. 5).

En cuanto a (iii), actualmente pocos se aventuran a dar la cara por ella. Louis Charland (2002) es un raro ejemplar de esta especie. Sugiere que el pesimismo no está justificado, y que podemos encontrar buenas razones para pensar lo contrario; esto es, que las emociones sí conforman una clase natural, ya que comparten características distintivas (estructuras explicativas), en virtud de las cuales están sujetas a sus propias leyes y regularidades. Charland hereda y utiliza mucha de la metodología que desarrolla Griffiths (1997) para defender la más ambiciosa tesis (iii): concibe la cuestión como una de índole *empírica*; adopta la noción boydiana de clase natural (clases de *Racimos de Propiedades Homeostáticas*) y, con ella, asume que la homología desempeña un papel crucial para individuar a las emociones. Los resultados de este experimento también serán evaluados.

IV. GRIFFITHS: EMOCIONES BÁSICAS COMO CLASES NATURALES

La concepción de clase natural que suscribe Griffiths incluye rasgos e implicaciones que ya se han identificado en la primera parte de esta investigación (I, v.II). En términos metafísicos, las entiende al estilo de Boyd, como clases de *Racimos de Propiedades Homeostáticas*; lo que significa que sus estructuras explicativas están constituidas por conjuntos (*indeterminados y dinámicos* —esto es, susceptibles de evolución a través del tiempo) de propiedades físicas causalmente vinculadas. En términos epistemológicos, sostiene que las clases naturales son indispensables para capturar y establecer generalizaciones en las diferentes ciencias. En este sentido, las clases naturales resultan fundamentales para la explicación y la predicción. Al discutir cómo debería entenderse la ‘naturalidad’ de las clases naturales (I, II.IV, 3), siguiendo a Platts (1997), se las llamó también *clases explicativas* por desempeñar el papel epistemológico que subraya Griffiths. En términos ontológicos, las clases naturales sirven para especificar la ontología de una ciencia. En este sentido, revelan los compromisos ontológicos de las diferentes ciencias.

IV.I. NATURALEZA DE LAS EMOCIONES BÁSICAS O PROGRAMAS AFECTIVOS

Al aplicar este paradigma de clase natural al caso de las emociones, Griffiths concluye que existen ciertas emociones básicas que son clases naturales⁸. Apoyándose en el trabajo e investigación de Paul Ekman, reconoce que existen seis emociones básicas o ‘Programas

⁸ Esto es, suscribe la postura (ii), expuesta en el primer capítulo de esta segunda parte.

Afectivos'. Griffiths los describe como programas o respuestas adaptativas a situaciones que son relevantes para la subsistencia y descendencia de un organismo. Son respuestas *complejas, coordinadas y automáticas*. Los programas afectivos involucran (a) expresiones faciales; (b) cambios musculoesqueléticos; (c) cambios vocales; (d) respuestas hormonales o endócrinas; y (e) del sistema nervioso (Griffiths, 1997, p. 77)⁹. Se dice que son respuestas *coordinadas* porque ocurren juntas, en patrones reconocibles. Son *automáticas* porque se despliegan de forma coordinada sin necesidad de activación consciente. Vale la pena señalar que Griffiths entiende que los cambios implicados en cada Programa Afectivo (a-e), los describen en un nivel *fisiológico*, y no en un nivel *neurológico* (*ibid.*).

IV.II. ARGUMENTOS, EVIDENCIA Y OBJECIONES EN TORNO A LOS PROGRAMAS AFECTIVOS

Griffiths considera que existe evidencia sólida para sostener que cada Programa Afectivo se individualiza por un patrón distintivo de respuestas (a-e). En este apartado, revisaremos los argumentos y evidencia que se ofrece para sustentar tal afirmación; así como algunas objeciones que se le han planteado.

(a) *Expresiones faciales*. Griffiths (1997, p. 79) considera que las investigaciones y experimentos de Ekman y sus colegas (Ekman, 1971; 1972; 1992; 1994; Ekman y Friesen, 1971; 1986; Ekman, Sorensen y Friesen, 1969), logran establecer la tesis de que cada Programa Afectivo involucra un patrón específico, complejo, coordinado y automático de expresiones faciales. James Russell (1994; 2006), por su parte, ha objetado (a) de diferentes e interesantes formas. Algunas de sus objeciones cuestionan aspectos metodológicos en los estudios de Ekman, otras se fundan en formas alternativas de interpretar los resultados obtenidos. Enseguida aparece una relación con las más representativas de dichas objeciones, junto con las acrimoniosas réplicas que, principalmente Ekman, ha desarrollado.

1. *Los resultados no muestran universalidad*. Las correlaciones entre patrones de cambios faciales y los programas afectivos, estadísticamente, no son lo suficientemente interesantes como para sostener su universalidad.

⁹ Griffiths considera la posibilidad de que aspectos fenomenológicos y procesos cognitivos se pudieran sumar a (a-e), pero no proporciona motivos adicionales para hacerlo o no hacerlo (*ibid.*).

Respecto de esta crítica, Russell trabaja con una noción de ‘universalidad’ que Ekman considera un ‘hombre de paja’. La universalidad que Russell ataca sería una en la que, para todos los individuos en todas las culturas, se presentan *exactamente* los mismos cambios faciales con las mismas emociones y, por otro lado, los observadores le dan *siempre* el mismo significado emocional a dichos cambios faciales. Así construida la noción de ‘universalidad’, cualquier anomalía entre cambios faciales y estados emocionales, o cualquier disminución en el acuerdo al interpretar dichos cambios faciales, cuenta como evidencia en contra de la supuesta universalidad de patrones de cambios faciales asociados con emociones específicas. Pero Ekman recuerda que no es necesario, ni él ha invocado jamás —de aquí la alusión al ‘hombre de paja’—, una concepción tan fuerte de universalidad para el caso de las emociones y los patrones de cambios faciales característicos. Reconoce que, debido a factores culturales, pueden presentarse divergencias en las situaciones que dan lugar a una emoción, en las reglas de despliegue y en las formas en las que se presentan los cambios faciales, y que pueden existir diferencias en cuanto a la interpretación de tales cambios de una cultura a otra y de un individuo a otro. Lo interesante, piensa el mismo autor, es que con toda la distorsión y ruido de fondo cultural la evidencia revela regularidades lo suficientemente notables (*op. cit.* pp. 271-2) como para sustentar la tesis (a) y su universalidad. Vale la pena subrayar cómo la respuesta de Ekman se alinea con la noción de clases naturales *RPH*. Ekman insiste en una correlación causal entre estados emocionales y patrones de cambios faciales, pero también reconoce que es posible la *imperfcción nomológica*: es decir, el vínculo causal entre un estado emocional y el patrón de cambios faciales puede no desplegarse del todo o ser interrumpido. En la misma dirección, también parece conceder que cada patrón de cambios faciales complejos está constituido por una multitud de movimientos en la musculatura facial que configuran, más bien, un *racimo indeterminado* (sin límites precisos) de movimientos (propiedades) que, en cada instancia particular, puede variar en su número y grado de activación.

2. *Los formatos de elección forzada amplifican el acuerdo.* Elegir dentro de un grupo limitado de términos emocionales para etiquetar un patrón de cambios faciales (por lo regular exagerado y alimentado por un contexto), favorece indiscriminadamente el acuerdo. Esto tiene como consecuencia que las correlaciones entre programas afectivos y patrones de cambios faciales parezcan, en algunos casos, estadísticamente interesantes.

Esta objeción suscita una cuestión ulterior. La objeción supone y acepta de algún modo la tesis (a): que existen correlaciones interesantes entre los patrones de cambios faciales y emociones específicas. En este sentido, ¿no es en sí mismo interesante que los sujetos, en todos los estudios y de todas las culturas incluidas, hayan estado de acuerdo (de forma estadísticamente interesante) acerca de cuál término elegir para categorizar una configuración de movimientos faciales? Ekman considera que esto, por sí mismo, aporta evidencia importante para la tesis (a). Pero, ¿piensa Russell que estas regularidades —que de hecho acepta y reconoce— se deben obviar o desacreditar? Russell considera que los resultados son engañosos y que las tendencias que apuntan a un acuerdo disminuyen si se realizan cambios en el formato de elección forzada. Por ejemplo, en el caso del patrón de cambios faciales que corresponde al enojo, si se elimina de la lista —a partir de la cual los sujetos categorizan las expresiones faciales— el término ‘enojo’, los sujetos escogerían otro término para etiquetar el patrón y, quizás, tendría lugar, o no, otra forma de correlación interesante entre el patrón de cambios que se relaciona con el enojo y una emoción distinta. Esto, piensa Russell, establecería que el formato de elección forzada vicia el estudio y deja sin mucho sustento la tesis (a). Ekman responde de la siguiente forma. Supongamos un experimento en el cual se presenta a un grupo de sujetos una muestra de color amarillo y una lista de ‘términos’ de colores para categorizarla en la que aparecen los términos ‘verde’, ‘rojo’, ‘azul’, ‘blanco’, ‘púrpura’ y ‘naranja’... pero no ‘amarillo’. En ausencia del término correcto, los sujetos tienden a elegir el término de color ‘naranja’. Pero esto no prueba que la muestra no sea amarilla o que no haya una correlación fiable entre la muestra y aquello a lo que llamamos ‘amarillo’. Todo lo que prueba es que los sujetos elegirán la segunda mejor opción en ausencia de la correcta. Ahora bien, en el caso de las emociones sucede algo parecido. Ekman las llama “confusiones comunes”, y tienen lugar cuando las fallas de los sujetos al categorizar una expresión emocional presentan cierta regularidad. Así, las personas a las que se les restringe el término ‘enojo’ para categorizar la expresión facial correspondiente, tienden a etiquetarla con la palabra ‘asco’ —este es un hecho que Russell reconoce. Claro que esto, como en el caso de la muestra de amarillo y el término ‘naranja’, no muestra que la expresión facial corresponda al asco o se correlacione adecuadamente con el término ‘asco’. Ekman (1972), ha mostrado que los sujetos a los que se les restringe el término ‘enojo’, tienden a categorizar como enojo a la expresión facial correspondiente una

vez que se les restituye el término; restableciendo así la correlación entre patrones de cambios faciales y emociones que postula la tesis (a). Incluso hay una explicación que emana de los estudios que fundamentan la tesis (a), para dar cuenta de la regularidad que se observa en las llamadas “confusiones comunes”. Ekman apunta que cada patrón de cambios faciales, cuando no es interpretado de la forma habitual, tiende a ser correlacionado con otra emoción o término emocional específico (“confusión común”) en virtud de similitudes anatómico-topográficas y de configuración: es decir, la emoción con la que se confunde la expresión facial, cuando se presenta, da lugar patrones de cambios faciales similares en términos anatómicos y de configuración. Así, los sujetos tienden a confundir la expresión facial característica del enojo con la de asco, porque en ambos casos se activan aproximadamente los mismos grupos musculares y, aproximadamente, de la misma forma (Ekman, 1992).

Russell, por su parte, interpone otra objeción al formato de elección forzada. Se puede nutrir la lista de términos a elegir, de suerte que se incluyan algunos que correspondan a emociones emparentadas pero distintas: por ejemplo, además de ‘enojo’, se podrían incluir otros como ‘frustración’, ‘rabia’, ‘irritación’, ‘disgusto’, etc. En este caso, el grado de acuerdo disminuye radicalmente, puesto que la correlación entre el patrón de cambios faciales y el enojo se verá seriamente afectada por la presencia de los demás términos. Los sujetos, que antes sólo contaban con la opción de ‘enojo’, reducirán el acuerdo toda vez que correlacionarán la expresión facial que antes sólo podían categorizar como enojo con los otros términos incluidos. Pero esto, en realidad, no resta apoyo a la tesis (a). Griffiths (1997, p. 78) cita una objeción similar avanzada por Ratner (1989), quien muestra sorpresa al notar que los defensores de los programas afectivos y la tesis (a) sostienen que los sujetos de un experimento califican como ‘alegría’ la expresión facial de un individuo en una fotografía, *no obstante que los sujetos, de hecho, describieron la imagen con palabras tan distintas como ‘diversión’, ‘gratitud’, ‘optimismo’ y ‘satisfacción’*. Pero, el punto es que los sujetos, en las diferentes culturas, tienden a relacionar el patrón de cambios faciales —más que con palabras específicas— con estados psicológicos o emociones positivas. Los experimentos están diseñados para mostrar que todos los observadores son sensibles a ciertos patrones de expresión facial y que son interpretados, aproximadamente, de la misma manera. Por esta razón se considera que existe acuerdo transcultural en las categorizaciones y

descripciones de las expresiones faciales, no obstante las diferencias terminológicas. De nuevo, la tesis (a) parece salir bien librada de los ataques de Russell.

3. *Los formatos de elección libre muestran resultados muy distintos.* Cuando el estudio permite que el sujeto elija de una lista más nutrida, o libremente, un término para etiquetar un patrón de cambios faciales, las correlaciones y acuerdo entre éstos decaen sensiblemente, al punto de dejarnos con relaciones estadísticas muy poco interesantes.

Para Ekman, esta objeción supone que los términos o descripciones que eligen libremente los sujetos remiten a distintos tipos de fenómeno. Pero esto, piensa Ekman, no necesariamente es el caso. En el contexto del formato de elección libre, los sujetos podrían responder con términos que remiten a emociones o estados de ánimo (enojo o irritabilidad), o con descripciones de lo que siente el sujeto o de lo que pasa en su cuerpo (que la sangre le hierve), de la situación que causa la expresión (lo han insultado), o de lo que debe estar pensando (que quiere golpear a alguien). La cuestión es si la variedad de posibles respuestas remiten, en efecto, a diferentes fenómenos. De ser así —y así es como piensa Russell—, tal diversidad, si tiene lugar, representaría una seria disminución en el acuerdo que presume la tesis (a), entre patrones de expresiones faciales y emociones específicas. En esta dirección, Russell menciona un experimento con formato de elección libre en el que se muestra a un grupo una fotografía con la expresión de enojo y se pregunta a los sujetos lo siguiente: ¿Qué estado emocional o anímico está sintiendo la persona de la fotografía? Los resultados indican que el 30.6 % de los sujetos utilizaron la palabra ‘frustración’ para describir la expresión; mientras que el 25.6 % utilizó la palabra ‘enojo’. Ya que Russell considera que ‘frustración’ y ‘enojo’ remiten a fenómenos distintos,¹⁰ concluye que los formatos de elección libre prueban su punto: esto es, disminuyen el acuerdo entre expresiones faciales y emociones y, por lo tanto, restan sustento a la tesis (a). Ekman opone a la de Russell, una lectura según la cual, las emociones son fenómenos sumamente complejos que apenas comenzamos a entender, los cuales —por lo que sabemos— pueden incluir todo aquello a lo cual remiten las variadas descripciones que los sujetos pueden ofrecer en un formato de elección libre. Así, los distintos términos emocionales y descripciones (de lo que siente un sujeto, de lo que le ha ocurrido y/o lo que piensa) igual pueden apuntar al mismo fenómeno.

¹⁰ Russell considera que la frustración remite, más que a un estado emocional, a una *situación* (Russell, 1994).

La idea es que, distintos términos o descripciones pueden, en principio, referir a lo mismo. Básicamente, pueden entenderse como ‘sentidos fregeanos’ que nos conducen al mismo fenómeno por distintas rutas. Si este es el caso, entonces la diversidad de términos o descripciones que se pueden encontrar en experimentos con formatos de elección libre, no necesariamente dan al traste con el acuerdo entre expresiones faciales y emociones que presumen los estudios que no utilizan este formato. Según Ekman, los resultados que utiliza Russell para hablar de reducción de acuerdo, de hecho lo evidencian, toda vez que describir la expresión facial de enojo como frustración —incluso si ésta se entiende como remitiendo a una situación, puede considerarse como la situación típica que desencadena el enojo— es sólo otra forma de referirse parcialmente al mismo fenómeno: un estado psicológico complejo que, entre otras cosas, puede incluir patrones de expresiones faciales, una fenomenología, una fisiología, situaciones activadoras y procesos cognitivos característicos. La evidencia que en las últimas décadas se ha colectado sobre las emociones, sugiere fuertemente que la definición anterior es aproximadamente apropiada para el fenómeno que se estudia. Todo esto significaría que el acuerdo que se alcanzó en el caso que presentó Russell es en realidad del 70.9 % —si sumamos también los porcentajes que obtuvieron los términos ‘irritado’, ‘enfurecido’, ‘molesto’ (Ekman, 1994, p. 275). En conclusión, los términos y descripciones que se utilizan para caracterizar las expresiones faciales en los estudios que utilizan el formato de elección libre, bajo una interpretación más amplia, siguen abonando a favor de la tesis (a).

Amerita agregar un par de corolarios sobre este punto. En primer lugar, que Rosenberg y Ekman (1993) aportan evidencia empírica adicional a favor de la tesis (a) que se basa en formatos de elección libre. En segundo lugar, que la forma en la que se aproxima Ekman al estudio de las emociones es consistente con la teoría de la referencia que se desarrolló en la parte I de este trabajo (I, v.II) basada en la propuesta de Boyd (1979/1993). Según la teoría de la referencia que propone Boyd, los términos de clase natural (científicamente interesantes) presentan rasgos como referencia parcial y acceso epistémico (entre otros). En cuanto al primero, significa que los términos de clase natural, por ejemplo, pueden en sus primeros usos (en ausencia de una ciencia plenamente desarrollada y establecida que estudie sus referentes) denotar sólo parcialmente al fenómeno al que refieren; es decir, denotar sólo unas pocas de las propiedades que constituyen la estructura explicativa de di-

cho fenómeno. La referencia parcial también implica que distintos términos pueden hacer referencia a distintos aspectos o propiedades de un mismo fenómeno. El segundo rasgo significa que un término de clase natural inaugura una empresa epistémica que se proyecta indefinidamente en el tiempo y que tiene como finalidad el estudio y descubrimiento empírico de las propiedades físicas relevantes que constituyen la esencia de la clase (de entidad, organismo o fenómeno) que designa el término; donde la finalidad de dicha empresa epistémica consiste capturar las leyes y regularidades que gobiernan a la clase. Los argumentos de Ekman para hacer frente a la objeción de Russell basada en los formatos de elección libre, suponen que los términos emocionales refieren a un conjunto complejo de propiedades que estamos aún en camino de comprender (acceso epistémico), y que los términos y descripciones que usamos para designarlas, refieren a distintos aspectos de las mismas (referencia parcial).

4. Interpretaciones alternativas de la evidencia. Aun si se acepta que existen patrones de cambios faciales universales, eso no significa que estén relacionados con emociones particulares. Los patrones de cambios faciales universales no necesariamente tienen que implicar o estar asociados con una emoción. Es probable que dichos patrones hayan sido evolutivamente seleccionados, exclusivamente, para favorecer la interacción social en ciertas circunstancias, no como componentes o acompañantes de un episodio emocional. Quizás los patrones faciales son sólo adaptaciones que promueven respuestas y conductas en el observador que favorecen a quien los despliega.

Esta objeción de Russell es interesante en la medida en que permanece infradescrita. Cuando atendemos a sus implicaciones, su plausibilidad se tambalea. Hume enseñó a los filósofos que se pueden disociar dos ideas, no importa qué tan unidas aparezcan en la experiencia y en nuestra mente. La objeción de Russell persevera en esta ruta al sugerir que los patrones de cambios faciales no necesariamente deben vincularse con las emociones. La conjetura no es insensata o incoherente, debe reconocerse, pero implica que los patrones de cambios faciales deben disociarse, a su vez, de la mayoría de los procesos que identifican los expertos de diversas ramas como constitutivos de las emociones. En esta dirección, la objeción e hipótesis alternativa de Russell significa que los patrones de cambios faciales no están causalmente ligados (como es propuesto regularmente por los teóricos de las emocio-

nes) a ciertos cambios fisiológicos, estados fenomenológicos, procesos cognitivos y respuestas típicas por parte de quien experimenta una emoción —o, si se quiere, (*b-e*). No es claro siquiera que, dada la hipótesis de Russell, los patrones de cambios faciales estén relacionados con ciertos estímulos o situaciones activadoras de emociones. En términos de la hipótesis alternativa de Russell, para que un observador se comporte de una manera que favorezca a alguien que exhibe cierto patrón de cambios faciales, no es ni siquiera necesario que infiera la situación antecedente de la persona que exhibe dicho patrón. Como quiera que sea, la objeción de Russell implica la empíricamente dudosa afirmación de que los patrones de cambios faciales no están vinculados con ninguno de los procesos arriba mencionados. No puede hacerlo porque ello implicaría que están vinculados con algún o algunos aspectos que se identifican como procesos constitutivos de las emociones. Independientemente de lo que Russell piense que es o no es una emoción, la teoría de los Programas Afectivos puede construir un caso sólido en la medida en que la evidencia empírica recabada en la actualidad sugiere correlaciones causales interesantes entre patrones de cambios faciales y otros procesos (*b-e*) constitutivos de lo que la propia teoría llama ‘emociones’ (Matsumoto, *et al.*, 2008). En resumen, aunque la objeción e hipótesis de Russell no sean incoherentes, no cuentan con credenciales suficientes como para competir o hacer vacilar la teoría de los Programas Afectivos.

Finalmente, se puede citar el trabajo de Matsumoto, Keltner, Shiota, O’Sullivan y Frank (2008), quienes ofrecen un compendioso pero iluminador recuento del trabajo desarrollado y la evidencia obtenida en torno a la tesis (a). Señalan que, en las últimas tres décadas, se han publicado al menos 25 estudios que muestran que, para las seis emociones básicas y una más (*desprecio/contempt*), se identificaron patrones reconocibles de expresión facial. En todos los casos se utilizó el Sistema de Codificación de Acción Facial (FACS, por sus siglas en inglés) para obtener y validar los resultados. Además, el rango de culturas que participaron en dichos estudios fue extensivo: incluyendo a individuos de los cinco continentes (Matsumoto, *et al.*; *ibid.*, pp. 214-17; Tabla 13.2). Las evidencias que enumeran Matsumoto y sus colegas en torno a la universalidad y especificidad de las expresiones faciales de las emociones básicas, no sólo abarca estudios con individuos humanos. También incluyen algunas investigaciones con primates no-humanos que sugieren similitudes morfológicas entre las expresiones emocionales humanas y algunos comportamientos

expresivos en los primeros. Desde una perspectiva evolutiva, es probable que las configuraciones faciales en primates no humanos sirvan a las mismas funciones adaptativo/expresivas para las que sirven sus homólogas humanas. Dicha homología se debe entender también en términos anatómicos: muchos de los músculos faciales que intervienen en las expresiones emocionales humanas, tienen homólogos en los primates. Principalmente en el chimpancé (Waller *et al.*, 2006).

(*d y e*) *Respuestas endócrinas y del Sistema Nervioso Autónomo (SNA)*. También existe evidencia para establecer que cada Programa Afectivo instancia respuestas específicas endócrinas (*d*) y del sistema nervioso (*e*). En apoyo de esta afirmación, Griffiths cita los experimentos de Ax (1953) y Joseph Schachter (1957), los cuales revelan que el miedo produce algunos síntomas cardiovasculares específicos asociados con el aumento en la producción de adrenalina (aumento del ritmo cardíaco y de la presión sistólica). Por su parte, el enojo da lugar a síntomas cardiovasculares asociados con la noradrenalina, y presenta similitudes con la respuesta fisiológica del dolor. Estos hallazgos son consistentes con los resultados obtenidos por Schwartz, Weinberger y Jefferson Singer (1981), quienes reportaron que el enojo (*anger*) difiere de la felicidad (*happiness*), la tristeza (*sadness*) y el miedo (*fear*) en términos de presión diastólica y variación del ritmo cardíaco. Ekman, Levenson y Friesen (1983) refuerzan la idea de que cada emoción básica o Programa Afectivo involucra un patrón específico de (*d*) cambios endócrinos y (*e*) respuestas del sistema nervioso. Al medir los niveles de excitación de diferentes indicadores del Sistema Nervioso Autónomo, SNA (ritmo cardíaco, resistencia eléctrica y temperatura de la piel, tensión muscular, etc.), en situaciones emocionalmente cargadas, encontraron que existen patrones de cambios (*d-e*) específicos para cada emoción básica. Indican que no sólo fue posible distinguir entre emociones positivas (alegría) y negativas (miedo, enojo y (*grief*) pesar), también lograron distinguir patrones distinguibles de cambios fisiológicos entre las emociones negativas (Griffiths, 1997, pp. 79-83). Una actualización de la literatura sobre este particular debe incluir los estudios de Levenson, Ekman, Heider y Friesen (1992). En su investigación, Levenson y compañía solicitan a individuos de la tribu Minangkabú (Sumatra) que muevan sus músculos faciales para configurar las expresiones prototípicas de las 6 emociones básicas. Al hacerlo, se midieron cambios en el SNA identificables y distintos para cada emoción; los cuales, además, se correspondían con los cambios en el SNA registrados en suje-

tos norteamericanos, con quienes se llevó a cabo el mismo procedimiento. Matsumoto *et al.* (2008, p. 220, Tabla 3.14) también da cuenta de numerosos estudios sobre la correlación entre (d) cambios endócrinos y (e) fisiológicos específicos en instancias de emociones básicas. Con este trabajo, Matsumoto llena una laguna metodológica que Russell (2006) había denunciado con anterioridad: la necesidad de realizar un meta-análisis de los estudios que muestre que, en efecto, en todos ellos se pueden identificar patrones de cambios en el SNA para cada emoción.

(b) *Cambios musculoesqueléticos*. Sobre este expediente, Griffiths reconoce que la evidencia está constituida, apenas, de un interesante anecdotario que correlaciona reacciones corporales como encorvarse (tristeza), mostrar el pecho (enojo), retroceder (miedo) u orientarse hacia lo que estimula la emoción (alegría), con los demás elementos que la constituyen (a, c-e). Es necesario realizar experimentos controlados para proporcionar más apoyo a esta parte de la teoría de los Programas Afectivos. Finalmente, en cuanto a (c) los cambios vocales, parece que se ha comenzado a consolidar un cuerpo de evidencia relevante. Scherer, Banse, Wallbott y Goldbeck (1991) encontraron que la gente es bastante capaz de distinguir entre distintas emociones básicas a partir de ciertas pistas (no semánticas) vocales.¹¹ El análisis acústico de las vocalizaciones de las distintas emociones reveló que existen patrones de rasgos vocales distinguibles para cada emoción con valores específicos para distintos aspectos: volumen, tono, ritmo, fluidez e inflexiones. No obstante los resultados, Griffiths apunta a una seria deficiencia en los experimentos de Scherer, *et al.*: se llevaron a cabo exclusivamente con sujetos educados en la tradición occidental. En 2001, Scherer y sus colegas publicaron los resultados de una nueva serie de experimentos con aproximadamente los mismos resultados, sólo que en esta ocasión los sujetos pertenecían a distintos países y culturas (norteamericanos, europeos y asiáticos). Así, parece articularse un panorama argumental y evidencial sólido para el caso de las emociones básicas o Programas Afectivos.

¹¹ En un 62.8 %, los sujetos del experimento acertaron a la emoción que se pretendía representar (miedo, enojo, alegría, tristeza) por medio de actores. El asco/repugnancia fue identificado en un número bastante inferior en proporción a al resto de las emociones básicas (Scherer, *et al*, 1991, p. 123).

IV.III. ¿SON REALMENTE LAS EMOCIONES BÁSICAS ‘PROGRAMAS’? LA OBJECCIÓN DE LA AGREGACIÓN ARBITRARIA Y LA ALTERNATIVA COMPOSICIONALISTA

Ahora bien, ¿qué evidencia existe para pensar que las respuestas adaptativas (a-e) que caracterizan a cada emoción básica están correlacionadas causalmente, de suerte que constituyan, precisamente, un programa o circuito?¹² Después de todo, es posible interpretar (a-e) como sistemas independientes que responden a distintos rasgos del entorno; los cuales, fortuitamente, co-ocurren con frecuencia en forma de situaciones estereotípicas del medio. Esta sería la razón por la que pensamos que las respuestas a las que dan lugar configuran un patrón de cambios fisiológicos complejo, coordinado y automático (programa). Pero esto representaría un caso de *agregación arbitraria*. Es decir, un caso en el que se asume que un conjunto de propiedades están causalmente correlacionadas sin evidencia suficiente¹³. Los problemas que suscita un caso de *agregación arbitraria* son de índole ontológica y epistemológica. En términos ontológicos, nos puede conducir a asumir la existencia de entidades teóricas sin fundamentos sólidos. La consecuencia epistemológica consiste en viciar nuestras aproximaciones científicas a ciertos tipos de fenómenos, desorientando programas de investigación y empobreciendo la proyectabilidad de nuestras teorías sobre dichos fenómenos. Barrett (2006, p. 33) menciona una pléyade de experimentos y estudios que muestran que las correlaciones que se tienden a establecer entre diferentes aspectos relacionados con las emociones son más bien débiles; lo cual implica que las razones para pensar a las emociones básicas como *programas afectivos* que involucran repuestas adaptativas *coordinadas* y *automáticas* son menos sólidas de lo que se presume. Esto enciende la alerta para un caso de *agregación arbitraria*. Dispuestas así las nociones, Barrett, McNaughton y Russell proponen una interpretación *constructivista* o *composicional* de las emociones, según la cual no existirían realmente las emociones básicas, sino sólo un conjunto de sistemas fisiológicos/psicológicos más básicos (entre los que figurarían las respuestas (a-e)) independientes que respondan a distintos rasgos del entorno y a partir de los cuales se constituyen lo que llamamos emociones básicas.

¹² La cuestión es si, en efecto, (a-e) exhiben o no homeostasis causal.

¹³ Barrett (2006), McNaughton (1989) y Russell (2006) lanzan variables de esta objeción a la tesis de las emociones básicas.

IV.IV. RESPUESTAS A LA OBJECCIÓN DE LA FALTA DE CORRELACIÓN

En cuanto a la supuesta falta de correlación causal, se han ofrecido distintas explicaciones. Se habla, por un lado, de que existen reglas de despliegue (*display rules*) relacionadas con cuestiones culturales que impiden que los sujetos de un experimento expresen naturalmente sus respuestas emocionales (expresiones faciales) o las reporten sinceramente. Por esta razón, pueden estar ausentes en un episodio emocional (Ekman, 1973; 1973; Matsumoto, Kasri y Kooken, 1999). En segundo lugar, pueden existir deficiencias en el protocolo experimental que impiden detectar las correlaciones, en virtud de que no consideran que los sistemas de respuesta podrían funcionar en diferentes tiempos (Bradley y Lang, 2000). Mientras las mediciones no tomen en cuenta esta posibilidad pueden ser poco confiables. Finalmente, es posible que las condiciones de laboratorio no sean apropiadas para desencadenar las respuestas emocionales en toda su potencia y naturalidad (Tassinari y Cacioppo, 1992; Elster, 1999, cap. 2). Por otro lado, como bien reconoce Barrett (2006, p. 34), un bajo nivel de correlaciones causales (si fuera el caso), no desconfirmaría la tesis de que las emociones básicas no son clases naturales. La definición de una clase natural puede fundarse en la homología, y no solamente en analogía. Esto significa que, lo que distingue una clase natural es un origen o derivación común, no solamente un conjunto de propiedades en común. Si una clase natural es definida, en parte, por homología —en oposición a ser definido por analogía— entonces las instancias de dicha clase no tienen que compartir *exactamente* las mismas propiedades observables. La idea es que existe un mecanismo causal homeostático subyacente para dicha clase de fenómenos que se puede descubrir empíricamente, y la constitución de dicho mecanismo puede tener límites vagos o imprecisos. La individuación de las clases naturales por homología es coherente con la concepción de clase natural *RPH*, ya que acepta indeterminación e imperfección nomológica.

IV.V. RESPUESTA AL RETO COMPOSICIONALISTA: LA DIVERSIDAD DE LOS ESTÍMULOS

En cuanto al reto que plantea la alternativa *composicionalista* a la teoría de las emociones básicas, se pueden ensayar, por lo menos, dos formas de enfrentarlo. Griffiths señala que las situaciones o estímulos que provocan una emoción básica son tan diversos que es difícil sostener que comparten un conjunto de rasgos comunes a los cuales responden los distintos sistemas fisiológicos/psicológicos independientes. Pero esto es precisamente lo que requie-

re la alternativa *composicionalista* de los sistemas independientes; pues sólo de esta manera se explicaría la aparente coordinación de sistemas independientes en la respuesta emocional. La teoría de los Programas Afectivos explica la cuestión de la diversidad de situaciones o estímulos que activan una emoción básica de la siguiente forma. Lo único que tienen en común las situaciones que, por ejemplo, dan lugar a la emoción básica del *miedo*, es la propiedad sumamente abstracta de que son *peligrosas*. Así, la detección de *un rasgo* en el entorno (que exhiba *peligrosidad*), desencadena el conjunto correlacionado de respuestas adaptativas (a-e) o *programa afectivo*. Esta interpretación no requiere que la amplia gama de situaciones que dan lugar a las emociones básicas compartan *un conjunto de rasgos* óptimos para activar, respectivamente, cada uno de los distintos sistemas independientes. El modelo *composicionalista* requeriría una arquitectura más compleja, en la que cada sistema fisiológico/psicológico se activaría como resultado de la percepción de algún rasgo apropiado en el entorno, lo que, a su vez, exigiría que cada sistema contara con un subsistema de detección de rasgos apropiados en el entorno para su activación. Aunque la complejidad del modelo no constituye un defecto en sí mismo, sí lo es el hecho de que la evidencia no parece acumularse a su favor. Simplemente nunca se han logrado aislar los rasgos que, de forma individual, activarían independientemente cada uno de los sistemas fisiológicos/psicológicos (Barrett, 2006). Barrett, para hacer frente a esta objeción, sugiere que la dificultad tiene que ver con que es muy difícil, y quizás imposible en la práctica, aislar en una situación concreta controlada (experimentalmente) los rasgos que activan cada sistema fisiológico. Pero, si es así, entonces vuelven menos eficiente su propio modelo, ya que cabría preguntar qué sentido tendría postular subsistemas de detección (de dichos rasgos) independientes. ¿No sería más simple y evolutivamente más eficiente tener un solo detector de un tipo de rasgo? Después de todo, los demás vendrían prácticamente juntos. De nuevo, los hechos tal y como los describen algunos composicionalsitas, parecen favorecer la teoría de los Programas Afectivos.

Este escenario es el que resulta de analizar la forma en que responden, la teoría de los Programas Afectivos y la alternativa *composicionalista*, a la cuestión de la diversidad de las situaciones y estímulos que desencadenan las emociones básicas. Pero este argumento no logra descartar la teoría *composicionalista*. Por lo tanto, no logra establecer que en efecto exista una correlación causal entre las distintas repuestas fisiológicas (a-e) que definen a

las emociones básicas, de suerte que se pueda afirmar que constituyen programas. El argumento sólo parece mostrar, por un lado, la *improbabilidad* de la alternativa *composicionalista* (de los sistemas independientes) y, por otro, reconocer el atributo epistemológico de la *simplicidad* en la teoría de los Programas Afectivos en relación al caso de la diversidad de situaciones que desencadenan las emociones básicas¹⁴.

IV.VI. OTRA RESPUESTA: LA MODERADA EFICIENCIA Y ESCASA FLEXIBILIDAD DE LOS PROGRAMAS AFECTIVOS

No obstante, podríamos obtener más elementos para decidir el caso entre los Programas Afectivos y la alternativa *composicionalista* a partir de las siguientes consideraciones. En términos de diseño o arquitectura cognitiva el modelo *composicionalista* parece ser más *flexible* y quizás, por la misma razón, más *eficiente* que el modelo de los Programas Afectivos. Si cada uno de los sistemas independientes que constituyen una emoción reacciona ante un tipo peculiar de rasgo del entorno, entonces podemos contar con que, en cada experiencia emocional, se desplieguen los sistemas independientes pertinentes para la situación. Muy eficiente, ¿no? Pero ese es el problema, las repuestas emocionales no son así de *eficientes* y *flexibles*. Pueden inhibirse o presentarse respuestas (a-e) que resultan poco convenientes en ciertas situaciones y para las que no se puede identificar un rasgo específico del entorno que las active. Por ejemplo, una persona, a solas, puede exhibir expresiones faciales características ante estímulos emocionalmente cargados —junto con otras respuestas fisiológicas (b-e) (Ekman, 1972). En tales casos, ¿qué rasgo del entorno sería el responsable de dicha respuesta? Las instancias particulares de emociones básicas *suelen* involucrar el paquete completo de respuestas fisiológicas (a-e); y no existe evidencia sólida que pruebe que cada una de tales respuestas (a-e) dependa de la presencia de algún rasgo en el entorno. Para bien o para mal (en términos individuales), el combo (a-e) se presenta, por lo regular, *completo*; aunque a veces no sea lo más *apropiado* en términos evolutivos. Luego, las emociones básicas no parecen exhibir mucha *flexibilidad*, o ser del todo *eficientes*, del modo en que deberían serlo si estuvieran constituidas por sistemas independientes.

¹⁴ Ni siquiera puede decirse que la teoría de los Programas Afectivos expresa de forma más fidedigna el modo en que funcionan y tienen lugar las adaptaciones evolutivas. Griffiths (*Ibid.* p. 85) reconoce que McNaughton ofrece buenas razones evolutivas para sustentar la tesis constructivista o composicionalista de los sistemas independientes.

En resumen, la falta de simetría en casos concretos entre sistemas supuestamente independientes y rasgos del entorno, por un lado; y la co-ocurrencia frecuente de distintas respuestas fisiológicas (a-e) en situaciones emocionalmente cargadas —en las que es difícil detectar rasgos apropiados para cada respuesta—, por otro, parecen favorecer la tesis de que las emociones básicas son *programas coordinados y automáticos*, poco *flexibles* y, aunque *eficientes*, no tanto como lo parece implicar la alternativa *constructivista*. El cargo por *agregación arbitraria* en contra de la teoría de los Programas Afectivos debe esperar por mejores argumentos, teorías y evidencia.

IV.VII. LA MODULARIDAD DE LAS EMOCIONES BÁSICAS

La discusión precedente, servirá para introducir otra característica relevante que reconoce Griffiths en los Programas Afectivos: la *modularidad* (*ibid.* pp. 91-97). Se ha señalado que las respuestas adaptativas (a-e) parecen precipitarse de forma completa *coordinada y automática*. En cuanto a las causas, se reconoce que algunas veces pueden estar involucrados *procesos cognitivos superiores*¹⁵. Pero, en muchos casos, los procesos cognitivos superiores no parecen tener injerencia. De hecho, parece que *no pueden* intervenir en la activación o desactivación de los programas afectivos. Así, algunas cadenas de pensamiento podrían desembocar en conclusiones que justificarían la presencia de una emoción, sin que esta se presente, después de todo. En otras ocasiones, procesos similares implicarían que no habría razones para experimentar una emoción, no obstante, tiene lugar una. Que las emociones básicas exhiban esta relativa disociación respecto de *procesos cognitivos superiores* sugiere que constituyen un sistema cognitivo modular. Los sistemas cognitivos modulares poseen, entre otras características¹⁶, la *obligatoriedad* (mandatorinnes), *opacidad* o *acceso limitado a los procesos que efectúa el módulo*, y *encapsulamiento informacional*. Las emociones básicas muestran *obligatoriedad*, en el sentido de que no está bajo nuestro control el activarlas o no. La gente responde con enojo o miedo ante ciertas situaciones lo quiera o no. También son *opacas*, lo que significa que no tenemos acceso consciente a los procesos cognitivos que tienen lugar mientras experimentamos una emoción; sólo podemos llegar a ser conscientes (nos resultan accesibles) las respuestas conductuales. Las emociones bási-

¹⁵ Griffiths (*ibid.* p. 92) define los procesos cognitivos superiores como aquellos en los que la información utilizada involucra contenidos de tipo proposicional (creencias y deseos).

¹⁶ Fodor (1983) enumera un total de nueve rasgos distintivos de la modularidad.

cas también parecen exhibir *encapsulamiento*, lo que quiere decir que no pueden acceder a la información que se procesa en otros sistemas cognitivos (los cuales pueden contradecir la información del módulo). Podemos, por ejemplo, tener creencias conscientes sobre la inocuidad de una serpiente, pero esta información no es tomada en cuenta cuando un programa afectivo modular ha activado ya el conjunto de respuestas *automáticas y coordinadas*.

Aunque Griffiths sólo menciona tres características de la modularidad, sus observaciones ayudan a identificar otras dos características de la modularidad presentes en los programas afectivos. Los módulos emocionales, también incluyen un *sub-sistema de aprendizaje enrutado* (biased), a partir del cual se aprende *cuáles* estímulos nuevos pueden ser apropiados para activar un programa afectivo. La idea es que no cualquier cosa o estímulo se considerará apropiado para dispararlos. En este sentido, Ohman (1976) mostró que es muy complicado lograr que las personas sientan miedo por estímulos neutros como flores y formas geométricas. Y, aunque puede lograrse mediante condicionamiento por medio de descargas eléctricas, la asociación de un estímulo neutro con una emoción básica puede suprimirse fácilmente con el mismo método. Algo distinto ocurre con estímulos de distinto tipo, como serpientes y precipicios. En tales casos, el condicionamiento con descargas eléctricas fue prácticamente innecesario; y los intentos por remover las respuestas emocionales asociadas con estos estímulos a partir de descargas, tomaron más tiempo y exposición. Las emociones también poseen la característica modular del *procesamiento rápido*. Esto es, su activación y despliegue global se lleva a cabo en un lapso bastante breve. El procesamiento rápido representa una ventaja evolutiva. Una respuesta rápida, no mediada por ningún proceso cognitivo superior, puede ser vital para la subsistencia de los individuos de una especie. Así, saltar *ipso facto* al detectar un objeto reptante y sinuoso —sin realizar ningún tipo de juicio complejo basado en las características que se perciben—, además de representar un tipo de conducta con muchos méritos evolutivos, sugiere que algún tipo de procesamiento informacional ha tenido lugar a una velocidad muy alta.

IV.VIII. PROGRAMAS AFECTIVOS COMO CLASES NATURALES RPH: HOMEOSTASIS Y PROYECTABILIDAD

Los Programas Afectivos o Emociones Básicas parecen satisfacer las condiciones que se requieren para calificar como clases naturales, entendidas como clases *RPH*. En apartados

anteriores, se ha citado evidencia que apoya la tesis de que una emoción básica involucra un patrón de respuestas distintivo (a-e) (Ekman, *et al.*, 1983; Schwartz, *et al.*, 1981; Ax, 1953; y Schachter, 1957). Se han propuesto respuestas al cargo de *agregación arbitraria* en contra de la teoría de los Programas Afectivos, y al reto que representaba la alternativa *composicionalista*. También se han ofrecido razones para afirmar que las emociones básicas, en su conjunto, constituyen un *módulo* psicológico. Todos los argumentos y evidencia citada al abordar las problemáticas mencionadas, en conjunto, también abonan para la idea de que las emociones básicas exhiben *homeostasis causal*: es decir, que comparten un conjunto de propiedades que se hallan correlacionadas por medio de mecanismos causales homeostáticos (a-e). Tales propiedades y mecanismos homeostáticos, podría decirse, constituyen la *estructura explicativa* de las emociones básicas. Como tales, aportan una explicación de dicho fenómeno y lo hacen proyectible: lo que significa que garantizan la existencia de generalizaciones y predicciones sobre las emociones básicas. Hasta aquí, puede verse que los Programas Afectivos exhiben la *homeostasis causal* característica de las clases naturales *RPH* y, al hacerlo, satisfacen el requisito epistemológico de *proyectabilidad*, el cual es también esencial y definitivo, como se ha indicado antes (I, V, II), de las clases naturales.

IV.IX. INDETERMINACIÓN Y CARÁCTER DINÁMICO DE LOS PROGRAMAS AFECTIVOS

Más aún, la teoría de los Programas Afectivos, está de acuerdo con otras demandas metodológicas e hipótesis metafísicas que se discutieron en la primera parte de este trabajo. Por un lado, la esencia o estructura explicativa de las emociones básicas (a-e), está constituida por propiedades físicas relevantes que sólo se pueden descubrir empíricamente. Por otro lado, la naturaleza de la estructura explicativa de los programas afectivos exhibe la *indeterminación* y carácter *dinámico* característicos de la concepción de clases naturales como clases *RPH*. Como lo evidencian Ekman (1972), e incluso los estudios que cita Barrett (2006, p. 33) para mostrar la falta de correlaciones constantes en las respuestas emocionales, es posible que instancias particulares de emociones básicas no desplieguen todas las propiedades o mecanismos homeostáticos característicos (a-e)¹⁷. Pero esto es un rasgo distintivo de las clases *RPH*. Las respuestas (a-e) que constituyen a las emociones básicas no se entienden como condiciones necesarias y suficientes. La imperfección o indeterminación

¹⁷ Ekman señala que las *reglas de despliegue* (*display rules*) pueden impedir que las expresiones faciales características se presenten de forma notoria en un episodio emocional.

nomológica es posible o real en la concepción de clases *RPH*. En resumen, se acepta que las esencias de las emociones básicas tienen un carácter *indeterminado*. Esta afirmación se entiende de dos formas relacionadas. En primer lugar, como significando que la *estructura explicativa* misma de las emociones básicas está constituida por un racimo *indeterminado* de propiedades homeostáticamente vinculado, y empíricamente identificado. Esta *indeterminación* se proyecta hacia las instancias particulares de emociones básicas: éstas pueden exhibir la mayoría o una parte de las propiedades vinculadas con el racimo de propiedades homeostáticas que empíricamente se ha logrado identificar.

Quizás valga la pena llamar la atención sobre el hecho de que nuestras intuiciones y prácticas cotidianas de identificación de experiencias emocionales auténticas y apócrifas, son consistentes con la imagen de las emociones como indeterminadas en su esencia. Estamos dispuestos a aceptar que otras personas y nosotros mismos podemos tener una experiencia emocional plena o auténtica aun cuando no se presenten algunas características que consideramos (intuitivamente) esenciales a una emoción: como (a) y (c). Nos hemos sentido alegres sin sonreír o gritar de gozo y, sin duda, otras personas han sentido miedo en ciertas ocasiones y no lo han evidenciado con un grito de ayuda o una configuración facial particular. También sabemos muy bien que la presencia de ciertas características esenciales de las emociones (a y c) no son suficientes para adscribir a alguien una emoción. ¿No es acaso un hecho que existen los actores, los sarcásticos y los mentirosos? Nuestras prácticas cotidianas para identificar emociones auténticas y apócrifas, suponen la *indeterminación* característica de las emociones. La ausencia de algunas propiedades o elementos que se suelen considerar como distintivos o representativos de una emoción no significa que ésta no tenga lugar; del mismo modo que su presencia no implica lo contrario.

En cuanto al carácter *dinámico* de las estructuras explicativas de las emociones básicas, está implicado en la caracterización de Griffiths. En tanto que respuestas adaptativas (complejas, coordinadas y automáticas) de un organismo ante situaciones relevantes para su subsistencia y descendencia, las emociones básicas se reconocen como manufacturadas evolutivamente *a través* del tiempo. Así, es plausible concebirlas como sujetas a *ajustes y cambios* en los últimos 10 000 años (*EEA: Environment of Evolutionary Adaptation*) que redituaron en los programas afectivos o emociones básicas que estudiamos hoy en día.

En esta dirección, Griffiths considera que las emociones básicas, en tanto que clases naturales, son más similares a las clases naturales de la biología que a las de la química o la física. Las últimas definen sus esencias en términos de su composición o microestructura *interna* y de *analogía*. Las esencias de las clases naturales biológicas no sólo toman en cuenta la composición interna, también apelan a factores *externos* como la descendencia y la *homología*. Una especie biológica es, en parte, individuada por el linaje del que procede, y este se especifica por su trayectoria dentro de la jerarquía taxonómica (género, familia, orden, clase, etc.); dicho linaje lo comparte con subespecies y especies de su mismo género. En este sentido, habrá mecanismos homeostáticos que compartirán, por ejemplo, las especies que pertenecen al mismo género, y en virtud de las cuales será posible articular generalizaciones e inducciones sobre dichas especies: sobra decir que la posibilidad de dicha proyectabilidad descansa en el hecho de que las especies dentro de un mismo género tienen un ancestro común del cual heredaron los mecanismos homeostáticos que comparten.¹⁸ Así, existe un factor histórico/*dinámico* crucial en la individuación de categorías (taxones) en biología y en la identificación de mecanismos homeostáticos. Las emociones básicas, en tanto que programas o respuestas adaptativas, pueden también ser individuados en términos histórico-evolutivos. Programas y respuestas adaptativas homólogas se pueden observar en distintas especies cercanas y no tan cercanas a los seres humanos en términos filogenéticos. Lo cual sugiere, precisamente, que son programas adaptativos *dinámicos*: es decir, que han evolucionado a través del tiempo en diferentes especies.

V. ARGUMENTOS Y OBJECIONES EN CONTRA DE UNA TEORÍA DE LA EMOCIÓN

V.I. EL ARGUMENTO DE LA DESUNIÓN DE LAS EMOCIONES

En el inicio de esta segunda parte se enunciaron tres tesis que representan posiciones distintas en torno a la cuestión del estatus de las emociones en tanto que clases naturales. Dichas tesis son:

i) Cada una de las emociones básicas es, individualmente, una clase natural.

¹⁸ Mientras más alto se encuentre una categoría en la jerarquía taxonómica, podremos hablar de mecanismos homeostáticos más específicos y de generalizaciones y predicciones más finas. En la medida en que descendemos en la jerarquía taxonómica la proyectabilidad es menos específica, pero no menos interesante. La replicación celular en un embrión humano ejemplifica un mecanismo homeostático común que no es en ningún sentido trivial.

ii) El conjunto de las emociones básicas constituye una clase natural.

iii) El conjunto de todo lo que llamamos emociones (lo cual incluye a las emociones básicas y a las que no lo son) constituye una clase natural.

Desde entonces, rehusamos profundizar en torno a (i), por no exhibir un compromiso ontológico lo suficientemente interesante en relación a las emociones. En lo que a (ii) respecta, el capítulo precedente intentó evidenciar que se dispone de evidencia y una interpretación de la misma que ayudan a establecer dicha tesis. En este sentido, se ha alegado a favor de la tesis de que, en efecto, existen seis emociones básicas o programas afectivos, los cuales son respuestas complejas, coordinadas y automáticas, evolutivamente diseñadas. Que sean respuestas coordinadas, automáticas y, en muchos casos, insensibles a procesos cognitivos superiores, sugiere que constituyen *un sistema modular psicológico*. Estas emociones básicas pueden considerarse clases naturales del tipo *Racimo de Propiedades Homeostáticas (RPH)*, toda vez que las propiedades (a-e)¹⁹ que las caracterizan constituyen un racimo de propiedades físicas que exhiben homeostasis causal: están causalmente correlacionadas. La identificación de dicho racimo de propiedades homeostáticas (a-e) compartidas por las emociones básicas, favorece la promulgación de generalizaciones, hipótesis e inducciones en torno a una parcela del fenómeno que llamamos ‘emoción’. Además, se ha argumentado que las emociones básicas —en tanto que clases naturales *RPH*— tienen esencias *dinámicas e indeterminadas*. Es decir, por un lado, dichas esencias han evolucionado con el tiempo, y la historia de dicha evolución es relevante para su individuación (homología). En otras palabras, las emociones básicas son el resultado de la evolución de rasgos o mecanismos adaptativos presentes en nuestros ancestros más remotos y que heredamos y compartimos con otras especies. Por otro lado, las propiedades homeostáticas que constituyen la esencia de las emociones básicas configuran un racimo indeterminado de propiedades que, en instancias particulares de emociones básicas, puede ser instanciado sólo parcialmente. Es una cuestión empírica decidir si un caso particular constituye una instancia de la clase natural o emoción básica.

¹⁹ (a) expresiones faciales; (b) cambios musculoesqueléticos; (c) cambios vocales; (d) respuestas hormonales o endócrinas; y (e) respuestas del sistema nervioso (*supra*, II, IV.1).

Pero la defensa exitosa de (ii), tiene repercusiones negativas para la más general tesis (iii) o, por lo menos, así lo cree Paul Griffiths (1997). Además de las emociones básicas (EB), existe lo que Griffiths denomina “Emociones Cognitivas Superiores” (ECS). Ejemplos de ECS son la vergüenza, el orgullo, los celos, la culpa, el amor, etc. Estos fenómenos, según el autor, parecen implicar mecanismos causales muy distintos a los que se ven envueltos en las emociones básicas. Griffiths persevera en la identificación de las diferencias que separan a las EB de las ECS, con la finalidad de construir un caso en contra de la posibilidad de una teoría general de las emociones o, en otras palabras, de la tesis (iii). En este sentido, enumera cinco aspectos cruciales en los que son distintas entre sí dichas emociones:

1. Las ECS involucran capacidades conceptuales y cognitivas bastante complejas, sin equivalentes en las emociones básicas. Por ejemplo, tener un concepto de la imagen pública de uno mismo parece ser un requisito conceptual para experimentar vergüenza (Solomon, 2007, p. 135).
2. Relacionado con lo anterior, está el hecho de que, mientras las EB están presentes desde el primer año de vida, las ECS aparecen entre los 18 meses y 4 años de edad, y sólo cuando ciertas capacidades cognitivas se han desarrollado (la capacidad de reconocer e interiorizar normas, la de atribuir pensamientos y sentimientos a los demás, etc.).
3. No hay evidencia de que las ECS involucren patrones de cambios fisiológicos distinguibles. Las ECS no parecen tener expresiones faciales características e, incluso, algunas instancias no involucran siquiera trastornos fisiológicos.
4. Las ECS no exhiben los rasgos característicos de la *modularidad* identificados en las emociones básicas: obligatoriedad, acceso limitado a los procesos que efectúa el módulo, encapsulamiento informacional, sub-sistema de aprendizaje enrutado y procesamiento rápido (*supra*, II, IV.VII).
5. Griffiths, siguiendo con cautela a Robert Frank (1988), reconoce que tanto las emociones básicas como las ECS pueden ser respuestas adaptativas y, además, que tienen en común el ser *patrones de motivación irruptoras*, lo que significa (brevemente) que ambas son mecanismos por medio de los cuales se puede interrumpir una tarea y reorientar la conducta co-

mo resultado de alguna propiedad (exigencia) detectada en el entorno (Griffiths, *op. cit.* pp. 115-122). No obstante, insiste en que hay diferencias cruciales entre las emociones básicas y las ECS. La función evolutiva de las últimas es bastante específica: *resolver problemas de compromiso*. Por ejemplo, ser capaz de sentir culpa, vergüenza y amor, garantizan que puedes asumir un compromiso y que te será complicado no honrarlo. Las emociones cognitivas superiores tienen una función colectiva en tanto que existen para favorecer vínculos y compromisos entre individuos. Así, mientras que las emociones básicas son programas rápidos, modulares y con contrapartidas *homólogas* en otras especies, las ECS son programas que interactúan con creencias, tienen una función social y no parecen encontrarse en otras especies.

Todas estas diferencias llevan a Griffiths a formular la *tesis de la desunión de las emociones*, según la cual las EB y las ECS son fenómenos que involucran mecanismos causales e historias evolutivas distintas, en virtud de lo cual constituyen dominios de investigación científica separados. En conjunto, no conforman una clase natural —ni siquiera en el sentido más liberal de *RPH*—, lo cual contradice explícitamente la tesis (iii). No es posible una teoría general de la emoción. En este punto, Griffiths (1997, pp. 14-17) sugiere *eliminar* el término genérico de *emoción* —toda vez que designa a un grupo heterogéneo de fenómenos psicológicos— o conservarlo sólo para el caso de las emociones básicas (*op. cit.* p. 230)²⁰.

V.II. MÁS OBJECIONES A LA POSIBILIDAD DE UNA TEORÍA GENERAL DE LAS EMOCIONES

Otros autores, con fundamentos distintos, han avanzado objeciones en contra de la tesis (iii), que sostiene que las emociones conforman una clase natural. Algunas son de carácter estrictamente *a priori* o conceptual, es decir, parten de lo que se entiende habitualmente por ‘emoción’ y de cierta concepción de ‘clase natural’, para negar que las emociones sean clases naturales. Otras, del mismo modo que las avanzadas por Griffiths, resultan de la asunción de ciertas concepciones de ‘emoción’ y ‘clase natural’, más cierta interpretación de los hechos y evidencia empírica. Enseguida se ofrece una breve relación de dichas objeciones:

²⁰ Nuestra aproximación promete ser menos pesimista que la de Griffiths. Más aún, nos valdremos parcialmente de la evidencia y argumentos que ofrece para mostrar que las emociones básicas son clases naturales para articular nuestro caso a favor de una teoría general de las emociones.

6. Le Doux (1996, p. 16), a partir de su lectura de la evidencia neurofisiológica disponible, señala que ‘no existe tal cosa como la facultad de la emoción’. No existe ‘porque no hay una estructura o sistema cerebral dedicado específicamente a esta función fantasma’.

7. Elster (1999), por su parte, aunque no niega categóricamente la verdad de (iii), se muestra escéptico respecto de la posibilidad de defenderla. Luego de analizar los métodos empíricos, evidencia e hipótesis disponibles, resuelve que no alcanzan para establecer o refutar (iii).

8. Solomon (1995, pp. 174-175), basándose en estudios antropológicos e interculturales, observa que el término ‘emoción’ está ausente en el léxico de muchas culturas y que, incluso en el idioma inglés, tiene límites bastante brumosos, de suerte que nunca es fácil decidir qué cuenta como una emoción. Así las cosas, la falta de universalidad del término ‘emoción’, y su vaguedad referencial, piensa Solomon, hacen poco plausible la posibilidad de que todo aquello a lo que etiquetamos como ‘emoción’ conforme una clase natural.

9. Amelie Rorty (1978, p. 141) observa que el conjunto de fenómenos que llamamos ‘emoción’ es demasiado heterogéneo, lo cual hace imposible identificar un grupo de propiedades distintivas (esencia) compartidas por todos. No existen fundamentos empíricos para considerarlas una clase natural.

10. Ben-Ze’ev (2000, p. 6) considera que las emociones no son (ni en conjunto, ni por separado) una clase, sino *prototipos*. Los prototipos, a diferencia de las clases, no se definen por conjuntos delimitados de propiedades compartidas. Los prototipos se definen en términos de grados de semejanza. Por ejemplo, se toma un caso *prototípico* de emoción y sólo consideraremos emoción a una instancia que se asemeje lo *suficiente*²¹ al prototipo. Esto vale, piensa Ben Ze’ev, tanto para las emociones en general como para los distintos tipos de emociones por separado. Ya que los prototipos no tienen límites claros en sus propiedades, no pueden ser clases naturales.

²¹ Ben Ze’ev no profundiza mucho sobre el grado de semejanza, en relación al prototipo, que es suficiente para contar como una emoción, en general o de un tipo específico. Sólo menciona ‘que es una cuestión criterial regida por normas contextuales diversas, que van de lo cultural, social y político, a lo académico’ (*ibid.*)

11. De Sousa (1987, p. 20) sostiene que las emociones no pueden constituir una clase natural porque no tienen un objeto formal al cual todas se dirijan. Las creencias son una clase natural porque tienen un objeto formal: a saber, la verdad.

Estas 11 objeciones y observaciones —en contra de (iii)—, compendiadas en el presente apartado y el anterior, pretenden estar ordenadas de forma decreciente en relación a su dependencia respecto de la evidencia empírica, de suerte que, las que figuran al final, podrían clasificarse como esencialmente conceptuales. Ahora bien, ya que el objetivo fundamental de esta investigación es profundizar en torno a la plausibilidad de la tesis (iii), que sostiene que las emociones, en su conjunto, conforman una clase natural y, por lo tanto, es posible una teoría general de las emociones, es menester responder a las 11 objeciones/observaciones. El desarrollo de nuestra investigación, a estas alturas, no ha generado los recursos suficientes para hacer frente a todas las objeciones: principalmente a las primeras siete, las cuales son, predominantemente, de índole empírica. En los capítulos subsiguientes intentaremos proveernos de los recursos conceptuales, teóricos y empíricos necesarios para articular respuestas satisfactorias a estas observaciones y objeciones. Por ahora, a partir de la concepción de clase natural que se discutió y defendió en la primera parte de este trabajo (clases naturales *RPH*) y contando con el apoyo de réplicas provenientes de Charland (2002) y Prinz (2004), es posible colegir respuestas breves y puntuales para las objeciones 8 a 11.

V.III. RESPUESTAS A ALGUNAS OBJECIONES

El escepticismo que muestra Elster en la objeción (7), en torno al alcance de los métodos disponibles para desarrollar un programa robusto de investigación científica sobre las Emociones Cognitivas Superiores (ECS) ha sido francamente refutado por los hechos. Elster considera que es impracticable, principalmente por razones morales y materiales, realizar investigaciones científicas sobre ECS. Este autor parece suponer que, para estudiar la vergüenza, por ejemplo, habría que inducir la emoción en una persona en una situación cotidiana real. Esto, considera Elster, tiene la desventaja inicial de parecer un abuso y, en segundo lugar, requeriría de la elaboración de inviables escenarios e inexistentes instrumentos de medición para identificar y cuantificar los procesos neurofisiológicos que tienen lugar durante el episodio emocional. Más adelante, mientras exponemos nuestro modelo de una

teoría general de las emociones, citaremos varios estudios que han superado la brecha metodológica que Elster considera insalvable. No entraremos en más detalles por el momento. Hacia el final del capítulo VII, al lector le habrá quedado claro que el ingenio de los estudiosos de las emociones es mucho mayor del que Elster les concede.

En cuanto a las impugnaciones de Solomon (8), Rorty (9) y Ben Ze'ev (10), en general, suponen una concepción de clase natural bastante estricta, según la cual éstas se definen en términos de condiciones necesarias y suficientes (Charland, 2002). La apelación a una noción más liberal de clase natural, como la de clases *RPH* —en términos de la cual las clases naturales se definen a partir de racimos indeterminados de propiedades causalmente relacionadas— puede ayudar a mitigar el pesimismo de estos autores. En el caso de la objeción (10) de Ben Ze'ev, su caracterización de las emociones como *prototipos*, de hecho, es compatible con la posibilidad de que las emociones sean clases naturales *RPH*. Las clases naturales *RPH*, como los prototipos, se individúan por racimos indeterminados de propiedades, de suerte que, instancias particulares que exhiban una porción importante de las propiedades identificadas en la clase natural *RPH* o prototipo, podrán considerarse como miembros de la clase natural *RPH* o casos emparentados con el prototipo. Lo que distingue a las clases naturales *RPH* de los prototipos, es que las primeras definen los criterios de membresía de una forma más interesante desde un punto de vista ontológico y epistemológico. Mientras que lo único que vincula a las instancias particulares de emociones con el prototipo es la semejanza, en el marco explicativo de las clases naturales *RPH* lo que uniría a las emociones sería la identificación (empírica) de un conjunto de propiedades físicas relevantes causalmente vinculadas (homeostasis). Así, un pequeño giro a la tuerca conceptual de clases naturales, y el pesimismo de Ben Ze'ev podría desaparecer. Claro que el resto de la cuestión parece ser preeminentemente empírica. ¿Conforman las emociones (básicas y cognitivas superiores) una clase natural *RPH*? Aunque Griffiths responde que no, más adelante intentaremos ofrecer las bases para una respuesta distinta.

La objeción (9) de Amelie Rorty se funda, además de en una concepción estricta de clase natural, en una falacia o figura retórica conocida como *demonización*, la cual consiste en describir de la peor manera posible —en este caso— un hecho, para inclinarnos a asumir una actitud pesimista o negativa sobre el mismo. En este sentido, Rorty, desespera porque,

al hablar de emociones, habitualmente nos vemos orillados a asentir a afirmaciones contradictorias entre sí. Así, llegamos a conceder simultáneamente que las emociones son algunas veces pasivas y otras activas; a veces las explicamos en términos estrictamente psicológicos y otras en términos fisiológicos; a veces como racionales y otras como irracionales; a veces como voluntarias y otras como involuntarias. Estas tensiones, piensa Rorty, obligan a una reclasificación de las emociones, en la implícita inteligencia de que no pueden constituir una clase natural (Rorty, 1978, p. 141). El problema es que Rorty no considera la posibilidad real de construir, sobre la base de estudios e investigación conceptual y empírica, explicaciones para todas estas aparentes contradicciones en torno a nuestro concepto de ‘emoción’. La búsqueda de tales explicaciones, de hecho, constituye el eje temático de un programa de investigación multidisciplinario que, en la actualidad, está en plena marcha. Rorty no puede vaticinar, *a priori*, el fracaso de dicho proyecto. Que a esta autora le parezca difícil o imposible la tarea, no es suficiente para establecer su punto. De nuevo, la cuestión que intenta resolver Rorty a partir sólo de observaciones superficiales, requiere de mayor esfuerzo empírico y del uso de recursos conceptuales y teóricos más adecuados. En este último sentido, consideramos que la aceptación de la noción de clase natural *RPH* constituye un paso crucial.

La objeción (8) de Solomon también echa de menos la falta de precisión en nuestra definición del concepto ‘emoción’. Ya hemos observado que una concepción de clase natural más liberal, como la de clases naturales *RPH*, ayuda a mitigar la angustia que este hecho lingüístico puede ocasionar. Pero, además, Solomon infiere ilícitamente que ‘emoción’ es una categoría que remite a un fenómeno cultural/socialmente dependiente, (en oposición a un fenómeno natural/independiente de la cultura), a partir del hecho de que no todas las lenguas tienen el término ‘emoción’. Debe quedar claro que, en absoluto, la sólo prueba de falta de universalidad léxica de un término es suficiente para mostrar que se refiere a un hecho o fenómeno cultural-dependiente. Términos como ADN, quásar o electrón, surgieron y figuran principalmente en las lenguas de las sociedades occidentales, no tienen equivalentes en los lenguajes de culturas alejadas espacial e ideológicamente de occidente, y no por ello nos atreveríamos a afirmar que se refieren a entidades cultural-dependientes. Además, no deberíamos esperar que las clasificaciones emocionales que se derivan de la terminología emocional occidental, fueran completamente reivindicadas o respetadas, ni por el léxico

de lenguas de otras culturas, ni por la ciencia y la investigación empírica. En efecto, tendríamos que estar dispuestos a renunciar a parte de (o a todas) nuestras clasificaciones emocionales si la investigación empírica termina concluyendo que es apropiado hacerlo (Griffiths, 1997). En este escenario, ni nuestro lenguaje emocional ni el de ninguna otra cultura hubiera jugado jamás un papel relevante en relación a la cuestión del estatus de las emociones en tanto que clase natural. No consideramos que un eliminativismo masivo terminará siendo el resultado de una ciencia acabada de las emociones —y quizás, ni siquiera sea posible—, sólo intentamos señalar que los estudios léxico-transculturales por sí solos no son suficientes para desacreditar la tesis (iii), que sostiene que las emociones constituyen una clase natural. Otra vez, son indispensables estudios y evidencia empírica en distintos niveles y de un marco teórico-conceptual adecuado para poder llegar a una conclusión o algo similar.

Finalmente, en cuanto a la objeción (11) que avanza de Sousa, resulta cuestionable que tener un objeto formal definido sea una condición necesaria o suficiente para ser una clase natural. Los tigres, el agua y las galaxias son ejemplos aceptables de clases naturales, y ninguno tiene objeto formal. En cuanto a la suficiencia, los enunciados afirmativos tienen el mismo objeto formal que las creencias, la verdad, pero, ¿estaríamos dispuestos a afirmar que constituyen una clase natural? (Prinz, 2004, p. 80). De Sousa podría replicar que, si bien no todas las clases naturales se caracterizan por tener un objeto formal, sí es propio de ciertas clases naturales psicológicas —estados mentales— tener un objeto formal. Aunque en sí mismo es problemático establecer qué es para un estado mental tener un objeto formal y explicar qué función juega para su individuación (Teroni, 2007), podemos conceder que cualquier tipo de estado mental que aspire a ser considerado clase natural debe tener uno bien definido. En tal caso, aún es posible pensar que las emociones satisfacen este requisito. Más adelante, siguiendo a Prinz (2004), intentaremos probar que las emociones involucran “temas relacionales”. Los temas relacionales son aquello que es representado por la emoción (en tanto que fenómeno eminentemente fisiológico), y remite a algún tipo de relación entre el organismo y el entorno; una relación cuya temática principal es el ‘bienestar’ (well-being) del organismo. Esto significaría que las emociones tienen un *objeto formal relacional*, el cual es el ‘bienestar’. Pero no es el ‘bienestar’ en un sentido general, es el bienestar en tanto que está en juego en una situación concreta, que implica al organismo y su relación

actual con el entorno (natural o social). Más aún, al igual que las creencias, las emociones también tienen un criterio compartido de éxito, el cual consiste en detectar las situaciones que exhiben algún tema relacional y desencadenar en el organismo respuestas favorables que aseguren o apunten hacia su ‘bienestar’. Más adelante se abordará más detallada y detenidamente este aspecto representacional de las emociones. Por ahora, este bosquejo sólo tiene por objeto amainar las inquietudes expuestas por de Sousa, mostrando que es posible identificar un objeto formal para las emociones, con lo que su afirmación en contra de que las emociones conforman una clase natural —por carecer de objeto formal—, pierde sustento.

En relación a las objeciones 7 a 11, podemos decir que no representan amenazas graves o insalvables para la tesis (iii) y para la posibilidad de una teoría general de las emociones. Pero, aún es indispensable hacer lo propio con las observaciones y objeciones (1-6) que han planteado Griffiths y LeDoux. La estrategia a seguir consistirá en generar un marco teórico-conceptual en términos del cual sea posible interpretar y ordenar la evidencia disponible de forma tal, que resulte apropiado seguir considerando a gran parte de lo que identificamos como ‘emoción’ dentro de un mismo dominio ontológico-epistemológico.

VI. UN MODELO PARA UNA TEORÍA GENERAL DE LAS EMOCIONES

Las objeciones planteadas por Griffiths y LeDoux, parten de cierta interpretación de la evidencia empírica que, de entrada, es discutible. En este capítulo propondremos un modelo (teórico-conceptual) alternativo, a partir del cual es posible interpretar y articular dicha evidencia de forma tal que no nos comprometa con el desmembramiento del conjunto de todo aquello a lo que llamamos emoción. Por lo menos, el modelo que propondremos, no conducirá a una ruptura tan radical como la que propone Griffiths, según la cual, sólo las emociones básicas (EB) habrían de ser consideradas, propiamente, emociones; mientras que, en relación a las emociones cognitivas superiores (ECS), tendríamos que encontrar otra terminología, e incluso, desarrollar otro marco teórico distinto para hablar de ellas y estudiarlas. En la medida en que se progrese en la exposición de nuestro modelo, nos proveeremos de los recursos teóricos y conceptuales indispensables para responder a las objeciones (1-6) que permanecen sin respuesta. De hecho, veremos que, a partir del mismo modelo, pueden desarrollarse dos programas con soluciones distintas para la supuesta desunión de las emo-

ciones. Pero, antes de ofrecer más detalles sobre dichos programas, es indispensable exponer los detalles del marco teórico-conceptual, del modelo, o de la forma y supuestos en los que debería descansar una teoría general de la emoción.

VI.I. MARCO CONCEPTUAL

I. CLASES NATURALES DE RACIMOS DE PROPIEDADES HOMEOSTÁTICAS

La posibilidad de una teoría general de las emociones requiere que reconozcamos que *la mayoría*²² de los fenómenos que así llamamos, pertenezcan a la misma clase o categoría psicológica: independientemente de que sea considerada Emoción Básica o Emoción Cognitiva Superior. Esta es, básicamente la tesis (iii) que se expone al inicio de esta segunda parte (II, II). Para defender esta tesis (iii) que pretende mantener unidas a las emociones en una sola clase natural, necesitamos criterios para identificar a las entidades teóricas relevantes de una teoría de las emociones: esto es, las entidades sobre las cuales se pretende identificar leyes y correr generalizaciones e inducciones. En este punto, resulta ser de mucha utilidad la noción de clase natural, ya que, a partir de ella, es posible establecer los criterios metafísicos y epistemológicos en términos de los cuales individuaremos las entidades relevantes de nuestra —y de cualquier— teoría. En la primera parte de este trabajo (I, v.II) expusimos y defendimos una concepción realista —pero liberal— de clase natural (Boyd, 1979; 1991; 1999) según la cual, muchas de las entidades con cuya existencia se comprometen las teorías científicas respetables tienen como estructuras explicativas racimos de propiedades homeostáticas, indeterminadas y dinámicas (clases *RPH*). El modelo de una teoría general de las emociones que tenemos en ciernes, adopta esta concepción de clase natural para individuar a las entidades teóricas que constituyen su objeto de estudio. En otras palabras, todo lo que una teoría general de las emociones reconozca como tal, debe compartir una misma esencia, la cual se entiende como un racimo indeterminado de propiedades homeostáticas. Enseguida presentamos un recuento de las características distintivas de una clase natural entendida como clase *RPH*.

²² Es importante el adverbio de cantidad. Nuestra intención en mantener unidas a las emociones en una sola clase natural. Esto no implica que *todo* lo que llamamos emoción sea una emoción realmente. Algunos términos emocionales sólo son sinónimos de otros términos o refieren, más que a emociones, a rasgos contextuales o conductas asociadas con la emoción. Por lo tanto, quizás no todo lo que llamamos ‘emoción’ lo sea, pero quizá sí *la mayoría* (*infra*.II, VIII.II).

En primer lugar, las clases *RPH* tienen estructuras explicativas constituidas por racimos de propiedades homeostáticamente vinculadas, lo que significa que dichas propiedades están unidas por medio de mecanismos causales que favorecen la integración de dichas propiedades en el racimo. Esto hace que la presencia de algunas propiedades favorezca la presencia de las demás. Por ejemplo, los mecanismos homeostáticos que, en el agua, favorecen la aparición de propiedades como ser líquido, insípido e incoloro, promoverán la presencia de la propiedad de ser inodoro. En segundo lugar, las propiedades que constituyen el racimo esencial de una clase *RPH* no configuran un conjunto bien delimitado de propiedades. Esto es, las estructuras explicativas de las clases *RPH* no tienen límites bien definidos, son *indeterminadas* por naturaleza.²³ Esto implica que los mecanismos causales homeostáticos que mantienen unidas a las propiedades de un racimo esencial exhiben imperfección nomológica; es decir, es posible que algunas veces no se presenten todos los efectos o propiedades identificados generalmente en el racimo. La imperfección nomológica de los mecanismos homeostáticos, que a su vez da cuenta de la indeterminación en las propiedades que constituyen la esencia de una clase *RPH*, puede ser un rasgo metafísico-ontológico de la realidad misma, que se proyecta en el carácter *aproximativo e imperfecto* de muchas leyes y generalizaciones e inducciones en distintas disciplinas científicas. Esta falta de exactitud metafísica en la realidad se filtra, entonces, y llega hasta nuestras teorías científicas en forma de definiciones vagas o indeterminadas de entidades teóricas y leyes imperfectas. Pero esto, más que un defecto en nuestras teorías científicas, representa un avance respecto de nuestra imagen metafísica de la realidad. Quizás la realidad no cuente con un laser de alta precisión para cortar sus coyunturas ontológicas de forma exacta. Quizás, son sólo nuestros prejuicios filosófico/definicionales los que nos han llevado a pensar que así tendría que ser. El modelo que proponemos para una teoría general de las emociones se apega a esta imagen metafísica y, en este sentido, concede que las entidades teóricas que le interesan (las emociones) pueden exhibir esta imperfección ontológica y nomológica.

Una tercera característica de las clases *RPH* consiste en su carácter *dinámico*. Esto quiere decir que el racimo de propiedades homeostáticas a partir del cual se define a una clase *RPH* se puede concebir como un proceso u objeto histórico sujeto a evolución. Cier-

²³ Esta característica de las clases *RPH* impide que sus estructuras explicativas se especifiquen en términos de condiciones necesarias y suficientes.

tos cambios, en las propiedades o mecanismos que constituyen la esencia de una clase *RPH*, pueden presentarse a lo largo del tiempo sin que deje de ser la clase que es. Las emociones básicas o programas afectivos ejemplifican bien esta característica de las clases *RPH*. Las emociones básicas (EB) están presentes en muchas especies de mamíferos y son el resultado de presiones evolutivas que favorecieron su diseño y replicación —como adaptaciones— en los últimos miles de años. Durante este tiempo, es probable que hayan estado sujetas a alteraciones en su diseño, sin que esto impida que sigamos hablando de ellas como perteneciendo a la *misma* clase *RPH*. Esto se debe, también, al hecho de que parece apropiado individuar a las EB como se individua a las especies biológicas; esto es, en términos de *homología*. La homología representa un criterio de individuación que, más que fundarse en la semejanza o *analogía* estructural y funcional, se basa en la historia evolutiva que comparten los rasgos u organismos, para identificarlos. En la biología, la clasificación de rasgos y organismos (especies) en términos de *homología* es coherente con la concepción de clases naturales *RPH* porque, además de suponer el carácter *dinámico* de los mismos, lo reconoce como un aspecto fundamental para su individuación. Según el modelo de una teoría general de las emociones que proponemos, concedemos la posibilidad de entender a las emociones como clases *RPH* dinámicas cuya individuación puede realizarse en términos de homología.

2. PROPIEDADES 'FÍSICAS' RELEVANTES

También es necesario decir algo en torno a las propiedades que configuran los racimos o estructuras explicativas de las clases *RPH*. Hemos venido hablando de que dichas propiedades deben ser 'propiedades *físicas* relevantes'. Esta caracterización de las propiedades que constituyen una clase natural plantea una dificultad a la propuesta que intentamos articular. El estudio de las emociones, en la actualidad, implica a muchas disciplinas, y el conjunto de propiedades que se suelen considerar como constitutivas y relevantes para comprender y explicar el fenómeno emocional —desde esta diversidad— es, también, variado. Se habla —de forma muy general— de que las emociones involucran propiedades fisiológico/neurológicas, expresivo/conductuales, funcional/psicológicas, y hasta propiedades fenomenológicas, como su *quale* característico y valencia. El problema se presenta porque el modelo de una teoría general de las emociones que proponemos pretende incluir propie-

dades funcional/psicológicas y fenomenológicas dentro del racimo de propiedades homeostáticas que configuran la estructura explicativa de las emociones en tanto que integrantes de una clase natural. Y es un problema porque, si bien no resulta problemático reconocer el carácter físico de las propiedades fisiológicas/neurológicas y expresivo/conductuales, así como su relevancia para entender el fenómeno afectivo, no sucede lo mismo en el caso de las propiedades funcionales y fenomenológicas. Desde luego, el testarudo problema metafísico-ontológico que plantea la necesidad de explicar la relación de la mente con el cuerpo, nos advierte e impide sentenciar categóricamente que los aspectos funcional/psicológicos (Searle, 1980) y fenomenológicos (Chalmers, 1996), involucrados en las emociones, sean de naturaleza física. Por otra parte, no parece viable una teoría general de las emociones que no incluya sus dimensiones funcional/psicológicas y fenomenológicas. A pesar de que no sea categóricamente demostrable que dichas sean propiedades físicas, para muchos, incluso para aquellos que estudian las emociones desde un enfoque primordialmente fisicalista (Damasio, 1994; 2005; Panksepp, 2000; LeDoux & Phelps, 2000) sí es claro que son *relevantes* para entender y explicar a las emociones. Luego, un modelo de una teoría general de las emociones que deja fuera estos aspectos, no parece ser muy prometedor.

Curiosamente, las limitaciones que imponen consideraciones metafísico-ontológicas desde el campo de la filosofía de la mente, no siempre son observadas cuando se trata de asuntos propios de la filosofía de la psicología. Proponer un modelo para una teoría de las emociones es una cuestión más propia de la última rama de la filosofía, que de la primera. Por esta razón, es habitual encontrar que muchas teorías de las emociones de base fisicalista —por llamarles de algún modo a las teorías que se fundan principalmente en aspectos fisiológicos/neurológicos y conductuales— tienden a reconocer alguna *forma de relación*, no esclarecida de un modo preciso, pero aceptablemente identificada, entre procesos físicos, por un lado, y procesos funcionales y estados fenomenológicos, por otro. Las teorías de James (1890), Damasio (1994; 2005), Panksepp (2000), LeDoux (1996), LeDoux & Phelps (2000), Prinz (2004, 2005)²⁴, Russell (2003), Rolls (2005) y Barrett (2005), ejemplifican

²⁴ En su más reciente libro, Prinz (2012) señala la innegable fertilidad que tiene en neurobiología el enfoque reductivo. Este enfoque, a pesar de pasar por alto los detalles explicativos y ontológicos sobre los que insisten los filósofos de la mente, siguen tomándose muy en serio las propiedades fenomenológicas y funcionales, y las incluyen como *explananda* en sus investigaciones. Prinz considera que esta estrategia, que es la que nosotros asumiremos, ya ha dado sus dividendos y lo seguirá haciendo. Finalmente, aunque por ahora no haya

esta suposición. Detrás de todos estos casos, en los que la cuestión que se atiende tiene que ver con la arquitectura cognitiva de las emociones, parece operar una suerte de *principio de caridad teórico*, el cual consistiría en obviar una cuestión —pasando por alto los detalles escabrosos de la misma— para atender a una más específica y distinta. En el caso presente, la cuestión que se obvia es la de *la naturaleza de la relación* entre las propiedades funcional/psicológicas y fenomenológicas de las emociones, con sus propiedades físicas: neurológicas, fisiológicas y conductuales. Los detalles exactos de la *relación* es lo que se pasa por alto. La autoconcesión, por parte de muchas teorías de las emociones, del *principio de caridad teórico*, parece ser un aspecto de la práctica y metodología científica contemporánea en el estudio de las emociones. Independientemente de que adopten o no alguna concepción de clase natural que los comprometa a incluir en sus teorías sólo propiedades físicas, o de que suscriban alguna forma de fisicalismo o dualismo, el hecho es que la mayoría de las teorías parecen reconocer la existencia de una *relación íntima* entre las propiedades funcionales-fenomenológicas de las emociones, y sus propiedades fisiológico/neurológicas y conductual/expresivas. Lo que es importante reconocer aquí, es que no sólo el modelo que proponemos —dado su compromiso con la concepción de clase natural *RPH*—, sino cualquier teoría de las emociones que se tome en serio las propiedades funcionales y fenomenológicas y las incluya en su explicación del fenómeno emocional, tendría que adoptar una postura en relación a las siguientes cuestiones interrelacionadas: (1) ¿cuál es la *naturaleza de la relación* que dichas propiedades mantienen con propiedades físicas (neurofisiológicas)? ¿Y (2) cuál es el estatus ontológico que le otorgan a las propiedades en cuestión? En resumen, no sólo es problemático para nuestro modelo de una teoría de las emociones incluir propiedades funcional/psicológicas y fenomenológicas como parte de la estructura explicativa de las emociones; es problemático, en términos metafísico-ontológicos, para cualquier teoría de las emociones en las que figuren.

¿Qué tipo de relación se sugiere que existe entre las propiedades funcional/psicológicas y fenomenológicas, y las propiedades fisiológico/neurológicas? ¿Qué estatus ontológico se le concede a las propiedades funcional/psicológicas y fenomenológicas?

respuestas satisfactorias para el problema difícil, el hueco explicativo o la computabilidad del pensamiento, sí hay muchas cosas que se pueden explicar y entender con los conceptos, teorías, métodos y herramientas de que disponemos.

En general, la comunidad científica que estudia las emociones (neurofisiólogos, psicólogos, filósofos, etc.) no repara en reconocer la importancia de las propiedades funcionales y fenomenológicas en el estudio de las emociones. Más aún, según cierto consenso implícito, basado en copiosa evidencia empírica, es bastante aceptable la idea de que las propiedades en cuestión, de las emociones y de la mente en general, se *realizan o implementan* en el tejido neuronal. El reconocimiento de esa *relación* entre propiedades funcionales y fenomenológicas, y neurológicas, favorece y justifica la inclusión de las primeras en la ontología de una arquitectura cognitiva de las emociones.

No obstante, la cuestión en torno a la *naturaleza de la relación* entre, por un lado, propiedades funcionales y fenomenológicas y, por otro, neuronales, es difícil de decidir, en virtud de que la investigación de las emociones desde los campos de las neurociencias no suele tomar en consideración las sutiles distinciones que los filósofos de la mente realizan al atender esta cuestión. No es nada habitual encontrar discusiones en el estudio de las emociones acerca de si sus propiedades funcionales y fenomenológicas *emergen, supervienen* o son *idénticas* a las propiedades físicas o neuronales, con las que se las asocia. Antes bien, uno tiene que leer entre líneas para derivar las posturas al respecto.

Por ejemplo, con su aproximación neurofuncional, Panksepp (1998, 2000, 2008) y Rolls (2005) parecen decantarse hacia una *relación de tipo causal*: el tejido neural (propiedades físicas) es la *causa* de ciertas funciones y propiedades (procesos psicológicos y sensaciones) asociadas con las emociones. De hecho, en estos autores son bastante comunes aseveraciones contrafácticas del tipo que, para algunos, definen la relación de causalidad: si no se activan ciertas estructuras neuronales (por algún tipo de lesión), no tienen lugar (se pierden) ciertos tipos de funciones asociadas con las emociones. En este sentido, resulta que la indagación empírica de Panksepp y Rolls sobre las emociones no es de talante claramente reduccionista. Ambos evitan los enunciados de identidad interteórica y, en su lugar, parecen más inclinados a pronunciamientos sobre la forma en que las propiedades físicas (neuronales) determinan causalmente las propiedades funcionales y fenomenológicas. Podríamos concluir, entonces, y si no malinterpretamos del todo a estos autores, que identificarían como una forma de *emergentismo débil* a la relación existente entre las propiedades funcionales y fenomenológicas, y las propiedades físicas o neuronales: las últimas

serían causalmente responsables de la existencia de las primeras, y sería una cuestión de tiempo e investigación empírica elucidar los mecanismos causales detrás de estos fenómenos emergentes. En cuanto al estatus ontológico de las propiedades funcionales y fenomenológicas de las emociones, aunque el *emergentismo* es compatible con el dualismo, ni Panksepp ni Rolls reparan al considerar las propiedades en cuestión como físicas. Para ambos es muy claro que se realizan y requieren de tejido neuronal.

Por otro lado, Damasio (1994, 2005), Damasio, et. al. (1998), LeDoux (2000), LeDoux y Phelps (2000), Cosmides y Tooby (2008), Evans (2001) y Prinz (2004)²⁵, parecen adoptar una posición fisicalista más radical. En todos los casos mencionados el talante es reduccionista. Los enunciados condicionales son remplazados por aseveraciones de la forma ‘las emociones (propiedades funcionales²⁶ y fenomenológicas) no son más que tal y cual patrón de activación neurofisiológico’. De nuevo, es complicado aseverar categóricamente que los autores aludidos suscriben alguna variante de la teoría de la identidad en relación a las propiedades funcionales y fenomenológicas, pero la forma lógica de algunas de sus aseveraciones lo sugiere²⁷. Si es así, entonces una porción representativa de la investigación de las emociones suscribiría la idea de que la *relación existente* entre, por un lado, propiedades funcionales y fenomenológicas y, por otro, físicas o neuronales, es de *identidad* y, por lo tanto, las primeras tendrían un estatus obviamente *físico*.

No obstante, habrá quienes —principalmente filósofos— no consentirán tan fácilmente la caracterización de las propiedades funcionales y fenomenológicas como físicas, a menos que no estén completamente esclarecidos e identificados los detalles de la relación, ya sea causal (*emergentismo*) o de *identidad*, entre las propiedades en cuestión y las propiedades físicas —neurológicas— de las emociones. Los particulares que complican la situación son varios. Respecto de las propiedades funcional/psicológicas, la aproximación computacional —la idea de que las capacidades y procesos psicológicos son computaciones que se realizan en el cerebro— está lejos de resolverse. El conexionismo y la teoría compu-

²⁵ Prinz (2004, p. 211) representa el único caso esclarecido en este sentido, ya que abiertamente reconoce que el enfoque neurofuncional que adopta es reductivo, y caracteriza a la conciencia en términos de identidad con ciertos sucesos neuronales-físicos.

²⁶ Que van desde la detección y respuestas emocionales (LeDoux), hasta procesos de toma de decisiones (Damasio), reorientación de la atención, memoria, reorganización de prioridades (Cosmides y Tooby), etc.

²⁷ Para ejemplos de este tipo de afirmaciones véase, Damasio (2005, p.55), LeDoux y Phelps (2000, p. 162), Cosmides y Tooby (2008, p. 116).

tacional de la mente llevan años disputándose el epíteto de ‘la teoría que mejor modela la mente animal’. Y, para complicar aún más la situación, estos modelos, a su vez, siguen intentando dejar atrás planteamientos como los de Searle (1980), Dreyfus (1999) y Penrose (1989)²⁸. En cuanto a las propiedades fenomenológicas y su relación con las propiedades físicas-neuronales del cerebro, para muchos, el problema se torna más complicado. Chalmers (1996) y Levine (1983), han planteado, respectivamente, los que se conocen como ‘el problema difícil’ y ‘el problema del hueco explicativo’. Ambos insisten en el hecho de que, mientras no queden claros los detalles y mecanismos causales que expliquen *cómo* produce conciencia el cerebro, no estamos autorizados para sostener relaciones como la *emergencia causal* y/o la *identidad* entre propiedades fenomenológicas y propiedades físicas-cerebrales. Algo peor, según estos autores, es que en la actualidad no tenemos la más mínima idea de la forma que tendría una teoría que explicara la relación. Por estas razones, se podría insistir en que los investigadores de las emociones deberían, más bien, asumir alguna forma más débil de relación entre propiedades funcionales y fenomenológicas, y propiedades físicas-neuronales. Por ejemplo, alguna forma de *superveniencia*. Por las mismas razones, insistirían en dejar indeterminado el estatus ontológico de las propiedades funcionales y fenomenológicas. No obstante, se podría apelar a consideraciones heurísticas y conceptuales para adoptar, como la relación entre propiedades funcionales y fenomenológicas, y las propiedades físicas, una menos débil y más robusta que la *superveniencia*. De paso, lograríamos abonar a favor de la reputación ontológica de las propiedades en cuestión.

En relación a las propiedades funcionales, existe un modelo computacional de las capacidades y procesos psicológicos cuya arquitectura cuadra bastante bien con las propiedades anatómico-estructurales del cerebro. Se trata del modelo conexionista de procesamiento vectorial (Churchland, 1995). Con este modelo se han desarrollado explicaciones interesantes sobre la forma en la que el cerebro, a partir de conexiones neuro-químicas entre grandes poblaciones de neuronas ubicadas en estructuras cerebrales diversas, efectúa funciones psicológicas como el aprendizaje, el reconocimiento, la percepción, la categorización, etc., muchas de las cuales se tienden a asociar con las emociones. En este sentido,

²⁸ Searle, con el experimento mental del cuarto chino, sugiere que la computación no es suficiente para rasgos mentales como la intencionalidad y la conciencia. Dreyfus ha insistido en que la IA no podría jamás replicar los factores históricos y contextuales que, en un sentido crucial, determinan los rasgos más prominentes de la mente animal. Penrose, finalmente, argumenta que la conciencia es una función no computable.

una asunción irrenunciable de este modelo, es la de *identificar* las funciones psicológicas con los sustratos neuronales (físicos) los que se implementan y las hacen posibles. De nuevo, el talante fisicalista en torno a las propiedades funcionales parece estar fundado en un buen cuerpo de evidencia empírica y el éxito explicativo de las aproximaciones neurofuncionales.

La situación en relación a las propiedades fenomenológicas no es tan prometedora. A pesar del talante fisicalista general entre los investigadores empíricos de las emociones, en realidad, no tenemos idea de qué clase de mecanismos causales se activan o se requieren para que tengan lugar los estados mentales que exhiben fenomenología. Todo lo que tenemos son correlaciones, no explicaciones causales interesantes. El mismo Jesse Prinz (2004) reconoce que la aproximación neurofuncional que tanto éxito ha obtenido en otros aspectos del estudio de las emociones, no tiene una respuesta satisfactoria para el problema difícil o el del hueco explicativo. Sin embargo, esboza un argumento que tiene como finalidad limitar la fuerza de argumentos como los de Chalmers y Levine. Prinz indica que existen muchas teorías neurofuncionales de la conciencia, las cuales avanzan cada día más en la identificación de las estructuras y procesos neuronales de los cuales parecen depender las experiencias conscientes. Pero, ninguna explica *por qué* ciertos estados físicos se *sienten como algo*. Prinz sugiere que esta es una cuestión que no debería preocuparnos demasiado. Considérese la hipótesis de que el agua es H₂O. Lo que se sabe del comportamiento de dicho tipo de moléculas, cuando se agrupan, explica por qué el agua es líquida a temperatura ambiente, sólida a cero grados centígrados, etc, pero no explica *por qué* un cierto tipo de molécula es agua. La pregunta, ‘¿cómo qué es que el agua sea H₂O?’, obviamente, no tiene sentido. La *única* pregunta que parece pertinente en este sentido es ‘¿qué nos hace creer que el agua es H₂O?’. La respuesta a esta cuestión es que los encontramos en los mismos lugares e identificarlos ayuda a explicar varias de las propiedades del agua. Prinz propone que enfrentemos con la misma reluctancia la cuestión equivalente para el caso de la conciencia y el cerebro: ‘¿cómo qué es que la conciencia sea cierta activación cerebral?’ o ‘¿por qué la conciencia es el cerebro?’. La *única cuestión* válida, y para tomarse en serio, sobre la relación de la conciencia y el cerebro es, ‘¿qué nos hace creer que la conciencia sea o consista en cierta activación cerebral?’ Y, piensa Prinz, una buena respuesta a esta cuestión tendría la forma que tiene la respuesta para la cuestión análoga sobre el agua y las moléculas de

H₂O. Debe comenzar por el hecho de que encontramos experiencias conscientes donde hay cierta activación cerebral, y dicha actividad cerebral ayuda a explicar otras propiedades de la conciencia²⁹.

Difícilmente las consideraciones precedentes podrían decidir la cuestión en torno a la *naturaleza de la relación* entre las propiedades funcionales y fenomenológicas, por un lado, y las propiedades físicas o neuronales, por otro. Menos aún el estatus ontológico de las primeras. Pero, por otro lado, los escrúpulos filosóficos tampoco han sido lo suficientemente sólidos como para debilitar o detener la empresa de investigación empírica de las emociones, con todo y su talante reduccionista respecto de sus propiedades funcionales y fenomenológicas³⁰. Así las cosas, y ya que no podemos simplemente ignorar la relevancia de los argumentos antirreduccionistas, como tampoco podríamos ignorar el progreso y éxito del enfoque reductivo-fiscalista neurofuncional, optamos por adoptar una postura que intenta ser lo más justa posible con ambos enfoques, en términos de la cual concebiremos la *naturaleza de la relación* de las propiedades funcionales y fenomenológicas con las físicas o neuronales. Desde un enfoque que debe ser compartido tanto por la aproximación reduccionista como por la antirreduccionista, aceptamos que las emociones involucran propiedades funcionales-psicológicas y fenomenológicas, y que son tan relevantes en la producción e individuación de las emociones que deben figurar en cualquier teoría general y comprensiva de las emociones. En clave fiscalista, consideramos a dichas propiedades como *físicas*, toda vez que la indagación empírica reconoce y ha comenzado a identificar y explicar —hasta cierto punto— la íntima relación que tiene lugar entre las propiedades en cuestión y las propiedades físicas. En clave antirreduccionista, nos abstenemos de postular la relación de *identidad* como aquella que existe entre las propiedades funcionales y fenomenológicas, y las propiedades físicas. Hablaremos, más bien —y para respetar la terminología que ya introdujimos al considerar como hipótesis que las emociones son Clases Naturales *RPH*—, de que las propiedades físicas-neuronales, expresivo conductuales, funcionales y fenome-

²⁹ Otras propiedades funcionales de la conciencia son: 1) los estados conscientes pueden ser reportados (si no hay lesiones cerebrales que lo impidan). 2) Todos y sólo los estados conscientes son candidatos para ingresar a la memoria episódica. 3) Todos los estados conscientes son afectados por la atención (Prinz, 2004 y 2012).

³⁰ Algunos autores como Dylan Evans (2001) muestran muy poca paciencia ante los escrúpulos de algunos filósofos. Evans es hostil en particular con los planteamientos de Searle y Chalmers, los cuales se basan en experimentos mentales: cuartos chinos y zombis. Evans señala que las intuiciones no ayudan a determinar la cuestión de si los robots podrían llegar a ser capaces de tener las experiencias subjetivas propias de las emociones; v. gr. *conciencia*. Evans considera que esta cuestión se decidirá empíricamente.

nológicas de las emociones, están relacionadas e interactúan por medio de *mecanismos homeostático-causales* que, en algunos casos, no están del todo esclarecidos. Esta afirmación no necesariamente comporta una distinción ontológica entre propiedades funcionales y fenomenológicas y las propiedades físicas que las realizan. Más bien, indica una *distinción epistemológica*. Concedemos que las emociones constituyen un fenómeno muy complejo —que se asocia con propiedades muy diversas— que requiere ser estudiado en distintos niveles explicativos. Nosotros aceptamos la diversidad de propiedades y, por lo tanto, de niveles explicativos y postulamos, basados en el acopio de evidencia que se tiene al respecto, la existencia de vínculos causales en los distintos niveles explicativos. Lo que rechazamos rotundamente es cualquier forma de dualismo ontológico: todas las propiedades relevantes en el estudio de las emociones pueden ser consideradas como físicas, en el sentido epistemológico limitado de que pueden estudiarse empíricamente. Esta aproximación y la pluralidad metodológica que supone, al incluir distintos niveles explicativos, se verá apoyada por la discusión siguiente, sobre la metodología que proponemos seguir para una teoría general de las emociones.



VI.II. METODOLOGÍA: SIGUIENDO EL CAMINO DE LA REFERENCIA

Para indicar el camino (método) que debe seguir nuestro modelo de una teoría de las emociones, vale la pena recordar algunas observaciones de Boyd (*supra* I, v.II, 2) en torno a la forma en que refieren los términos de clase natural. Desde el punto de vista de Boyd, la introducción de un término de clase natural, por medio de una definición ostensiva y la acuñación de un término para designar un fenómeno, sustancia, proceso u organismo, da lugar a lo que este autor llama *apertura inductiva*. La *apertura inductiva* consiste, básicamente, en la inauguración de un programa de investigación que tiene como objetivo descubrir *empíricamente* la estructura explicativa —propiedades esenciales— del referente del término acuñado. La *apertura inductiva* es un proceso que se prolonga indefinidamente y que, con el tiempo, puede pasar de ser un programa de investigación informal a uno formal. Por ejemplo, pensemos en el oro y el término que en castellano utilizamos para referirnos a esta sustancia: ‘oro’. Para empezar, sin un término para referirse a dicha sustancia, sería imposible pensar en la articulación de una empresa informal cuyo objetivo fuera identificar sus propiedades esenciales. La acuñación de un término para el oro hace posible una empresa epistemológica tal, y es seguro pensar que dicha empresa —*apertura inductiva*— comenzó de manera informal con esfuerzos pre-científicos aislados llevados a cabo, principalmente, por especialistas primitivos —herrereros, joyeros y alquimistas. La *apertura inductiva* que, sobre su referente propició el término ‘oro’, se prolonga hasta nuestros tiempos y, desde hace un par de siglos al menos, es auspiciada y monopolizada formalmente por los expertos de las ciencias empíricas —y no más por alquimistas, por ejemplo.

Antes de seguir, es menester reconocer un par de rasgos relevantes de la *apertura inductiva*. Por un lado, es sensible a la selección de los métodos más apropiados para la identificación de las propiedades del referente de un término de clase natural; por otro, en distintas etapas exhibe cierta tolerancia hacia la pluralidad de enfoques para lograr dicha empresa epistemológica. En relación al primer rasgo, debe notarse que los métodos seleccionados son predominantemente empíricos, y con esto nos referimos a que se trata de métodos cuyos estándares de calidad están determinados por su alineación con los hechos y la experimentación. Respecto del segundo rasgo, es evidente, sobre todo, en las etapas previas a la consolidación de una o un cuerpo de disciplinas científicas. Este hecho se expresa clara-

mente en el caso de la *apertura inductiva* inaugurada por términos como ‘mente’ y ‘emoción’. En la medida en que no es apropiado hablar de una ciencia consolidada sobre los fenómenos a los que se refieren dichos términos, la investigación en torno a los mismos parece estar abierta a enfoques y aproximaciones diversas, las cuales incluyen niveles de análisis y de explicación distintos: neurológico, fisiológico, conductual, expresivo, psicológico y fenomenológico³¹. No obstante, hay que tener claro que esta diversidad de enfoques es deferente al primer rasgo mencionado: si bien se acepta cierta apertura de enfoques y aproximaciones, sólo los que respetan, en alguna medida, la autoridad de los hechos y la evidencia son bienvenidos al campo del estudio de la mente y las emociones.

Sobre el segundo rasgo característico de la *apertura inductiva*, hay que añadir que involucra lo que Boyd (1979, p. 510) denomina *división del trabajo cognitivo*. La *división del trabajo cognitivo* se refiere a las formas en que la gente —dentro de la comunidad en la que tiene lugar la *apertura inductiva*— divide responsabilidades para reunir información acerca de diferentes aspectos del mundo y, cuando se trata de un aspecto específico del mundo, se refiere a la forma en que la gente divide tareas y acumula información distinta pero, en apariencia, relevante, sobre el aspecto específico del mundo que resulta de su interés. Por ejemplo, en el caso de las emociones, la *división del trabajo cognitivo* remite a las tareas e información que aportan antropólogos, psicólogos, psiquiatras, etólogos, neurofisiólogos, médicos, filósofos, y hasta comunicólogos y publicistas. Todos los datos y evidencia que provienen de los enfoques de los especialistas enlistados, configuran y retratan la etapa actual de la *apertura inductiva* que comenzó cuando se acuñó el término ‘emoción’ o, quizás, antes, con la aparición de términos como ‘pasión’ y ‘sentimiento’.³²

³¹ Esto explica parcialmente la postura que adoptamos hacia el final del apartado anterior. Aceptamos que las emociones involucran propiedades muy diversas y, por lo tanto, su estudio requiere un abordaje desde distintos niveles explicativos.

³² Dixon (2003) señala que el término ‘emotion’ aparece relativamente tarde en la lengua inglesa (s. XVII), y sólo hasta la segunda mitad del siglo XIX se generaliza. Además, muestra que el término ‘emotion’ llega a sustituir o abarcar a un racimo de términos antes bien diferenciados: ‘passion’, ‘feeling’ y ‘affect’. En este caso, el término ‘emotion’ parece cubrir como una sombrilla los fenómenos designados por los términos mencionados. Esto significa, a la luz de la teoría de la referencia de Boyd, que, de algún modo, la *apertura inductiva* propiciada por el término ‘emotion’ heredó parte de las datos e información previas, obtenidos por la *apertura inductiva* a que dieron lugar los términos ‘passion’, ‘feeling’ y ‘affect’. Soca (2004) relata una historia bastante similar sobre el término castellano ‘emoción’. Aparece registrado por primera vez en diccionarios del idioma español del siglo XVII, proveniente del vocablo francés ‘émouvoir’, que significa conmovirse o emocionarse. No obstante, la palabra se generaliza sólo hasta el siglo XIX, al tiempo que caen en desuso palabras como ‘pasión’ y ‘afección’.

El modelo que proponemos para una teoría general de las emociones intentará seguir el camino que señala la *apertura inductiva*. Hablamos de que nuestro modelo debe vigilar que sean elegidos los métodos más apropiados para identificar las propiedades que constituyen las estructuras explicativas de las distintas emociones —lo que implica que debe favorecer métodos y aproximaciones eminentemente empíricos—, pero, al mismo tiempo, ha de mostrar tolerancia hacia la pluralidad de enfoques desde los cuales es posible, hoy en día, obtener datos e información —posiblemente— relevantes para identificar las propiedades relevantes de las emociones. En pocas palabras, se concede que un modelo prometedor para una teoría general de las emociones se funda en la *división del trabajo cognitivo*.

El enfoque metodológico que proponemos para construir una teoría de las emociones es consistente con el que es sugerido por otros autores para el caso particular de las emociones (Charland, 2002; y Prinz, 2004), y, en general, para cualquier asunto o temática en filosofía de la psicología. Bermudez (2005, p. 8), por ejemplo, insiste en la necesidad de que, incluso la reflexión más general y abstracta sobre la arquitectura de la mente, esté científicamente informada, y sea sensible a los hechos y evidencia disponible. Este requisito, llamémosle, de ‘deferencia empírica’, válido para cualquier teoría general o de algún aspecto particular de la mente, resulta indispensable dada la esterilidad explicativa derivada de la utilización de métodos que tienden a restarle importancia a los datos y evidencia empírica. En el caso de las emociones, las mal llamadas teorías cognitivas³³ —en filosofía— ofrecen un claro ejemplo. Estas teorías suelen³⁴ abordar la cuestión de las emociones con la convicción de que no es necesario, para entenderlas, apelar a los resultados y evidencia empírica emanada de ciencias como la neurofisiología y la psicología experimental. Otra convicción crucial vinculada con la anterior, es la que supone que el análisis del marco terminológico-conceptual de la “psicología popular” (creencias, deseos, intenciones) sería suficiente para llegar a desarrollar una aceptable teoría de las emociones. En varios casos de estas teorías, se puede identificar la polémica opinión de que la naturaleza de las emo-

³³ Griffiths (1997, p. 2) insiste en que el nombre más apropiado para estas teorías es ‘teorías de actitud proposicional’, toda vez que tienden a identificar a las emociones con actitudes proposicionales como creencias y deseos o, por lo menos, como esencialmente ligadas (las emociones) con actitudes proposicionales.

³⁴ Deberíamos hablar de estas teorías en pretérito, dado que cada vez son menos populares en su forma más pura.

ciones se podría elucidar —independientemente de la investigación empírica— realizando un minucioso y pormenorizado análisis de la forma en que usamos los términos y conceptos emocionales de la psicología popular.³⁵ Sobre la insuficiencia explicativa de esta metodología para el estudio de las emociones, Prinz (2004, p. 26-28) nota que la evidencia empírica está en contra de la tesis de que algún tipo de creencia o deseo o cualquier otra actitud proposicional, sea central para experimentar una emoción —la tesis nodal de toda teoría cognitiva de las emociones. Cualquiera que sea el tipo de actitud proposicional esencial para una emoción, es un hecho empírico que puede tener lugar dicha actitud proposicional y no presentarse la emoción. En cuanto a la metodología, como tal, Griffiths (1997, pp. 6-7) advierte que es riesgoso apelar exclusivamente a la información anidada en nuestros conceptos y las circunstancias en que usamos los términos que los refieren. Esto se debe a que nuestros conceptos sobre las emociones, tal y como están, pueden no capturar las determinaciones esenciales del referente. Analizar el concepto de una emoción no es otra cosa que exhumar nuestras creencias relativas a dicha emoción. Tales creencias pueden fundarse en experiencias, o en alguna suerte de atenta contemplación, o en lo que nos han enseñado sobre ellas. En todos estos casos existe la posibilidad de que nuestros conceptos ordinarios fallen para capturar los rasgos esenciales de su referente. Lo que sabemos sobre las emociones, basados en lo que se nos ha dicho o lo que hemos colegido a partir de nuestras propias meditaciones, puede ser inadecuado. Si extraemos conclusiones acerca de la naturaleza de las emociones sobre la base de un cuidadoso análisis de nuestros conceptos emocionales ordinarios, no hay garantía de que lleguemos a una teoría confiable de lo que son las emociones. Acaso sería posible una teoría de lo que está implicado en nuestros conceptos emocionales ordinarios. Peor aún, es posible terminar regurgitando algún lugar común de la psicología popular que resulte ser falso. La falta al requisito de ‘deferencia empírica’ puede desembocar en estas ruinosas consecuencias. Dadas estas posibilidades, nuestro modelo refrenda su compromiso con la selección de los enfoques más adecuados para entender y explicar qué son las emociones, lo que a su vez nos obliga a honrar el requisito de ‘deferencia empírica’³⁶.

³⁵ Kenny (1963), Solomon (1976; 2007), Lyons (1980) y Nussbaum (2001) son ejemplos de autores que, implícita o explícitamente se atienen a la metodología delineada.

³⁶ Autores como Elster (1999/2001) y Goldie (2000) han señalado que la literatura puede proporcionar una fuente muy relevante, si no es que la más fiable, de información acerca de la naturaleza de las emociones. Si

En el mismo sentido, Bermudez favorece una aproximación pluralista —*división del trabajo cognitivo*— a los problemas de la filosofía de la psicología. Sugiere que ni la mente en general, ni ninguna capacidad particular o fenómeno relacionado con ella, logrará elucidarse si optamos por posturas extremistas, según las cuales, o decidimos proveernos sólo de los datos y evidencia que provenga de disciplinas estrictamente empíricas como las neurociencias, olvidándonos de lo que puedan aportar disciplinas como la psicología y la filosofía, o viceversa (*op. cit.* p. 13). Respetar el requisito de ‘deferencia empírica’ no implica descartar *a priori* aproximaciones vinculadas, de forma menos evidente, con la evidencia empírica. De nuevo, el análisis filosófico puede seguir siendo de mucha utilidad. De hecho, en este trabajo, ha sido fundamental para revisar y cuestionar cierta concepción esencialista de clase natural que resulta demasiado estricta y problemática para abordar la cuestión del estatus de las emociones en tanto que clase natural. Más aún, el análisis filosófico —celoso del requisito de ‘deferencia empírica’— parece ser necesario para evaluar y reinterpretar —si es necesario— los marcos conceptuales, así como las implicaciones metafísicas y los compromisos ontológicos operantes en las teorías y prácticas científicas vigentes (Boyd, 1992). Por tal motivo, un modelo aceptable de las emociones debe observar cierta tolerancia a la pluralidad metodológica —*división del trabajo cognitivo*—, siempre que cada método o enfoque considerado dignifique, a su vez, del requisito de ‘deferencia empírica’.

VI.III. TEORÍA: LA COMPLEJA ESTRUCTURA EXPLICATIVA DE LAS EMOCIONES

Para aspirar a construir una teoría general de las emociones, es indispensable identificar la estructura explicativa (propiedades esenciales) de las emociones, ya que, a partir de estas, se propondrán enunciados válidos sobre todas ellas (leyes), los cuales harán posibles, a su vez, generalizaciones e inducciones. La identificación de la estructura explicativa de las emociones y las leyes que las gobiernan debe realizarse, como se ha sugerido en los apartados anteriores, teniendo presente la concepción de clase natural *RPH* y, además, siguiendo los compromisos metodológicos adquiridos, según los cuales hemos de considerar distintos tipos de aproximación al fenómeno de las emociones —*división del trabajo cognitivo*—, siempre y cuando respeten el requisito de ‘deferencia empírica’. En esta dirección, proponemos que una teoría general de las emociones debe comenzar estipulando que, todo aque-

bien no se descarta *a priori* tal posibilidad, nuestro enfoque exigiría la complementación y comparación de la información proveniente de la literatura con la que se obtiene por medio de métodos experimentales.

lo que haya de ser clasificado como una emoción, tendría que exhibir la mayoría de las propiedades enumeradas en un *racimo identificado de propiedades homeostáticas*. Pero, ¿qué propiedades constituyen el *racimo de propiedades o estructura explicativa* de las emociones? Distintas teorías de las emociones, con credenciales epistemológicas aceptables, proponen grupos de propiedades que, aunque difieren entre sí, se traslapan continuamente. A partir de una revisión de dichas propuestas es posible detectar la recurrencia de diez propiedades que, tentativamente, podemos considerar como candidatas a figurar en lo que podría considerarse como el esquema de la estructura explicativa de las emociones. Las diez propiedades aludidas se enlistan enseguida y son clasificadas, por su naturaleza, en cinco categorías:

I. Propiedades neurofisiológicas y expresivo/conductuales.

II. Propiedades biológicas.

III. Propiedades funcional-psicológicas.

IV. Propiedades o rasgos de su arquitectura.

V. Propiedades fenomenológicas.

Esta clasificación de las propiedades que, tentativamente, consideramos como constituyentes de la estructura explicativa o esencia de las emociones, será útil para abordar, más tarde, algunas cuestiones ontológicas y epistemológicas a la cuales debe responder una teoría general de las emociones.

I. Propiedades neurofisiológicas y expresivo/conductuales

1. Estructuras cerebrales especializadas. Una teoría general de las emociones debe presentar un caso sólido a favor de la tesis de que existe un conjunto de estructuras cerebrales especializadas, y evolutivamente diseñadas, para las emociones. Esto significa que cada emoción involucraría la activación de un circuito neuronal identificable y, además, que las estructuras cerebrales en las que se realizan dichos circuitos sean, aproximadamente, las mismas (Panksepp, 2000, 2008; Damasio, 2005). Sólo de esta forma, podría sostenerse que todas las emociones están emparentadas neurológicamente y que, por lo tanto, es razonable

pensar en un conjunto de *estructuras cerebrales especializadas* para su producción. En relación a las EB, son señaladas estructuras del sistema límbico como la amígdala, la cíngula anterior y el hipotálamo (entre otras), como candidatos viables para constituir las *estructuras cerebrales especializadas* para dichas emociones (Damasio, 1994, 2005; Panksepp, 1998, 2000; LeDoux & Phepls, 2000). En cuanto a las ECS, aunque parecen involucrar estructuras corticales que no figuran en el caso de las EB —como la región ventromedial del lóbulo prefrontal—, parecen utilizar gran parte de las estructuras del sistema límbico implicadas en la producción de las EB (Shin, *et. al.*, 2000). Quizás, a partir de esta sobreposición entre las EB y las ECS, se pueda defender el caso de un conjunto de estructuras cerebrales especializadas para las emociones y, de esta forma, responder al reto que plantea la objeción 6 planteada por LeDoux³⁷.

2. *Patrones de cambios fisiológicos.* Desde el tiempo de James (1884) se ha considerado como una característica distintiva de las emociones el que cada una involucra un patrón de cambios fisiológicos y endócrinos distintivo. Aunque esta tesis parece estar bien fundada para el caso de las emociones básicas (EB) (véase II, IV.II), hay quienes piensan (Griffiths, 1997; Harré, 1986; y Solomon, 1876; 2007) que este rasgo no está presente en las emociones cognitivas superiores (ECS)³⁸. Una teoría general de las emociones, que incluya a las EB y las ECS en su jurisdicción teórica, debe mostrar que ambas exhiben esta propiedad.

3. *Expresión y conductas características.* Se ha supuesto durante mucho tiempo que una propiedad distintiva de las emociones es la de involucrar expresiones faciales características y respuestas conductuales estereotípicas. De nuevo, en relación a las expresiones faciales características, existe un cuerpo de evidencia robusto a favor de que este es el caso, por lo menos para las EB (Ekman, 2003, 1994, 1973, 1972; Ekman & Friesen, 1971). En cuanto a las ECS, se tiende a suponer, a veces de forma precipitada, que no exhiben expresiones características, universales e innatas (Solomon, 1976, 2007). Una teoría general de las emociones debe mostrar que esta afirmación es falsa o explicar por qué es así. Algo similar sucede en el caso de la conducta estereotípica. Mientras se piensa que las EB implican el despliegue de conducta estereotípica, universal e innata, hay quienes sostienen que la conducta asociada con las ECS podría estar más condicionada por factores sociales y culturales (Els-

³⁷ Véase II, v.I.

³⁸ Véase la objeción 3 de Griffiths, capítulo anterior (IV).

ter, 1999; Solomon, 2007). De igual forma, una teoría general de las emociones debe refutar o explicar esta divergencia. Cumple agregar que la expresión y conducta características asociadas a las emociones están íntimamente relacionadas con las demás propiedades neurofisiológicas (1 y 2). De hecho, son el resultado de la activación de dichos mecanismos neurofisiológicos y tienen funciones adaptativas; esto es, su despliegue representa una ventaja evolutiva ante ciertas circunstancias recurrentes en el entorno (Griffiths, 1997; Ekman, 2003, 1999, 1992; Larsen, *et. al.*, 2008; Matsumoto, *et. al.*, 2008; y Cosmides y Tooby, 2000)

II. Propiedades biológicas

4. *Función adaptativa.* Como se acaba de mencionar, se asume que las emociones tienen una función adaptativa general y otra específica. En general, existen porque, ante situaciones de relevancia evolutiva, despliegan mecanismos específicos de respuesta (1-3) que favorecen la supervivencia y oportunidades reproductivas de los organismos de una especie (fitness). En términos particulares, cada emoción desencadena dichos mecanismos, en patrones específicos, para responder a distintos tipos de situaciones evolutivamente relevantes (Cosmides y Tooby, 2000). Pero, en este nivel, se suele sugerir que las funciones adaptativas de las EB y las ECS, divergen. Mientras la función adaptativa de las EB involucra una relación organismo-entorno natural, y dado que parecen estar presentes en especies distintas a la humana, pueden individuarse por homología, la función adaptativa de las ECS tiene que ver con lo que Frank (1988) llama *problemas de compromiso*. La función adaptativa de las ECS involucra una relación organismo-entorno social, dado que favorecen el cumplimiento de compromisos entre individuos de un grupo (amor, vergüenza, culpa, venganza). Dado que los *problemas de compromiso* son situaciones evolutivamente relevantes que sólo tienen lugar en los grupos humanos, las ECS deben ser específicas de los seres humanos, sin homólogos en otras especies. Esto implica que las EB y las ECS tienen historias evolutivas distintas.³⁹ Una teoría general de las emociones tiene que refutar estas observaciones u ofrecer una explicación para esta divergencia en la función adaptativa de las EB y las ECS.

³⁹ Véase objeción 5 de Griffiths II, v.I.

III. Propiedades psicológico/funcionales

5. *Carácter representacional.* Algunas teorías de las emociones subrayan el carácter representacional de las emociones. Este es el caso de las teorías valorativas (*appraisal theories*) en psicología (Lazarus, 1991), según las cuales las emociones involucran distintos tipos de juicios, sobre situaciones evolutivamente relevantes, el resultado de los cuales da lugar a una valoración o representación global de la situación. Prinz (2006, 2004, 2003a y 2003b) también defiende una teoría según la cual las emociones son representaciones corporeizadas, parecidas a las percepciones. Charland (2002), en la misma clave, sostiene que las emociones tienen un carácter representacional, y que dicha propiedad es un elemento crucial para defender la tesis (iii) que sostiene que las emociones constituyen una clase natural. Una teoría general de las emociones puede ahondar en torno a esta propiedad para alegar a favor de la unidad de las emociones.

6. *Efectos psicológicos característicos.* También se habla de que las emociones, además de desencadenar patrones de cambios fisiológicos, expresivos y conductuales, promueven variaciones en procesos psicológicos. Por ejemplo, Cosmides y Tooby (2000 y 2008) proponen que cada emoción es un programa que coordina otros subprogramas como el redireccionamiento de la atención y las motivaciones; formas de recuperar información y de acceder a información almacenada en la memoria; así como a sistemas especializados de inferencia y comunicación. Una teoría general de las emociones debe aportar argumentos y evidencia para sostener que todo lo que se ha de llamar emoción involucra algún patrón de efectos psicológicos como los mencionados.

IV. Propiedades o rasgos de su arquitectura

7. *Modularidad.* Algunos rasgos de las emociones —algunos identificables a primera vista y otros descubiertos mediante la investigación científica— sugieren que las emociones son modulares en el sentido de Fodor (1983)⁴⁰.

a) *Obligatoriedad.* Se activan automáticamente y no pueden interrumpirse voluntariamente.

⁴⁰ En el siguiente capítulo renunciaremos a esta versión fodoriana de modularidad, a favor de la versión de quasi-modularidad que propone Prinz (2004) para hacer frente a algunas dificultades relacionadas con el supuesto encapsulamiento de las emociones.

- b) *Opacidad*. No tenemos acceso consciente a los procesos cognitivos que tienen lugar mientras experimentamos una emoción.
- c) *Encapsulamiento o impenetrabilidad cognitiva*. En tanto que módulos, las emociones no tienen acceso a la información que se procesa en otros sistemas cognitivos.
- d) *Sub-sistema de aprendizaje enrutado*. Existe un sistema que facilita el aprendizaje de estímulos nuevos apropiados para activar la emoción.
- e) *Procesamiento rápido*. La activación y despliegue global de los procesos implicados en una emoción se llevan a cabo en un lapso bastante breve.
- f) *Arquitectura neural fija*. Se trata realmente de la propiedad 2, y supone que cada módulo emocional se asocia con actividad en ciertas estructuras cerebrales.
- g) *Patrones de disfunción específicos*. Dada la dependencia de los módulos emocionales de ciertos circuitos neuronales, el daño en las regiones cerebrales que implementan dichos circuitos implicaría la disfunción de dichos módulos.

El problema con atribuir modularidad a las emociones es el mismo que se presenta con otras propiedades. Existe cierta inclinación a pensar que, si bien las EB exhiben de forma clara los síntomas de la modularidad, las ECS no. Si se ha de sostener que las EB y las ECS conforman una sola categoría psicológica o clase natural, es necesario refutar o articular una explicación para la aparente divergencia.

8. *Sistema de recalibración de estímulos*. Las emociones se presentan ante una variedad de estímulos de distintos tipos. Una misma emoción puede activarse tanto por una percepción como por una imagen, recuerdo o razonamiento. Así, es posible que el miedo se experimente tanto al ver una serpiente como al considerar la posibilidad de que la situación financiera de la empresa donde laboramos resulte en un recorte de personal que podría incluirnos.

Una teoría general de las emociones debe explicar cómo es que las emociones pueden ser causadas por distintos tipos de estímulos y activadores (percepciones en distintas modalidades, fantasías, sueños, razonamientos), y debe hacerlo sin llegar a la conclusión de que esto se debe a que se trata, en realidad, de dos fenómenos completamente distintos: una

son Emociones Básicas y otras Emociones Cognitivas Superiores. En otras palabras, una teoría general de las emociones tiene que explicar por qué las ECS parecen requerir, para su activación, de procesos cognitivos más complejos, sin que dicha explicación implique que las ECS constituyen una categoría psicológica aparte, en virtud de que involucran mecanismos causales diferentes a las EB.⁴¹

Para explicar la plasticidad de las emociones respecto de la variedad de estímulos que las pueden llegar a desencadenar, algunos autores como Prinz (2006, 2004, 2003a y 2003b) y Cosmides y Tooby (2000 y 2008) han postulado la existencia de sistemas de recalibración emocional, a partir de los cuales es posible actualizar (eliminar o agregar) los estímulos que son apropiados, o dejan de serlo, para desencadenar una emoción. El problema con la postulación de sistemas de recalibración es que parece estar en tensión con la tesis de la modularidad de las emociones; en particular, con su supuesto encapsulamiento⁴² (Jones, 2006). Una teoría general de las emociones que apele a un sistema de recalibración para explicar la plasticidad de las emociones y, al mismo tiempo, sostenga que son modulares, debe explicar esta aparente inconsistencia.

9. *Duración breve.* Se piensa que las emociones aparecen de forma repentina y que su duración es corta (Elster, 1999; Griffiths, 1997; James, 1884). Pero, de nuevo, esta propiedad no parece generalizarse a todas las emociones. Se habla de emociones a largo plazo, como el amor, el rencor o el odio. Una teoría general de las emociones debe contar con una explicación para esta disparidad.

V. Propiedades fenomenológicas

10. *Fenomenología de las emociones.* Para autores como James (1884), la forma en la que se experimenta cada emoción constituía un rasgo distintivo de la misma. La dificultad para individuar cada emoción en términos de su *fenomenología característica*, hacía muy complicado incluir dicha propiedad en una teoría científica de las emociones (Schachter & Singer, 1962). No obstante, simplemente no podría hablarse de una teoría completa de las

⁴¹ Véase objeción 1 de Griffiths en II, v.I.

⁴² La tensión se presenta porque el encapsulamiento supone que las emociones, en tanto que módulos, no pueden ser influidos por otros módulos. No reciben información externa de otros sistemas o subsistemas. Esto, naturalmente, parece excluir la posibilidad de un sistema de recalibración de estímulos. Esto es, un sistema que entre a un módulo emocional y, de algún modo, altere y/o reorganice su funcionamiento.

emociones si ésta no le hace un espacio a su fenomenología. Algunos autores han hecho esfuerzos importantes para implicar este aspecto de las emociones en sus teorías. Los modelos circunflejos de Russell (1980) y Plutchik (1984), por ejemplo, otorgan un lugar preeminente a la fenomenología de las emociones por medio de las nociones de ‘valencia’ y ‘excitación física’ (*arousal*). Los modelos circunflejos —que organizan las emociones en un espacio ortogonal— consideran que la valencia y excitación son propiedades esenciales de las emociones porque los estudios revelan que son fundamentales para nuestras clasificaciones ordinarias de las emociones. Otra forma de entender la relevancia de la fenomenología emocional sugiere que ésta no radica en su utilidad para clasificar nuestras experiencias emocionales, sino en la *función* que desempeñan dentro de los estados emocionales mismos. Para estas aproximaciones, asociadas con Damasio (1999, 2005) y Prinz (2004), lo que debe explicar una teoría de las emociones en relación a su fenomenología, es cuál es su función y cómo surge o se relaciona con otras propiedades asociadas con las emociones.

VII. UNA TEORÍA GENERAL DE LAS EMOCIONES

VII.I. NATURALEZA DE LAS EMOCIONES

En el capítulo precedente se enunciaron cinco grupos o tipos de propiedades que, en la medida en que se tienden a asociar con lo que llamamos ‘emoción’, tentativamente, podrían constituir la estructura explicativa de las emociones. En este apartado, ofrecemos nuestra versión de los cinco tipos de propiedades que constituyen la estructura explicativa de las emociones: propiedades neurofisiológicas y expresivo/conductuales; propiedades biológicas; propiedades funcional-psicológicas; propiedades o rasgos de su arquitectura; y propiedades fenomenológicas. Al tiempo que las introducimos, tenemos en mente dos objetivos: en primer lugar, proporcionar buenas razones para caracterizar dichas propiedades de la forma en que lo hacemos y, en segundo lugar, generar una imagen global de las emociones que nos permita ver cómo se articulan los distintos tipos de propiedades en un solo fenómeno.

I. PROPIEDADES NEUROFISIOLÓGICAS Y EXPRESIVO/CONDUCTUALES

Las emociones son patrones distinguibles de cambios en individuos (pertenecientes a alguna especie de mamífero, principalmente), que se presentan en distintos niveles: (a) neuro-

nal-endócrino; (b) fisiológico (visceral y musculo-esquelético); y (c) expresivo-conductual. En términos causales, se entiende que son fenómenos cuyo desarrollo en tiempo real se origina *desde abajo*: en palabras más simples, se trata de procesos que comienzan en ciertas estructuras cerebrales (a) y extienden sus efectos en los niveles superiores mencionados (b-c). Según nuestra propuesta de una teoría general de las emociones, tanto las emociones básicas (EB) como las emociones cognitivas superiores (ECS) o sociales, no sólo deben instanciar (a), (b) y (c), además, debe existir cierta cartografía cerebral compartida: esto es, deberíamos esperar que activen aproximadamente las mismas estructuras cerebrales.

1. Estructuras cerebrales especializadas en la producción de emociones

La investigación neurofuncional de las emociones en los últimos años comienza a prefigurar un escenario en el cual es posible identificar un grupo de estructuras cerebrales que parece ser esencial en la producción y regulación de respuestas emocionales. Con base en dicho consenso es posible proponer la existencia de un racimo de estructuras cerebrales, integrado verticalmente, que evolucionó para y se especializó en la producción de respuestas adaptativas en situaciones recurrentes de relevancia evolutiva. Estas respuestas adaptativas las llamamos ‘emociones’. Por supuesto, no suponemos que dichas estructuras, por sí solas, sean responsables de producir o regular emociones. Como recién se indicó, consideramos que dichas estructuras están integradas verticalmente, desde regiones en el tallo cerebral conectadas con estructuras del sistema límbico y de la corteza cerebral, de forma tal que permiten la inscripción en ellas de circuitos cerebrales que activan complejos sistemas de respuesta ante situaciones evolutivamente relevantes: esto es, emociones. La idea, expuesta en otros términos, es que ciertas estructuras cerebrales describen un espacio anatómico-funcional integrado verticalmente en el cerebro, el cual es ocupado por distintos circuitos emocionales —emociones básicas y cognitivas— sobrepuestos en dicho espacio. Las estructuras tallo-cerebrales, límbicas y corticales identificadas en varios estudios tienen bajo su control la activación y regulación de respuestas viscerales, motoras y endócrinas, además de estar asociadas con la presencia de experiencias subjetivas y la reorientación y modulación de habilidades cognitivas como la atención, la memoria, el aprendizaje, etc. (Tucker, et.al., 2000). Sobra decir que todas estas funciones fisiológicas, expresivo-conductuales y psicológicas figuran en nuestra propuesta del racimo de propiedades home-

ostáticas que constituyen la estructura explicativa (esencia) de las emociones, y que, como se señaló antes, se activan *desde abajo* —nivel neuronal.

También vale la pena subrayar que el racimo de estructuras cerebrales especializadas⁴³ en la producción de emociones que estamos por introducir es consistente con la noción de clases naturales *RPH*. Las regiones y sistemas cerebrales a los que aludiremos como esenciales para la producción de emociones, acaso configuran un racimo indeterminado de estructuras de las que se sabe que pueden llegar a activarse y causar una emoción. Se acepta, en el mismo sentido, que *algunos* de estos sistemas o estructuras pueden no activarse o activarse escasamente en casos particulares. Pero esta indeterminación anatómica (en este caso) no afecta en absoluto el modelo. De hecho, es de esperarse cierta indeterminación. La identificación de sistemas cerebrales integrados en diferentes niveles especializados en la producción de emociones apunta específicamente a la posibilidad de construir un mapa (abstracto) en el que se delimita el espacio topológico/anatómico que, potencialmente, se puede ocupar o activar en el cerebro durante una emoción básica o cognitiva superior. Es natural esperar que cada emoción tenga su rúbrica (con su propio grado de indeterminación) y que esta consista en un circuito que activa, de cierta forma, sólo algunas de las estructuras cerebrales especializadas que se han identificado. Enseguida, ofrecemos un bosquejo del modelo de integración vertical del que venimos hablando.

Nivel del tallo cerebral. Ciertos grupos de neuronas en el **tallo cerebral** son activados indefectiblemente durante las experiencias emocionales. Los núcleos celulares en el tallo cerebral son responsables de respuestas somáticas (movimiento del tronco axial, los músculos de las extremidades, músculos de la cabeza, vocalización, expresiones faciales y movimientos del ojo) y autónomas en el sistema simpático (dilatación de pupilas, aumento de fuerza y frecuencia del ritmo cardíaco, dilatación de los bronquios y disminución de las contracciones estomacales). Además, en ciertas regiones del tallo cerebral se promueve la producción de **serotonina**, **norepinefrina** y **dopamina**, las cuales inhiben que información del cuerpo ascienda a las regiones somatosensoriales. Esta actividad neuroquímica propicia la analgesia que tiene lugar en episodios afectivos estresantes. Más aún, otros componentes del tallo cerebral envían señales neuroquímicas al **procencéfalo**, desde donde es posible

⁴³ Aunque no exclusivas. Las estructuras toman parte en la realización de otras funciones psicológicas.

modular la atención. La **sustancia gris periacueductal** está situada en lo profundo del tallo cerebral y se asocia con sensaciones agradables y desagradables, y patrones motivacionales relacionados con el dolor y el miedo. La estimulación en distintas regiones promueve conductas defensivas, conductas defensivas agresivas, conducta de escape y conductas de reposo. Todas parecen ser respuestas específicas a distintos niveles de dolor.

En la medida en que, según nuestra teoría, estas respuestas motoras, autónomas y reguladoras son parte de toda emoción, proponemos que algunos núcleos celulares y componentes del tallo cerebral —no por sí solos, si no en la medida en que se integran con estructuras límbicas y corticales específicas— son parte del conjunto de *estructuras cerebrales especializadas* dedicadas a la producción de emociones.

Nivel del sistema límbico. En cuanto a su contribución a las emociones, el **hipotálamo** tiene una función reguladora e integradora con el tallo cerebral. Recibe información sensorial exteroceptiva y propioceptiva a partir de la cual envía señales al tallo cerebral para orquestar respuestas motoras, autónomas o reguladoras adecuadas. Se comunica también con la glándula pituitaria para regular la actividad endócrina a la luz de las respuestas motoras y autónomas antedichas. Además, es capaz de inhibir las respuestas de defensa agresiva que se desencadenan en la **sustancia gris periacueductal**. Panksepp (1998) piensa que esta inhibición de la agresión de origen tallo-cerebral ha sido fundamental para conductas favorables para la socialización como la confianza, los juegos y el afecto.

En la medida en que el **hipotálamo** desempeña un papel crucial en la activación y regulación de respuestas motoras, autónomas y reguladoras (endócrinas), es plausible considerarlo como una estructura cerebral especializada para producir emociones —de nuevo, no por sí sola. En particular, su contribución en el caso del miedo está bastante bien esclarecida y establecida. Gray (1996) sugiere que el **hipotálamo**, la **sustancia gris periacueductal** y la **amígdala** trabajan en conjunto para elegir el tipo de respuesta defensiva adecuada en situaciones de peligro: pelear (en caso de que la amenaza sea próxima e ineludible) o escapar (en caso de que la amenaza sea distante o haya posibilidad de huir). Gray también sugiere que ciertas regiones corticales pueden activarse y formar parte del circuito del miedo en casos en que el peligro exija una coordinación más compleja de respuestas apropiadas. Más todavía, el **hipotálamo** forma parte de lo que se conoce como **eje hip-**

talámico-pituitario-adrenal, el cual está detrás de gran parte de la neuroquímica de muchas emociones cognitivas superiores, como el amor (Esch y Estefano, 2005), los celos (Markova, et. al., 2010), la vergüenza (Clark, 2010) y el orgullo (Tracy, 2010).

La **amígdala**, dada su comunicación con el **hipotálamo** y el **tallo cerebral**, está asociada con la organización y activación automática (no mediada por cognición alguna) de respuestas (emocionales) motoras y autónomas a partir de estímulos asociados (de forma innata o aprendida) con el peligro. Pero también puede realizar esta misma función a partir de información procedente de estructuras corticales como el lóbulo frontal en su región ventromediana (Damasio, 1996). Su activación no sólo ha sido comprobada para el caso del miedo (LeDoux, 1996 y 2000), sino también en emociones negativas básicas como el enojo, el asco y la tristeza (Damasio, 1996); y en casos de emociones cognitivas superiores como los celos (Markova, et. al., 2010) y la vergüenza (Clark, 2010). En todas estas emociones, su función es la misma, coordinar respuestas motoras (músculo-esqueléticas) y autónomas (cambios a nivel visceral y endócrino). Y, como menciona Rolls (2005), es muy probable que investigaciones subsecuentes revelen que otras regiones de la amígdala, no estudiadas en detalle aún, se hallen involucradas en la ocurrencia de otras emociones cognitivas. En este mismo sentido, Rolls (2000) y LeDoux (1996) coinciden en que la amígdala puede estar involucrada incluso en emociones positivas como la alegría y el orgullo, aunque reconocen que no se sabe aún en qué consistiría exactamente su función en estos casos. Por esta razón, LeDoux llega a afirmar que, si ha de existir una pieza central de un sistema cerebral emocional, el mejor candidato sería la amígdala (*op. cit.* p. 222). Después de todo, si las emociones involucran cambios a nivel músculo-esquelético, visceral y endócrino, es natural que señalemos a la amígdala como la principal responsable de ellos y, por lo tanto, de las emociones; finalmente, orquestar dichos cambios es parte de su trabajo.

Antes de pasar a la siguiente estructura límbica especializada en la producción de emociones, vale la pena reconocer otras funciones, propias de la **amígdala** que son de carácter estrictamente emocional. Por un lado, se sabe que la amígdala es responsable de lo que se conoce como aprendizaje emocional. El aprendizaje emocional (Gallagher, 2000) es una forma de asociación condicionada, en la que un tipo de estímulo que regularmente ha involucrado experiencias desagradables, (aunque también puede tratarse de un estímulo

agradable) tiende a activar con su percepción posterior una respuesta emocional específica: miedo. Por ejemplo cuando a una rata se le expone a un sonido con un tono particular que, en casos anteriores ha sido seguido de una descarga eléctrica, la información auditiva (ya asociada con experiencias desagradables) llegará al núcleo lateral de la amígdala y ésta desencadenará (con apoyo del hipotálamo) respuestas musculo-esqueléticas, viscerales y endócrinas en la rata. Se sabe que la amígdala es crucial para este tipo de aprendizaje. La **amígdala** también es importante para el reconocimiento de las emociones a partir de las expresiones faciales, estímulos léxicos (palabras u oraciones con significado emocional) y estímulos auditivos (tono, volumen y otros rasgos de la voz). La investigación de casos de pacientes con lesiones selectivas bilaterales en la amígdala sugiere que la capacidad de reconocer emociones depende de un modo crucial de la amígdala (Adolphs y Tranel, 2000).

El **hipocampo**, como la **amígdala**, es crucial para la identificación y anticipación de situaciones emocionalmente relevantes pero, a diferencia de la última, que establece asociaciones entre objetos o rasgos específicos del entorno y experiencias desagradables y agradables, el **hipocampo** puede establecer asociaciones entre rasgos estructurales y contextuales en el entorno, asociados a experiencias agradables o desagradables. Esto significa que el **hipocampo** aporta otra vía de identificación de rasgos emocionalmente relevantes en el ambiente; en este caso, a partir de rasgos contextuales se logra anticipar y preparar respuestas del organismo para situaciones emocionalmente relevantes (LeDoux, 1996). También se sabe (Gray y McNaughton, 1996) que el **hipocampo** puede inhibir respuestas motoras defensivas (involucradas en emociones como el miedo, el enojo y los celos) activadas desde la **amígdala** hacia el **hipotálamo** y ciertas regiones del **tallo cerebral**. Esto es posible en virtud de que las asociaciones entre estímulos emocionalmente relevantes y experiencias (emocionales) agradables o desagradables del **hipocampo** —basadas en rasgos contextuales del entorno y no en tipos de rasgos u objetos— pueden llevar a una suspensión momentánea de las respuestas ordenadas desde la **amígdala** —que tienen como causa asociaciones basadas en tipos de objetos o rasgos específicos. La idea es que puede presentarse un conflicto sobre la relevancia emocional de una situación, toda vez que el estímulo (objeto o rasgo específico) que activa la **amígdala** y el que activa **hipocampo** (rasgos contextuales) pueden no tener la misma relevancia y urgencia emocional. Por esta razón el **hipocampo** puede influir en la inhibición de las respuestas defensivas (motoras, autónomas y endó-

crinas) que orquestan la **amígdala** y el **hipotálamo**. Por ejemplo, el miedo ante la presencia de un león podría desencadenarse desde la **amígdala** y el **hipotálamo**. Ambas estructuras reaccionarían a rasgos específicos de un objeto: en este caso, el león. Si lo vemos en un zoológico, a través de rejas, cristales blindados y acompañados de otros visitantes y vigilantes, se inhibirían parcialmente respuestas como huir y gritar por ayuda. Esto se explicaría porque, por medio del **hipocampo**, se detectan rasgos contextuales (rejas, cristales, vigilantes) que inhiben las respuestas orquestadas por **amígdala** y el **hipotálamo**.

Nivel cortical. Tucker, Derryberry y Luu (2000) destacan un par de circuitos cortico-límbicos que son cruciales para el aprendizaje y la memoria emocional. El primero, llamado **circuito ventral** involucra el **núcleo mediodorsal del tálamo**, la **amígdala** y algunas regiones del **lóbulo frontal ventromedial**. Sobre estas estructuras, Damasio (1994) ha mostrado que también son cruciales en la producción de emociones cognitivas. El segundo circuito implica el **núcleo anterior del tálamo**, el **hipocampo**, la **amígdala** y la región **dorsomedial del lóbulo frontal**. A este circuito de aprendizaje y memoria emocional se le conoce como **circuito dorsal**. El **circuito ventral** se relaciona con la memoria de objetos emocionalmente relevantes. Nadel (1992) ha sugerido que este circuito es importante para la memoria categorial; es decir, la memoria basada en la identificación de objetos. El **circuito dorsal**, por su parte, se asocia con la memoria emocional codificada en términos de rasgos contextuales como características espaciales, procesos y conductas presentes en situaciones emocionalmente significativas. Así, la alegría de ver a un viejo amigo en la calle, después de mucho tiempo de no hacerlo, en tanto que involucra la identificación de un objeto emocionalmente relevante, involucraría la activación del **circuito ventral**. Sentir miedo al recorrer los pasillos de un hospital en el que hace tiempo sufrimos un procedimiento quirúrgico y una rehabilitación muy dolorosa, en tanto que implica el reconocimiento de rasgos espaciales/contextuales emocionalmente relevantes, implicaría la activación del **circuito dorsal**. La identificación de estos circuitos de memoria y aprendizaje emocional muestra que ciertas regiones del lóbulo prefrontal tienen una profunda influencia en la producción y regulación emocional, toda vez que son capaces de activar o modular las respuestas emocionales que se disparan desde el sistema límbico. El papel productor e inhibitorio del lóbulo prefrontal en respuestas emocionales está bien documentado por diversas investigaciones, entre ellas, las de Damasio (1994), sobre los déficits emocionales que dejan le-

siones en el **lóbulo frontal en la región ventromedial**; las de Vrana, et. al. (1988), sobre la influencia que la activación de vastas zonas de la corteza tienen en la inhibición del reflejo de sobresalto; y en las de Ploog (1981), que muestran que la activación del lóbulo prefrontal contribuye en la regulación de las vocalizaciones emocionales en primates.

La **ínsula** es otra estructura cortical que ha sido ampliamente reconocida como esencial para las emociones. En particular, la **ínsula**, junto con la **corteza cingulada anterior** (Damasio, 2005), parecen ser cruciales para los sentimientos (fenomenología) asociados con una gran variedad de emociones: alegría, tristeza, asco, amor romántico, culpabilidad, lujuria. Incluso, la **ínsula** específicamente, se considera crucial para lo que Hsu, Anen y Quartz (2008) llaman “sentimientos de justicia y de utilidad” y “repugnancia moral” y, en general, para las emociones sociales (Lamm y Singer, 2010). A partir de estas observaciones se tiende a pensar a la **ínsula** y a la **corteza cingulada** como estructuras cruciales para la experiencia subjetiva (fenomenología) característica de las emociones. La asociación se explica porque también se sabe que la **ínsula** recibe información de los cambios que tienen lugar a nivel intra-abdominal y visceral, los cuales también se consideran constitutivos de las emociones (Duque, et. al. 2004). Resulta interesante notar que la **ínsula** es crucial para la experiencia subjetiva tanto de emociones básicas (alegría, tristeza, asco) como de emociones cognitivas superiores (amor romántico, culpabilidad y repugnancia moral)⁴⁴. Además, se sabe que la **ínsula** se relaciona funcionalmente con el control del ritmo cardíaco y la presión arterial, y el control visceromotor y de las funciones gástricas (Duque, *op. cit.*). Sobra decir que todas las funciones recién descritas también figuran como componentes esenciales de las emociones, en tanto que respuestas viscerales.

Las Emociones Cognitivas Superiores y el Sistema Cerebral Emocional. Algunos estudios recientes abonan para la idea de que algunas de las que podríamos catalogar como Emociones Cognitivas Superiores (ECS), al activarse, instancian circuitos que utilizan las estructuras cerebrales que hemos señalado como especializadas para la producción de emociones. El amor romántico (Esch y Stefano, 2005), por ejemplo, implica actividad a nivel amigdalario, hipotalámico y tallo-cerebral. La rica neuroquímica del amor también involucra la activa participación del eje hipotalámico-pituitario-adrenal (HPA), el cual es responsable de

⁴⁴ *Infra*, II, vii.i, V.

segregar los neurotransmisores definitorios del amor: oxitocina, vasopresina y, en menor medida, dopamina. Estos procesos en el nivel límbico son responsables de respuestas fisiológicas y conductuales asociadas con el amor: estrés por la separación, ansiedad, nerviosismo ante el objeto amado, protección. La corteza anterior insular es crucial también para la experiencia subjetiva del amor. En el caso del amor, la neuroquímica que lo soporta ayudaría bastante para resolver problemas de individuación en términos neurofisiológicos. Los celos (Markova, 2010) tienden a activar estructuras límbicas asociadas también con emociones negativas como el miedo y la tristeza. El eje HPA es el que media y regula la química de tales respuestas. La circuitería neuronal de otras ECS apenas comienza a ser develada. La vergüenza (Clark, 2010) también involucra estructuras límbicas y actividad del eje HPA. Tracy (2010) refiere a estudios sobre la neurofisiología del orgullo que señalan al la **corteza prefrontal en la región ventromedial** como involucrada en la emoción, junto con otras regiones corticales como el surco temporal superior posterior y el lóbulo temporal izquierdo. Aunque los detalles son más bien limitados en el caso de las ECS estudiadas en términos neurofisiológicos y resultaría muy complicado hablar de circuitos neuronales responsables de dichas emociones, es interesante identificar en cada uno de los estudios la participación recurrente de las mismas estructuras cerebrales.

Consideraciones finales. El conjunto de estructuras que se ha mencionado, junto con su contribución en la producción y regulación de las emociones, apenas logra ser una representación general y minimalista de lo que podríamos pensar que es, en realidad, un sistema de estructuras cerebrales integrado verticalmente especializado en las emociones. Una caracterización más adecuada y precisa podría incluir más estructuras cerebrales, más especificaciones sobre qué regiones de las estructuras mencionadas desarrollan qué función, más detalles sobre las reacciones neuroquímicas que pueden desencadenar dichas estructuras y más información acerca de la forma en que estas mismas estructuras se comunican y regulan o son reguladas por áreas del cerebro especializadas en funciones cognitivas más complejas; como el lenguaje, el razonamiento, la toma de decisiones, etc. Aunque dispersa, existe muchísima evidencia e información que, hoy en día, alcanzaría para desarrollar una representación más pormenorizada de las estructuras cerebrales especializadas en la pro-

ducción de emociones, que la que arriba se propuso⁴⁵. Por cuestiones expositivas —pensamos que sería innecesario y tedioso ahondar en detalles técnicos—, optamos por ofrecer tan sólo una imagen general de un sistema emocional integrado en el cerebro que, hasta donde la evidencia lo sugiere, es responsable de desencadenar la mayoría de las emociones que conocemos, básicas o cognitivas. Consideramos, junto con varios especialistas, que hay razones para ser optimista respecto de la posibilidad de lograr identificar un sistema afectivo/emocional en el cerebro. Si ajustamos ciertas tuercas conceptuales y metodológicas, de hecho, es bastante viable. Una estrategia conceptual y metodológica indispensable para tener éxito en la empresa mencionada la han sugerido y seguido —y nosotros con ellos— especialistas en la neuroanatomía y neurofisiología de las emociones como Damasio (1994, 2005), Rolls (2000; 2005), Panksepp (2000) y Tucker, et. al. (2000), entre otros. Se trata de intentar comprender las bases neurofisiológicas de distintos tipos de conductas complejas partiendo de la combinatoria de la anatomía y fisiología del cerebro. La idea es pensar en sistemas (conjuntos de estructuras cerebrales) integrados verticalmente. La integración vertical puede ayudar a explicar cómo muchas capacidades psicológicas se presentan en distintas advocaciones, desde formas simples hasta formas más elaboradas y complejas: por ejemplo, el aprendizaje y las emociones. Las formas simples implicarían estructuras cerebrales ubicadas en las regiones filogenéticamente más antiguas del cerebro —tallo cerebral y sistema límbico—, mientras que las formas más complejas ocuparían estructuras cerebrales filogenéticamente más modernas —corticales— que se montan sobre las más antiguas para lograr el mismo género de respuestas adaptativas, sólo que de maneras más flexibles y complejas. Un sistema emocional, entendido en este sentido, ha de entenderse como un espacio anatómico-fisiológico cerebral que involucra un racimo de estructuras integradas verticalmente que tienen como función —ya que se trata de emociones— orquestar y regular patrones complejos de respuestas fisiológicas y psicológicas —con diversos grados de complejidad— que tienen como objetivo preparar al organismo a enfrentar situaciones de relevancia evolutiva. Pensar los sistemas neuro-funcionales en estos términos representó, como señala Damasio (1994), crucial para superar la frenología, y seguir

⁴⁵ Incluso algunos estudios (Damasio, et. al., 2000 y Phan, et. al., 2002) han mostrado que distintas emociones activan patrones distintivos de activación neural a través de las estructuras y sistemas mencionados en este apartado.

haciéndolo será crucial para lograr mejores caracterizaciones de las capacidades psicológicas humanas.

También hay que insistir en el hecho de que el racimo de estructuras cerebrales especializadas en la producción de emociones que hemos propuesto, es compatible con la noción de clases *RPH*. Dichas estructuras configuran un racimo indefinido de regiones cerebrales vinculadas causalmente. Cada emoción instancia un patrón o circuito específico de actividad e interacciones causales en las distintas estructuras especializadas. La activación de cada circuito también se da dentro de límites de interacción causal más bien difusos. En este caso, la indeterminación anatómica y nomológica —que aquí se entiende como la variabilidad de las estructuras del circuito que se activan y su nivel de actividad— se explica porque, el circuito emocional, como un todo, se auto-regula homeostáticamente: es decir, dependiendo de la urgencia, persistencia e inminencia que exhiba el estímulo que activa al circuito, este regulará y actualizará continuamente la actividad de las estructuras que involucra. En este sentido, los circuitos emocionales son literalmente homeostáticos.

Sólo resta hacer un comentario en torno a la sexta observación⁴⁶ en contra de una teoría general de las emociones, avanzada por LeDoux (1996), según la cual habría muy pocas posibilidades de identificar un sistema neuronal dedicado exclusivamente a la producción de emociones. Pero este no sería el caso si las regiones cerebrales tipificadas en este apartado logran, en efecto, referir a un conjunto de estructuras especializadas en la producción de emociones: y, afortunadamente, consideramos que hay buenas razones para pensar que es así o que, por lo menos, representan un paso sólido en esa dirección.

2. Patrones de cambios fisiológicos

La añeja tesis de que las emociones activan patrones de trastornos a nivel fisiológico parece estar relativamente bien establecida —al menos para el caso de las emociones básicas⁴⁷. De hecho, debería ser una consecuencia obvia derivada de la seria posibilidad de que cada emoción instancia un circuito dentro de un sistema cerebral especializado en producir emociones (véase apartado anterior). Ya que los circuitos neuronales emocionales movilizan estructuras —tallo-cerebrales y límbicas— que activan cambios a nivel somático, visceral y

⁴⁶ Véase II, v.II.

⁴⁷ Véase II, IV.II.

autónomo, es natural esperar que cada circuito emocional desencadene un patrón de respuestas fisiológicas adecuadas para enfrentar la situación evolutivamente relevante para la que fue diseñado, como solución, el circuito emocional. Pero, antes que obviar nada, en este apartado intentaremos dar respuesta a algunos posibles cuestionamientos dirigidos a la tesis que ahora discutimos. En primer lugar, habría que especificar qué tipos de respuestas fisiológicas involucran las emociones. En segundo lugar, es necesario decir algo sobre la variedad de respuestas fisiológicas —con distintos resultados conductuales— que una sola emoción puede instanciar. Por ejemplo, el miedo puede activar conductas de huída, de pelea o de inmovilidad —dependiendo de la situación— y se sabe que cada una de las conductas requiere de la activación de distintos patrones de cambios fisiológicos. Hay que ofrecer una explicación para esta diversidad. Finalmente, es indispensable mostrar que las Emociones Cognitivas Superiores —en contra de lo que algunos piensan⁴⁸— también instancian patrones de cambios fisiológicos característicos. Lo contrario representaría un revés severo contra nuestra propuesta de una teoría general de las emociones.

¿Qué tipo de respuestas fisiológicas instancian las emociones? Las regiones y estructuras cerebrales especializadas en la producción de emociones activan respuestas a nivel del sistema nervioso autónomo (SNA). Las respuestas en el SNA implican cambios en sistema simpático y parasimpático. El primero, a grandes rasgos, prepara al cuerpo para la acción (p. ej. estimulando la frecuencia cardíaca). El último ayuda en funciones reparadoras (p. ej. estimulando la digestión). Los estudios que tienen como finalidad identificar patrones de cambios fisiológicos para las distintas emociones regularmente se enfocan en las siguientes variables: ritmo cardíaco, presión sanguínea, resistencia galvánica y temperatura de la piel, transpiración y tensión muscular (Ekman, Levenson y Friesen, 1983; Levenson, Ekman, Heider y Friesen, 1992; Levenson, Ekman, Friesen y Wallace, 1990; Levenson, 1992; Levenson 2003). Otros estudios incluyen actividad de las glándulas lagrimales, salivales y sudoríparas (Matsumoto, 2008). En menor medida, se han estudiado los trastornos en el sistema digestivo asociados con las emociones: como la secreción de jugos gástricos y movimientos en el estómago y tracto digestivo. Estas variables de cambios en el SNA configuran el espacio dentro del cual se han de identificar patrones de cambios fisiológicos específicos para las distintas emociones. Nuestra propuesta de una teoría general de las emociones

⁴⁸ Véase II, v.I.

se compromete con la idea de que no sólo debe existir un caso sólido respecto de las emociones básicas y patrones de cambios fisiológicos, también debe existir evidencia para sostener un caso favorable similar para las emociones cognitivas superiores.

Variedad de respuestas fisiológicas para una sola emoción. Antes de intentar construir un caso tal para las emociones cognitivas superiores, es indispensable responder a ciertas incertidumbres que surgen incluso para el caso de las emociones básicas. A pesar de que el trabajo de Levenson y sus colegas abona bastante a la tesis de que las emociones —básicas— instancian patrones distinguibles de cambios fisiológicos, aún se presentan ciertas dificultades. Por ejemplo, el miedo, dependiendo de la situación desencadenadora, puede activar patrones de cambios fisiológicos distintos que darán como resultado conductas distintas —escapar, pelear, inmovilidad. Además, Larsen, et. al. (2008) aluden a estudios en los que el tipo de estímulo altera el patrón de cambios fisiológicos de una misma emoción. Por ejemplo, cuando se induce miedo y enojo por medio de imágenes, el aumento de presión sanguínea es mayor en el caso del miedo, mientras que, cuando una situación real es la que induce ambas emociones el aumento de presión sanguínea es similar.

En relación al primer planteamiento es posible colegir dos posibles explicaciones. La primera dependería de llevar a sus límites la noción de clase natural *RPH*. En tanto que clase natural *RPH*, una emoción como el miedo, por ejemplo, se definiría —entre otras propiedades— por instanciar un patrón indefinido o indeterminado de cambios fisiológicos. La indefinición o indeterminación nomológica, que es un rasgo regular en las clases *RPH*, estaría en la base de la variabilidad de patrones de cambios fisiológicos que caracterizan a una emoción como el miedo. Además, hay que recordar que los patrones de cambios fisiológicos no son por sí mismos suficientes para individuar una emoción. Hemos propuesto otras propiedades —neuronales, psicológicas, biológicas— que están vinculadas con las propiedades fisiológicas y que, en conjunto instancian un racimo de propiedades causal y homeostáticamente articulado. En este escenario, podríamos insistir en la necesidad de individuar el miedo no sólo sobre la base de un conjunto indeterminado —variado, dentro de ciertos límites nomológicos— de respuestas fisiológicas, sino apelando, también, a otro tipos de propiedades que definen la estructura explicativa del miedo. Por ejemplo, en términos psicológicos, habría que aludir a aquello que representa la emoción: en el caso del

miedo, el objeto formal sería el peligro. En términos expresivo-conductuales, sería necesario considerar la expresión característica del miedo. En términos neuronales, habría que identificar el circuito cerebral que activa los cambios fisiológicos. El racimo causal y homeostáticamente articulado que configuran estas propiedades, junto con la variedad de respuestas fisiológicas típicas del miedo, podrían ayudar a integrar una imagen más comprensiva de dicha emoción, en cuyo contexto resulta menos problemático identificarla. Atender al racimo completo de propiedades que constituyen la estructura explicativa, en este caso, de la emoción del miedo en tanto que clase *RPH*, ayuda a disipar el aire paradójico que queda si sólo se considera la variada e indefinida fisiología de la emoción.

La liberalidad implícita en la noción de clase natural *RPH*, también es útil para construir una respuesta al segundo planteamiento. El margen de indefinición causal que distingue a las clases naturales *RPH* hace esperable que los sistemas fisiológicos —que configuran el patrón de respuestas fisiológicas que caracteriza a cada emoción— se activen dentro de ciertos rangos de intensidad. Es característico de los sistemas biológicos homeostáticos —la regulación de temperatura en un mamífero, por ejemplo— que las variables o subsistemas que los componen regulen y oscilen la intensidad de su actividad a partir de una retroalimentación constante entre el sistema y el entorno. Ya que el racimo de propiedades fisiológicas que forman parte de la estructura explicativa de las emociones, son en sí mismas homeostáticas, es de esperarse que se regulen y oscilen dentro ciertos rangos de activación e intensidad a partir de su retroalimentación con el entorno.

La otra vertiente de respuesta a los planteamientos, aunque no presiona tanto el carácter indeterminado de las clases naturales *RPH*, sí promueve una reforma relevante en nuestras categorías emocionales de sentido común. Para agilizar la exposición, llamaremos a esta posibilidad la *Estrategia de la Ramificación de las Emociones (ERE)*. Según esta tesis, la variedad de respuestas fisiológicas de una emoción se explicaría porque, en realidad, se trataría de distintas *especies* de emociones que pertenecen a un mismo *género* emocional, por decirlo de alguna forma. Según la *ERE*, la familia de las emociones puede tener géneros emocionales que se ramifican en especies emocionales. El miedo, por ejemplo, sería un género emocional que abarcaría, tentativamente, tres especies que se distinguirían, en parte, por sus patrones de respuestas fisiológicas —aquellas que preparan para

escapar, las que hacen lo propio para pelear y las que desencadenan el modo de inmovilidad. Evidencia para apoyar la ramificación del miedo, en este caso, la pueden aportar diversidades similares en relación a otras propiedades del racimo *RPH* que delimita la estructura explicativa del miedo. Por ejemplo, junto con la variedad de respuestas fisiológicas características del miedo, está una variedad de cosas representadas por las tres formas de miedo. Así, las distintas especies de miedo que activan mecanismos fisiológicos de escape, pelea e inmovilidad, representarían *peligro evitable*, *peligro inevitable-enfrentable* y *peligro inevitable no enfrentable*, respectivamente. Aunque el tema general de todas estas especies de miedo sería el peligro, cada una representaría una forma particular y, por lo tanto, sería de mayor valor adaptativo que cada una desplegara mecanismos fisiológicos y conductuales apropiados para cada tipo de peligro. Más aún, una matriz fisiológica distintiva para cada especie de miedo, debe implicar una matriz neuronal distintiva. Y este parece ser el caso. En el apartado sobre las estructuras cerebrales especializadas en la producción de emociones, se indicó que diferentes regiones del **hipotálamo**, la **sustancia gris periacueductal** y la **amígdala**, determinaban el tipo de respuesta fisiológica y, por lo tanto, conductual —huída o pelea— que se activaría en cada caso. En cuanto a la conducta de inmovilidad, se registraba una mayor actividad del **hipocampo**. Entonces, según la *ERE*, lo que mantendría unidos como genero a las distintas especies de miedo serían el *tema evolutivo* que representan (peligro), una expresión facial característica en la mayoría de los casos de miedo, y un conjunto indefinido —pero enmarcado dentro de ciertos confines nomológicos— de patrones fisiológicos y circuitería neuronal —entre otras propiedades. La *ERE*⁴⁹ sugiere que estemos abiertos a la posibilidad de reorganizar la taxonomía de las emociones a partir de una forma alternativa de integrar la evidencia en torno a la diversidad que, en términos de respuestas, presentan distintas emociones en diferentes niveles: fisiológico, neuronal, expresivo-conductual y psicológico.

Las Emociones Cognitivas Superiores instancian patrones de cambios fisiológicos. Sólo recientemente han comenzado a estudiarse los correlatos neuronales y fisiológicos de lo que podríamos llamar Emociones Cognitivas Superiores (ECS). Las ECS que más interés han

⁴⁹ Prinz (2005/2010) ha propuesto una variedad de *TRE*, a partir de consideraciones más bien psicológicas que fisiológicas. Platts (1997) reconoce la ramificación taxonómica como una estrategia ontológica y metodológica sumamente valiosa.

despertado por parte de los investigadores son el orgullo, la vergüenza, la culpa, los celos y el amor. El amor (Esch y Stefano, 2005) da lugar a respuesta a nivel del SNA que involucra las glándulas sudoríparas, aceleración de la frecuencia cardíaca, contracciones intestinales y hasta diarrea. Los celos (Markova, 2010) pueden implicar las glándulas lagrimales, aceleración en el ritmo cardíaco y presión sanguínea y segregación de jugos gástricos. La vergüenza (Clark, 2010) desencadena trastornos fisiológicos dentro de la misma matriz del SNA. De nuevo, es complicado hablar a estas alturas de *patrones* de cambios fisiológicos para las ECS. Incluso en un escenario optimista en el que se lograran identificar patrones más específicos de respuestas fisiológicas para las ECS, es muy probable que estos patrones terminen siendo muy similares a los patrones que se han identificado para otras emociones (Emociones Básicas, por ejemplo). Esto daría lugar al problema que ya hace tiempo había señalado Cannon (1915/1984), y que recientemente han vuelto a detectar en un meta-análisis Larsen et. al. (2008), en relación a la insuficiencia de los patrones de cambios fisiológicos para individuar emociones. La estrategia que sugerimos para enfrentar este problema consiste en recordar que los patrones de cambios fisiológicos que caracterizan a una emoción (Básica o Cognitiva) son apenas un tipo de entre un racimo más heterogéneo, pero homeostáticamente vinculado, de propiedades de distintos niveles que configuran la estructura explicativa de las emociones. En este supuesto, nuestra teoría general de las emociones no pretendería que sólo los patrones de cambios fisiológicos serían suficientes para individuar las emociones. La individuación de emociones, según nuestra teoría, exigiría la integración causal/homeostática de las propiedades fisiológicas con propiedades neuronales, biológicas, psicológicas y hasta fenomenológicas. Inmersos en este marco ontológico y explicativo más amplio es en donde los patrones de cambios fisiológicos de las distintas emociones adquieren un significado explicativo más preciso. Como veremos más adelante, la tipificación de propiedades biológicas, funcionales y psicológicas será de suma relevancia para lograr criterios de individuación más detallados y comprehensivos de las emociones. Pero esto no significa que deberíamos desistir o relegar la investigación empírica de las emociones. Al contrario, tal y como lo estipulamos en nuestra discusión sobre los compromisos metodológicos a los cuales nos ceñimos, la complejidad del fenómeno de las emociones exige de nosotros, además de deferencia a la investigación empírica, actitudes de pluralidad metodológica y ontológica. Nuestra teoría no es una teoría reductivista, es una

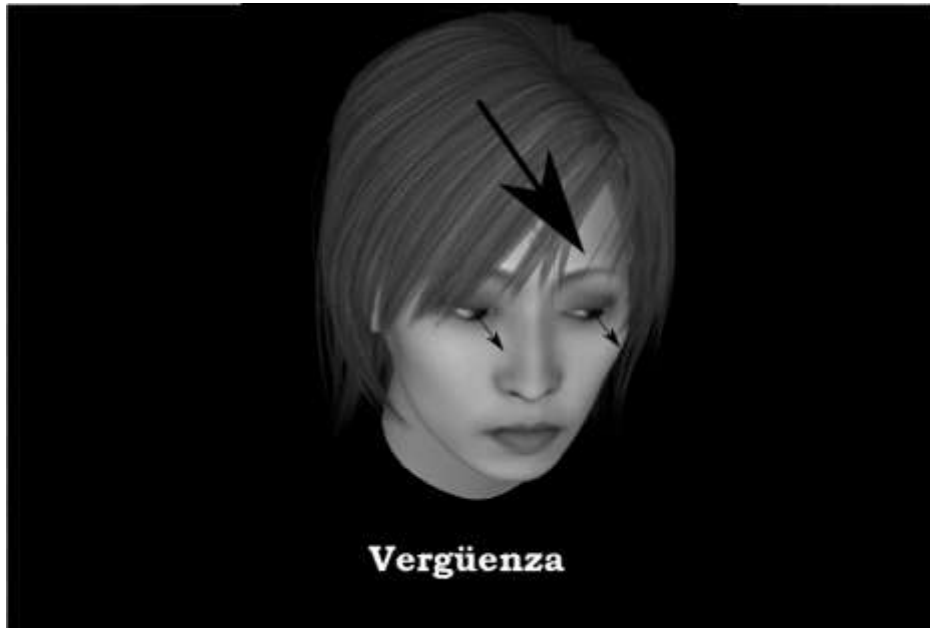
teoría que intenta integrar distintos niveles de explicación con la finalidad de proporcionar una teoría general de las emociones lo más comprensiva y explicativa posible.

Finalmente, ya que es bastante plausible que existan patrones de alteraciones fisiológicas en distintas Emociones Cognitivas Superiores —incluso si no resultaran suficientes para individualizarlas por sí mismas— entonces deja de ser del todo correcta la observación, avanzada por Griffiths (1997), según la cual dichas emociones no instanciarían patrones de cambios fisiológicos.⁵⁰ De hecho, tal y como señala Clark (2010), esta observación en contra de la tesis de que las ECS instancian cambios fisiológicos nunca se fundó en evidencia empírica, simplemente se avanzó sin ningún tipo de prueba o estudio que la sustentara.

3. Propiedades expresivo-conductuales

Antes (II, IV.II) hemos discutido extensamente la tesis según la cual cada emoción básica tendría una expresión facial característica. Si las respuestas que expusimos para hacer frente a las objeciones planteadas a dicha tesis (Russell, 1994) no fueron del todo inapropiadas, se logró establecer un caso favorable para las emociones básicas, por lo menos. La teoría general de las emociones que proponemos tendría que aportar evidencia similar para el caso de las Emociones Cognitivas Superiores (ECS). Recientemente se ha copilado evidencia sugerente que abona para la idea de que las ECS también exhiben respuestas expresivo-conductuales características. Clark (2010) cita varios estudios realizados con participantes de los cinco continentes en los que se muestra que la vergüenza da lugar a conductas como: encogimiento corporal, evasión de mirada, retroceso o alejamiento, andar encorvado (rodillas flexionadas), evitar contacto social y esconderse. Clark incluso llega a sugerir que eventualmente sería posible identificar una expresión posturo-facial característica.

⁵⁰ Véase II, v.I.



En cuanto al orgullo, estudios recientes (Tracy, 2010; Tracy y Robins, 2004, 2007) muestran que está asociado fiablemente con expresiones reconocibles que pueden considerarse como innatas y universales. La expresión prototípica del orgullo incluye el cuerpo en postura erguida —expansión de la caja torácica—, cabeza inclinada ligeramente hacia atrás, brazos en posición de jarras apoyados en las caderas o alzados en forma de ‘v’ con los puños apretados y una sonrisa incipiente en el rostro.



Orgullo: expresión A



Orgullo: expresión B

La expresión de orgullo es fiablemente distinguida de emociones similares (como la alegría o la excitación) por individuos desde Estados Unidos de América e Italia, hasta habitantes de comunidades analfabetas aisladas de Burkina Faso. Los índices de reconocimiento de la expresión de orgullo a partir de fotografías e imágenes van del 80 al 90 %: porcentajes muy similares a los que se registran para las reconocidas emociones básicas (Tracy & Robins 2008). La conducta expresivo-conductual característica de orgullo se despliega de forma espontánea y rápida en situaciones apropiadas (éxito o preeminencia). Se ha observado en niños desde los tres años, adolescentes, ciegos de nacimiento y atletas de 80 nacionalidades durante los juegos olímpicos. Tracy (2010) considera que la diversidad de culturas que exhiben y reconocen las expresiones, aunado al hecho de que personas invidentes de nacimiento también las despliegan (naturalmente, sin haberlas observado), sugiera que son universales.

Menos se ha investigado en torno a respuestas expresivo-conductuales de otras ECS. Del amor romántico (Bartels y Zeki, 2000) se sabe que promueve conductas de aproximación, imitación de gestos y acicalamiento. También se sospecha de un tipo de sonrisa característica que implicaría —a diferencia de la sonrisa típica de la alegría— dilatación de las pupilas. A los celos (Prinz, 2004) se los asocia con expresiones faciales típicas del enojo, el asco y, en menor medida, la tristeza. También hay evidencia que indica que los celos activan conductas de búsqueda, confrontación y pelea (Panksepp, 1998). Que algunas ECS, cuando involucran expresiones y conductas características, parezcan instanciar las que se reconocen como típicas de Emociones Básicas ha llevado a algunos a pensar que, en realidad, las ECS son combinaciones de EB (Plutchik, 2001 y Prinz, 2004). Esta posibilidad no es del todo descabellada. Después de todo, no es sólo el caso que las ECS instancian algunas de las propiedades expresivo-conductuales de las emociones básicas —los celos parecen tomar prestada su expresión al enojo, el asco y la tristeza; mientras que el orgullo y el amor a la alegría—, también parecen utilizar sus mismos recursos fisiológicos y neuronales. No descartamos *a priori* la posibilidad de que algunas ECS pudieran ser, en realidad, compuestos de emociones más básicas. No obstante, consideramos que antes de explorar la posibilidad de que una ECS sea una mezcla de emociones básicas, es indispensable considerar las propiedades biológicas, funcionales y psicológicas de la ECS en cuestión, ya que éstas podrían sugerir que se trata, en realidad, de una emoción con una historia evolutiva y

funciones adaptativas y representacionales distintas a las que tienen las emociones básicas que supuestamente la constituyen. Así, aunque por el momento no descartamos la tesis de que las ECS sean mezclas de EBs, posponemos nuestro veredicto debido a que, en este momento, no tenemos los datos necesarios para decidirlo.

II. PROPIEDADES BIOLÓGICAS.

4. Función adaptativa de las emociones.

Cada emoción, en tanto que involucra un *patrón de cambios en el organismo*, constituye una respuesta adaptativa del organismo —evolutivamente diseñada⁵¹—, para enfrentar situaciones evolutivamente relevantes. Lo que se considera como situación de relevancia evolutiva es un escenario en el que la sobrevivencia del individuo (de su descendencia o linaje) y posibilidades reproductivas están en juego (*fitness*). Las situaciones de este tipo, para el caso de muchos mamíferos (Panksepp, 1998), engloban relaciones del organismo con el entorno **natural** y **social**. Por ejemplo, el miedo que sobreviene al estar al borde de un acantilado y la vergüenza que se experimenta al caer en público, ejemplifican, respectivamente, cada especie de relación.

Griffiths (1997)⁵² está en contra de la idea de que las emociones —básicas y cognitivas— compartan una misma historia evolutiva y funciones adaptativas, por lo tanto, en contra de la idea de que conformen un conjunto homogéneo de fenómenos psicológicos o biológicos que se puedan explicar a partir de una misma teoría. Griffiths está de acuerdo en que las emociones básicas y las ECS son respuestas adaptativas. También concede que tienen en común el ser *patrones de motivación irruptoras*, lo que significa que ambas son mecanismos por medio de los cuales se puede interrumpir una tarea y reorientar la conducta como resultado de alguna propiedad (exigencia) detectada en el entorno (Griffiths, *op. cit.* pp. 115-122). No obstante, advierte diferencias cruciales entre las emociones básicas y las ECS. La función evolutiva de las últimas es bastante específica: *resolver problemas de compromiso* (Frank, 1988). Por ejemplo, ser capaz de sentir culpa, vergüenza y amor, garantizan que puedes asumir un compromiso y que será complicado no honrarlo. Las ECS

⁵¹ El que las emociones sean respuestas adaptativas evolutivamente diseñadas sugiere que se trata de mecanismos innatos y, muchas de ellas, universales en muchas especies.

⁵² Véase también, II, v.i.

tienen una función colectiva en tanto que existen para favorecer vínculos y compromisos entre individuos. Así, mientras que las emociones básicas son programas rápidos, modulares y con contrapartidas *homólogas* en otras especies, las ECS son programas que interactúan con creencias, tienen una función social y no parecen encontrarse en otras especies.

En este apartado no trataremos la cuestión de si las ECS son rápidas y modulares. Estas propiedades tienen más que ver con la arquitectura de las emociones y, por lo tanto, las preguntas que suscitan en relación a las ECS las abordaremos más adelante. La observación de Griffiths a la que intentaremos dar una respuesta es a la que sugiere que las ECS y las Emociones Básicas no comparten la misma historia evolutiva, ni las mismas funciones adaptativas, por el simple hecho de que las últimas tienen homólogas en otras especies y las primeras no. Si nuestra respuesta a esta objeción es satisfactoria, entonces habríamos anulado una razón más para pensar que la tesis de la desunión de las emociones representa el mejor camino a seguir.

Las ECS como formas homólogas de Emociones Básicas. La idea de que las ECS y las EB comparten una misma historia evolutiva ha sido desarrollada de forma novedosa por Jason Clark (2010 y 2012). Clark piensa que muchas de las que se consideran Emociones Cognitivas Superiores paradigmáticas tienen, en contra de lo que se suele pensar, contrapartidas homólogas en especies emparentadas. Es decir, desde el enfoque que ofrece Clark, emociones como la vergüenza, el orgullo y la culpa humanas, entre otras candidatas, tendrían advocaciones homólogas en otras especies y, por lo tanto, no serían resultado de procesos evolutivos específicos y exclusivos del ser humano. De ser así, tendríamos que comenzar a pensar en la posibilidad de rastrear las homólogas Básicas de ECS. Esto significaría que la indagación empírica tendría que conducirnos, en algún momento, a descubrir, por ejemplo, la vergüenza o el orgullo en sus versiones Básicas, a las que podríamos denominar Vergüenza-B u Orgullo-B, para distinguirlas de sus derivados Cognitivos Superiores. Clark considera e intenta probar que ya hay evidencia para probar su tesis y que, con ello, mantendría unidas en el mismo dominio teórico a las EB y las ECS.

Si es posible probar que existen contrapartidas homólogas de EB para algunas ECS, entonces tendríamos razones para dudar de la tesis de la desunión de las emociones. Deberíamos hablar, más bien, de cierta continuidad y parentesco evolutivo entre EB y ECS, lo

que haría plausible seguir hablando de un solo dominio ontológico y epistemológico (de una sola matriz teórica y explicativa) en virtud de que comparten las mismas propiedades evolutivas.

La vergüenza como una emoción básica. Clark (2010) intenta abonar a las ideas anteriores ofreciendo un estudio de caso de una ECS particular: a saber, la **vergüenza**. Clark señala que la vergüenza no debe considerarse una ECS pura (esto es, que no tiene una EB como correlato). La vergüenza, en tanto que ECS, tiene una EB como correlato (homóloga) y comparte semejanzas (mecanismos) causales profundas con ella, por lo que sería adecuado agruparlas dentro de una misma categoría psicológica (guiándonos por el criterio de clasificación basado en la homología).⁵³

La vergüenza, en su forma actual en los seres humanos, parece derivarse de una emoción que expresa/comunica relaciones de rango o estatus en mamíferos no humanos. Su función adaptativa primaria consiste en indicar subordinación frente a individuos dominantes con la finalidad de apaciguarlos. Puede activarse, en individuos de rango inferior ante la mera presencia de uno dominante. Es importante notar que la situación evolutivamente relevante que activa la vergüenza remite a propiedades del entorno social. Clark, además, cita estudios que muestran que esta advocación no cognitiva de la vergüenza incluye (a) patrones distintivos de respuestas conductuales y expresivas como: encogimiento corporal, evasión de mirada, retroceso o alejamiento, andar encorvado (rodillas flexionadas), evitar contacto social y esconderse; (b) cambios a nivel fisiológico que incluyen variaciones específicas en lo que se denomina *eje hipotalámico-pituitario-adrenal* (HPA) y el sistema inmune, entre otros; (c) patrones característicos de actividad neural; y (d) secuelas afectivas y cognitivas como sentirse pequeño, evitar ser visto por otros y (en tanto que estado motivador) buscar rutas de acción que reditúen en no perder rango social, conservarlo o mejorarlo. Clark considera que esto es sintomático de que la vergüenza no es una ECS pura. Tiene una EB homóloga como correlato.⁵⁴

⁵³ Una ventaja adicional del estudio de Clark consiste en que proporciona más evidencia a la idea de que las ECS tienen muchas de las propiedades que los defensores de la tesis de la desunión les niegan; a saber, respuestas fisiológicas, conductuales y expresivas características.

⁵⁴ Esta evidencia ha llevado al propio Griffiths (Griffiths y Scarantino, 2009) a incluir en su lista de EB a la vergüenza.

La vergüenza como una emoción cognitiva superior. En la medida en que las relaciones de estatus y rango siguen teniendo un papel importante en el ser humano, la forma humana de vergüenza (ECS) retiene muchas de las características y funciones adaptativas de la vergüenza en su modalidad básica. No obstante, es indispensable señalar que su función se ha expandido en dos sentidos. En primer lugar, si bien las relaciones de dominancia siguen siendo fundamentales en las sociedades humanas, se ha pasado de la dominancia entendida como ‘poder coercitivo sobre los demás’, a las ‘competencias por prestigio entre los demás’, la cual involucra la atracción de atención social positiva y la posesión de rasgos u objetos socialmente valorados. En este sentido, la vergüenza se asocia con las derrotas en estas competencias de prestigio. En segundo lugar, siguiendo a Frank (1988), Clark considera que la función de la vergüenza es resolver problemas de compromiso. Esto es, ayuda a facilitar la cooperación en la medida que señala la capacidad de detectar nuestras violaciones a normas o convenios sociales y, en la misma dirección, puede motivar nuestro deseo de adherirnos a dichas normas. A pesar de que ha ampliado sus funciones (responde a estímulos noveles) la vergüenza humana (ECS), sigue exhibiendo las mismas respuestas conductuales, expresivas y fisiológicas características de la vergüenza como emoción básica. Individuos que experimentan vergüenza en situaciones en las que su prestigio se deteriora o son sorprendidas violando un convenio, exhiben patrones expresivos y de conducta característicos de la homóloga básica, y probablemente también los mismos cambios fisiológicos.

Hasta este punto, las ideas de Clark apuntan a la existencia de dos formas de vergüenza, una Básica y otra Cognitiva Superior, que se distinguen por su función adaptativa. Mientras que la primera forma básica es útil —entre otras cosas— para indicar relaciones de dominancia en términos coercitivos, la forma cognitiva superior preserva el mismo tema evolutivo —indicar relaciones de dominancia— pero en relación a ‘competencias de prestigio’. La vergüenza ECS, en este sentido, se presentaría en situaciones en las que hemos sufrido un revés en dichas competencias de prestigio. Otra diferencia de función adaptativa entre la vergüenza-B y la vergüenza-CS, según Clark, consistiría en que la vergüenza-CS serviría para resolver *problemas de compromiso*, en el sentido descrito por Frank (1988). No obstante estas diferencias funcionales de detalle, las dos formas de vergüenza siguen

compartiendo una misma función adaptativa general y los mismos mecanismos causales fisiológicos y neurológicos.

El caso a favor de la homologación de ECS y Emociones Básicas. Según Clark, las consideraciones precedentes muestran que la ‘vergüenza-CS’ tiene un vínculo profundo y sistemático con la ‘vergüenza-EB’. La relación que comparten sugiere que son homólogas, según los criterios operacionales que Ereshefsky (2007) y Love (2007) han propuesto para identificar y probar relaciones de homología entre rasgos estructurales de nivel superior. Los criterios para juzgar la homología son los siguientes:

1. *Posición*: la colocación espacio-temporal o rol funcional que desempeña dentro de un nivel de organización más general.
2. *Cualidad especial*: la complejidad, particularidad o especialización del rasgo.
3. *Continuidad*: la identificación de un continuo de propiedades evolutivas; desde su estado más simple y primitivo a un estado más complejo y derivado del mismo rasgo.

Las observaciones de Clark sugieren que la vergüenza-CS satisface los criterios respecto de la vergüenza-B. En cuanto al criterio 2 —*cualidades especiales*—, la vergüenza-CS preserva los rasgos presentes en la vergüenza-B: su complejidad en términos neurofisiológicos, funcionales y adaptativos; y su especialización y particularidad en tanto que mecanismo psicológico diseñado evolutivamente para enfrentar un tipo específico de reto adaptativo utilizando, aproximadamente, los mismo recursos. En cuanto a 1, la *posición*, tanto la vergüenza-CS, como la vergüenza-B se ubican, aproximadamente, en la misma posición funcional dentro de un nivel de organización más general. Si bien las funciones de la ‘vergüenza-CS’ (señalar derrotas en competencias de prestigio, etc.) se han expandido y van más allá de las que desempeña la ‘vergüenza EB’ (señalar relaciones de rango y dominancia), el tema general de ambas sigue siendo el mismo: establecer nuestra ubicación en el organigrama social y motivar conductas que favorezcan una mejor colocación⁵⁵. Esto significa que tienen la misma *posición*. Sobre 3, la *continuidad*, hay que comenzar a reconocer que la ‘vergüenza-CS’ no surge de una vez y súbitamente en los seres humanos. Los mamíferos

⁵⁵ Incluso la persona avergonzada que se retrae y evita por un tiempo el contacto con los demás, pensará en acciones y conductas que lo ayuden a superar el revés social que sufrió su imagen. Estas estrategias para enfrentar la situación son motivadas por la vergüenza (Cosmides y Tooby, 2008).

superiores con mayores capacidades cognitivas (primates) parecen mostrar respuestas afectivas más complejas que las que se pueden observar en ratas, por ejemplo.

Clark menciona otro factor, no funcional, que respalda la homología. Se trata del carácter vestigial de algunas respuestas involucradas en la ‘vergüenza-CS’ que se derivan del correlato básico. Tales respuestas pueden llegar a representar desventajas, más que ventajas adaptativas. Por ejemplo, las respuestas expresivas y conductuales de vergüenza pueden deteriorar más aún nuestra imagen social en una situación desfavorable de tipo ‘competencia por prestigio’. En este sentido, Clark señala, además, que es posible que la homología se presente de forma individual o selectiva, en estructuras organizacionales complejas, involucrando, a veces, sólo algunos mecanismos. Esto implica que la evolución de rasgos complejos estructurales no necesariamente involucra a todos los mecanismos/componentes que los constituyen. La evolución de un rasgo, de nuevo, puede tener lugar de forma parcial o mixta. Este hecho, conduce a la necesidad de reconsiderar nuestras afirmaciones en torno a la homología o no-homología entre dos rasgos complejos. La realidad, puede ser más complicada y requerir de investigaciones que discriminen los distintos componentes de un rasgo complejo para detectar si todos han evolucionado o sólo algunos. En el caso de las emociones, esto significa que las Emociones Cognitivas Superiores podrían haber evolucionado (ser homólogas) de Emociones Básicas en virtud de haber modificado sólo parcialmente el conjunto de propiedades neuronales, fisiológicas, expresivo-conductuales, funcionales o psicológicas que caracterizaban a la Emoción Básica homóloga original.

Ahora bien, la homología entre vergüenza-B y vergüenza-CS se puede establecer con relativa seguridad en los niveles expresivo-conductual, fisiológico y social-adaptativo —la naturaleza de la relación social para la cual evolucionaron las emociones homólogas. Pero es más complicado establecer la homología en el nivel neuronal. Una posibilidad consistiría en reconocer que la *encefalización* (aquí entendida como la tendencia de algunas especies a que aumente el tamaño de su cerebro), por lo regular, implica la expansión o migración de mecanismos que antes sólo ocupaban regiones inferiores del cerebro, hacia regiones superiores. Esto da lugar a la re-escritura, a nivel neurofisiológico, de los mecanismos primitivos. Así es como se puede entender la homología en el nivel neuronal. Esta sugerencia para establecer la homología en términos neuronales de las ECS con sus formas

Básicas, es consistente con el hecho de que las ECS involucran procesos cognitivos más complejos, anidados en regiones más modernas —corticales— del cerebro. En este sentido, el que las ECS impliquen aspectos cognitivos superiores, ausentes en sus contraparte Básicas —en virtud de que ocupan estructuras corticales que tienen bajo control dichos aspectos— no debe representar un razón de peso para pensar que las ECS configuran un dominio teórico-explicativo distinto al resto de las emociones básicas⁵⁶. Se trata, acaso, de un rasgo menor de divergencia entre las ECS y las Emociones Básicas que no es suficiente para promover la desunión de las emociones. El peso que aportan el resto de propiedades que comparten ECS y Emociones Básicas —historia evolutiva común, propiedades neurofisiológicas, expresivo conductuales, funcional/adaptativas y, como veremos más adelante, psicológicas, de arquitectura y fenomenológicas— nos debe impulsar a mantenerlas juntas, bajo el mismo marco teórico-explicativo.

Homología de 'rasgo individual' y homología serial. Pero la homología alegada en el caso de la vergüenza-B y la vergüenza-CS aún deja preguntas pendientes. ¿Es la vergüenza-CS la misma emoción que la vergüenza-B, sólo que evolucionada? ¿O es la vergüenza-CS una emoción distinta que evolucionó a partir de la vergüenza-B? En el primer caso se trataría de homología de '*rasgo individual*': la emoción de la vergüenza ha sufrido una transformación evolutiva que va de su advocación EB a la más compleja ECS. En tal caso no sería correcto decir que los humanos tengamos dos emociones de vergüenza. Sólo tenemos una emoción de vergüenza-CS heredada de nuestros ancestros. Una respuesta afirmativa a la segunda cuestión apuntaría a lo que se conoce como *homología serial*. Según la *homología serial* las dos formas de vergüenza son homólogas en tanto que la versión Cognitiva Superior es una especie de réplica no exacta de su correlato Básico, con funciones y respuestas muy similares, y útiles, para contextos ligeramente distintos. A esta característica se la conoce como 'heterotropía' o 'heterocronía'. Ejemplos de este tipo de homología son las vértebras de la espina dorsal, los segmentos en el cuerpo de los artrópodos y, en el caso de la conducta en animales como las aves, los rituales de cortejo.⁵⁷ Si la vergüenza-B y la vergüenza-CS ejemplifican un caso de *homología serial* entonces es posible pensar que ambas emociones

⁵⁶ Solomon (2007) presenta esta objeción y se expone en II, v.I.

⁵⁷ Picotear originalmente les sirvió a las aves para encontrar comida, luego para comunicar a una hembra que había encontrado comida y, finalmente, es útil sólo como una conducta de cortejo disociada de la búsqueda de alimento.

co-existen en nuestro acervo efectivo, del mismo modo en que la tercera y cuarta vértebra co-existen en nuestra espina dorsal.

Clark favorece la hipótesis de la homología serial. En primer lugar, porque este tipo de homología es muy común en el campo de la conducta. En segundo lugar, las respuestas fisiológicas, conductuales y expresivas de la vergüenza-B, exhiben la capacidad de ser útiles en contextos noveles (heterocronía) relacionados o parecidos a los que originalmente desencadenaban la emoción básica. Esto es, la vergüenza-B parece seguir siendo útil a los seres humanos: piénsese en la utilidad que tendría en ciertas situaciones (como cuando nos confronta un superior) experimentar vergüenza-B, en la medida en que involucra mostrar sumisión. Por otra parte, si las emociones en cuestión son homólogas seriales, ayudaría a explicar la mayor participación de procesos cognitivos por parte de la forma CS, ya que su *posición* (relaciones con otros rasgos psicológicos) debe ser distinta de la que tiene la forma Básica con dichos rasgos. Una ventaja que identifica Clark en esta idea, es que parece respetar la intuición de que las EB y las ECS son en algún sentido independientes, sin que sean en absoluto fenómenos diferentes, en la medida en que hay un conjunto mecanismos causales que comparten. Mientras las ECS son desarrollos ulteriores, las EB siguen funcionando como lo hacen en otras especies, lo cual no es nada desventajoso. Siempre es útil contar con mecanismos de respuesta rápidos e innatos para enfrentar retos recurrentes en el entorno.

Proyección de la homología serial a otras ECS. La posibilidad de que otras ECS tengan homólogas Básicas ha sido explorada por el mismo Clark para el caso del orgullo (Clark, 2012). Clark cita evidencia (Tracy, 2010) a favor de la existencia de dos formas de orgullo que comparten mecanismos causales neuronales, fisiológicos y expresivo conductuales, pero que difieren ligeramente en sus funciones adaptativas. La forma básica de orgullo, conocida como *orgullo-hubrístico* tendría la función de expresar cierta preeminencia o dominancia social basada en la fuerza y la coerción; mientras que la forma CS, *orgullo-auténtico*, tendría una función similar sólo que basada en el *prestigio*. Clark alude a estudios recientes que sugieren la posibilidad de identificar relaciones de homología entre ECS —como la culpa, los celos, el placer estético, la repugnancia moral— y contrapartidas más básicas. Por ejemplo, el estudio de Markova, et. al. (2010) sugiere que la emoción de los

celos en adultos puede derivarse de cierta respuesta emocional en niños desde un año ante la exclusión o desatención por parte de sus protectores habituales.

Implicaciones de la homologación de ECS con Emociones Básicas. La estrategia que propone Clark, en la medida que funcione para otras ECS, representa una seria objeción en contra de la idea de que las ECS configuran un dominio ontológico y explicativo distinto al de las Emociones Básicas⁵⁸. En un sentido positivo, la homología entre ECS y Emociones Básicas abona a la causa de una teoría unificada y general de las emociones...o, al menos, de mucho de lo que llamamos emociones. En realidad, las ECS formarían y participarían de la misma historia evolutiva —y, por lo tanto, del mismo dominio teórico-explicativo— que sus contrapartidas básicas, y esto explicaría por qué utilizan, con ciertas modificaciones funcionales, los mismos recursos neurofisiológicos y expresivo-conductuales. El reconocimiento de lo anterior, restaría también peso a la objeción en contra de la unión de las emociones que apunta a la aparición tardía de las ECS en relación de las EB⁵⁹. El rezago, finalmente, puede entenderse por el hecho de que evolucionaron más tarde, *a partir de los mismos recursos neurofisiológicos y con la misma función adaptativa*. En otras palabras, el rezago ontogenético de las ECS no obedece a una diferencia filogenética en relación a las EB. Por otro lado, la estrategia de la homologación de las emociones nos obliga a replantear la taxonomía de las emociones básicas y, seguramente, a revisar nuestras categorías emocionales habituales. Pero esta implicación no representa en realidad una objeción a nuestra teoría general de las emociones. Muchas teorías nuevas, mientras se abren paso, tienden a exigir ajustes ulteriores en el léxico vernáculo.

Finalmente, hay que dejar claro que la estrategia de la homologación de las emociones constituye un enfoque que ayuda a salvaguardar la unidad de las emociones. En este sentido, aporta sustento a nuestra definición de las propiedades biológicas que atribuimos a todo lo que haya de ser considerado como una ‘emoción’. En términos de sus propiedades biológicas, ya sea Básica o Cognitiva Superior, una emoción se entiende como una respuesta adaptativa del organismo —evolutivamente diseñada— ante situaciones relevantes en términos evolutivos —que involucren su sobrevivencia y/o la de su linaje y oportunidades reproductivas. Las situaciones de este tipo, en contra de lo que los objetores de una teoría

⁵⁸ Véase II, v.I, observación 5.

⁵⁹ Véase II, v.I, observación 4.

general de emociones alegan, engloban relaciones del organismo, no sólo con su entorno **natural**, sino también con su entorno **social** —como lo muestra la advocación básica de la vergüenza y, muy probablemente, las versiones básicas de otras ECS.

III. PROPIEDADES PSICOLÓGICO/FUNCIONALES.

5. Carácter representacional de las emociones

Las emociones consiguen responder con mayor eficiencia a las situaciones evolutivamente relevantes en la medida en que ellas mismas son *representaciones* (Charland, 2002). Pero, ¿cómo desempeñan su papel representacional las emociones? Para responder a esta cuestión, seguiremos a Prinz (2003a; 2003b; y 2004) y su caracterización de las emociones en tanto que *valoraciones corporeizadas*. Esta propuesta se basa en la teoría co-variacional del contenido, según la cual, un estado mental obtiene su contenido representacional en virtud de (1) que es fiablemente causado por x , y (2) que existe (evoluciona o se adquirió), precisamente, para detectar x . Un requisito adicional —derivado de (1)— consiste en que (3) sea posible el error: esto es, que algunas veces el estado mental no sea causado por aquello que fiablemente lo causa. Por ejemplo, mi concepto de GATO representa a los *gatos* en virtud de que las instanciaciones de dicho concepto en mi pensamiento han sido causadas fiablemente, bajo ciertas condiciones, por la presencia de *gatos* y sólo por la presencia de *gatos*. Además, existe, precisamente, para eso: el concepto se originó (en este caso) para detectar y activarse ante la presencia de *gatos*. Finalmente, también es posible que el concepto se active de forma errónea (3): por ejemplo, el concepto se puede activar una noche oscura al percibir un mapache. Así, se piensa que algunos estados mentales adquieren su contenido en virtud de alguna relación causal con el entorno (Prinz, 2004, p. 54).

En la misma dirección, Prinz —quien también acepta que cada emoción involucra un conjunto de perturbaciones fisiológicas— sostiene que los cambios corporales involucrados en cada emoción *representan* algo, en el sentido estipulado por la teoría co-variacional. La idea es que las emociones son *fiablemente causadas* por la presencia de ciertas propiedades presentes en situaciones recurrentes. Además, también parecen existir para detectar, precisamente, dichas propiedades; y, finalmente, pueden hacerlo de forma errónea. Considérese el miedo. Lo experimentamos al percibir un determinado *patrón* de

cambios corporales. Un patrón de cambios fisiológicos de este tipo es activado cuando, por ejemplo, el sistema auditivo detecta un ruido fuerte y repentino o cuando el sistema visual detecta un objeto largo y enroscado en una esquina, etc. Los cambios fisiológicos *representan* peligro porque son causados fiablemente por instancias que exhiben la propiedad ‘*peligro*’. El peligro es la propiedad en virtud de la cual dichas condiciones de activación (el ruido súbito y fuerte, y el ver un objeto largo y enroscado) provocan la perturbación de nuestros cuerpos. Además, es razonable pensar que la perturbación fisiológica característica del miedo, en tanto que representación de peligro, existe precisamente para eso; para representar peligro. Finalmente, es posible que se active de forma errónea o que no se active: por ejemplo, puede activarse al percibir un cable enroscado en una esquina, o no activarse ante una situación que implica peligro real, como estar en un lugar con niveles de radiación muy elevados. Así, para Prinz, las emociones son representaciones de propiedades, no porque impliquen juicios o valoraciones en formato conceptual-proposicional, lo son porque involucran *valoraciones corporeizadas*: el cuerpo, por sí mismo *representa* propiedades del entorno.

¿En qué consiste el contenido representacional de las emociones? Bueno, el contenido representacional de las emociones es bastante general: es sobre *propiedades generales* u *objetos formales*. Si distinguimos (como Kenny, 1963) entre un objeto formal y otro particular, las emociones sólo representan el objeto formal. Por ejemplo, en la experiencia de miedo que nos embarga al estar al borde de un precipicio, los cambios fisiológicos no representan la situación u objeto particular ‘estar al borde de un precipicio’, representan sólo una propiedad general de dicha situación; a saber, su *peligrosidad*, la cual podemos entender aquí como el objeto formal de la emoción. Esto tiene como consecuencia que las emociones ejemplifican un tipo de contenido representacional bastante general (representan *objetos formales*⁶⁰), y no tienen, de hecho, objetos particulares. Prinz piensa que se trata de una confusión común, incluir al objeto particular como parte constitutiva de la emoción, cuando en realidad es sólo su causa: ‘estar al borde de un precipicio’ incluye la propiedad que causa y que representan los cambios fisiológicos característicos de la emoción del mie-

⁶⁰ Los objetos formales o temas relacionales son rasgos bastante generales que exhiben ciertas situaciones u objetos *desde* la perspectiva del sujeto. En tanto que generales, pueden ser compartidos por cosas muy distintas. Una serpiente en nuestro camino y un auto a toda velocidad hacia nosotros, pudieran tener muy poco en común, excepto por el hecho de que ambos parecen (a nuestro cuerpo) peligrosos desde nuestra perspectiva.

do, pero la situación misma no es representada por el cuerpo. La situación de ‘estar al borde del precipicio’ puede ser representada perceptualmente o proposicionalmente (por medio de un juicio), pero no se representa por medio de un estado fisiológico.

Expuesta de otra manera, la idea es que la forma en que reportamos nuestras emociones, nos lleva a pensar que el carácter representacional de la emoción abarca, no sólo al objeto formal, sino también al objeto particular. Considérese el reporte emocional:

i) Estar triste porque la niña ha muerto.

El contenido representacional de (i) suele interpretarse (en la tradición cognitivista) en términos del modelo de las actitudes proposicionales. ‘Estar triste’ es una actitud psicológica dirigida hacia el hecho de ‘que la niña ha muerto’, objeto particular. Y se agrega que el objeto particular ejemplifica el objeto formal de la tristeza; a saber, la *pérdida*. Entonces, se termina asumiendo que las emociones involucran dos niveles intencionales: el del objeto particular y el del objeto formal (anidado en el primero).

La teoría del carácter representacional de las emociones que propone Prinz lo facilita para rechazar esta interpretación. Prinz acepta que las emociones pueden involucrar dos niveles de intencionalidad (un objeto formal y otro particular), pero especifica que la emoción, en sí misma, sólo es intencional en la medida en que *representa* al objeto formal. El objeto particular, remite a un estado mental distinto (percepción o juicio) que lo representa. Así, el análisis correcto reconocería que (i) es, de hecho, un estado mental complejo que involucra dos estados mentales. Por un lado, está el objeto particular ‘que ha muerto la niña’, y que se representa por medio de una percepción o un juicio. Y, por otro lado, tenemos al ‘estar triste’ que está representado por un patrón de cambios fisiológicos —la emoción misma—, el cual es causado por una propiedad detectada en el objeto particular.

Activadores, recalibración y la unidad de las emociones. Recapitulando, Prinz considera que las emociones son *valoraciones corporeizadas*. Esto es, cada emoción se define en términos de la percepción de un conjunto de cambios fisiológicos que, a su vez, *representa* una propiedad general (objeto formal) instanciada por una situación (objeto particular). La teoría del carácter representacional de las emociones de Prinz, por lo tanto, proporciona más razones para mantener unidas en un dominio teórico-explicativos a las emociones

¿Cómo, entonces, explica el hecho de que algunas emociones parezcan ser reacciones compulsivas —emociones básicas— y otras parezcan ser el resultado de discriminaciones y valoraciones altamente cognitivas? De hecho, instancias particulares de un mismo tipo emocional, exhiben esta diferencia. Considérese el miedo que resulta de ver una serpiente y el que resulta de ver los resultados electorales. ¿Por qué tendríamos que seguir hablando de un solo tipo de fenómeno afectivo dadas estas diferencias?

Según Prinz, la causa de la aparente disparidad entre ECS y Emociones Básicas tiene que ver, más que con la naturaleza de las emociones mismas, con el tipo de *mecanismo representacional* que las activa: puede ser (a) una percepción, (b) una asociación adquirida entre un rasgo del entorno y una experiencia afectiva agradable o desagradable, o (c) un juicio o valoración articulada en términos proposicionales/conceptuales (Prinz, 2004, pp.74-77). Para entender cómo estas distinciones entre mecanismos representacionales explican la aparente diferencia entre fenómenos emocionales, es necesario explicitar algunos detalles adicionales de la arquitectura psicológica que Prinz le atribuye a las emociones.

Prinz considera que nacemos con una especie de carpeta mental (para cada emoción) que contiene los rasgos distintivos del entorno que, al ser percibidos, activan la respuesta emocional. Así, para la emoción del miedo, rasgos como ‘tener la apariencia de un objeto largo y enroscado’, ‘oscuridad’, ‘pérdida súbita de la estabilidad’ o ‘una gran altura’, serían algunos de los elementos archivados (de fábrica o de forma innata) en la carpeta de los *activadores* del miedo que, al ser detectados en el entorno por nuestro sistema perceptual, desencadenarían la emoción. Los miedos compulsivos, como el que se desencadena y experimenta al ver una serpiente, tienen como *activadores* rasgos que se encuentran de forma innata en la carpeta ‘activadores de miedo’, y que son *percibidos* en el entorno. Aquí, el *mecanismo representacional* que causa la emoción —el cambio fisiológico percibido— es (a) una *percepción*. Ahora bien, las carpetas de activadores de cada emoción pueden recalibrarse: agregar nuevos *activadores* y eliminar otros. Una forma de agregar elementos a una carpeta de activadores emocionales, es mediante la asociación de un rasgo del entorno con una experiencia afectiva agradable o desagradable. Para seguir con el mismo ejemplo, el miedo a las habitaciones pintadas de color amarillo se pudo haber adquirido después de haber pasado, durante la infancia, penosos días de enfermedad y agonía en una habitación

amarilla. En este caso, el activador también es percibido, pero no se trata de un rasgo incluido de nacimiento en la carpeta, se ha agregado mediante (b) una asociación entre dicho rasgo y una experiencia afectiva desagradable⁶¹. Finalmente, con la adquisición del lenguaje, es posible conceptualizar el objeto formal que ejemplifican los *activadores* de cada emoción. *Peligro*, en el caso del miedo. Esto hace posible que, al considerar situaciones en términos (abstractos) de juicios y creencias —como la de que ‘ganará las lecciones un candidato con tendencias fascistas’— se extraigan algunas implicaciones a las cuales podríamos juzgar como exhibiendo el objeto formal de la emoción del miedo: *peligro*. En estos casos, el mecanismo representacional por medio del cual se agrega un elemento a la carpeta de *activadores* de miedo, y que desencadena la emoción es (c) un juicio o valoración en términos proposicionales⁶².

Dispuestas así las nociones, podemos ver por qué Prinz considera que hay que mantener las emociones unidas. No existe una diferencia profunda —en términos psicológicos— entre el miedo a la serpiente, a las habitaciones amarillas y a que gane las elecciones un candidato fascista. En tanto que miedos, los tres implicarían el mismo patrón de cambios fisiológicos, el cual *representaría* el mismo objeto formal (el *peligro*), detectado, mediante diferentes mecanismos de representación, en diferentes situaciones. Los tres casos de miedo difieren en un aspecto extrínseco —no intrínsecamente emocional: a saber, distintos *mecanismos representacionales* los desencadenan. En conclusión, las emociones configuran un mismo tipo de fenómeno. Son intencionales —en un sentido general— en tanto que son

⁶¹ El trabajo de Öhman (1993) ofrece evidencia a favor del enfoque de *activadores* innatos y de la noción de *calibración* de las carpetas que almacenan dichos *activadores*. Öhman muestra que, mientras es muy sencillo lograr que un macaco desarrolle una respuesta de miedo ante la presencia de serpientes, arañas y precipicios, se requiere de un largo condicionamiento para que lleguen a temer a cosas como flores, hongos o figuras geométricas. Inversamente, mientras que unos días eliminan los miedos adquiridos (condicionados) a flores, hongos y figuras geométricas, los miedos a serpientes, arañas y precipicios persisten a lo largo de la vida. Estos hallazgos sugieren, en primer lugar, que es plausible la tesis de que existen elementos innatos almacenados en las carpetas de activadores de emociones (son aquellos a los que se reacciona con mayor facilidad y que es muy complicado eliminar). Por otro lado, abonan a la hipótesis del sistema de calibración (actualización) de la carpeta de activadores emocionales mediante la asociación o condicionamiento.

⁶² Debe notarse que, en muchos casos, la agregación de elementos a las carpetas de activadores de emociones mediante juicios y valoraciones proposicionales, requieren del mecanismo de asociación de rasgos de la situación, con experiencias afectivas agradables o desagradables. Así, para extraer el rasgo de peligrosidad implicado en la creencia ‘ganará las elecciones un candidato con tendencias fascistas’, de antemano, tendríamos que haber asociado a un gobernante con tal característica (fascismo), con la medida en que pone en riesgo valores civiles y políticos como la libertad, la justicia y la igualdad.

valoraciones corporeizadas: cada tipo de emoción se distingue por un patrón específico de cambios fisiológicos en virtud del cual *representa un objeto formal*.

Las valoraciones corporeizadas y la Estrategia de la Ramificación de las Emociones (ERE). Aunque Prinz desarrolló su teoría de las emociones como valoraciones corporeizadas con la finalidad, no sólo de mantenerlas unidas, sino también para respetar las clasificaciones emocionales vigentes en el lenguaje ordinario (Prinz, 2005/2010), vale la pena hacer notar que el propio Prinz ha reconocido que quizás sean necesarias ciertas distinciones taxonómicas ulteriores. Tales distinciones se fundarían en pequeñas variaciones en torno al objeto formal que instiga o activa a un género de emociones. Por ejemplo, Prinz considera que es posible que la alegría —cuyo objeto formal podría caracterizarse como ‘cualquier situación que favorezca las expectativas de sobrevivencia o reproducción’— se subdivide en especies, en virtud del tipo específico de situación que instancia el objeto formal. Así, quizás tendríamos que hablar, al menos, de tres formas de alegría: una relacionada con los placeres sensitivos, otra con la satisfacción asociada a logros y metas, y otra más vinculada con el gozo de los juegos. Cada una de estas especies de alegría instancia de forma peculiar el mismo objeto formal general de la alegría. En este escenario, tendríamos que hablar de que la alegría es un género emocional que incluye tres especies emparentadas por su objeto formal —aunque cada una lo represente de forma peculiar. Más aún, es posible que las sutiles diferencias representacionales entre las especies emocionales que pertenecen al mismo género reflejen también diferencias sutiles en términos de sus patrones de cambios fisiológicos —y, por lo tanto, también en el nivel de circuitería neuronal y respuestas expresivo-conductuales.

Ahora bien, vale la pena reconocer que la sugerencia de Prinz es básicamente la misma a la que se sugirió antes para explicar la diversidad de patrones de cambios fisiológicos en una sola emoción. En aquel caso, la que denominamos Estrategia de la Ramificación de las Emociones (*ERE*) ayudó a explicar el hecho de que ciertas emociones instanciaran patrones diversos de cambios fisiológicos —y respuestas expresivo-conductuales. Lo que es crucial señalar, es que la *ERE* no amenaza en absoluto los principios y compromisos ontológicos y metodológicos de nuestra teoría general de las emociones. No representa una amenaza en contra de la unidad de las emociones. El reconocimiento de géneros emociona-

les y especies emocionales es consistente con nuestra definición de la estructura explicativa de una emoción.

Las ECS y sus homólogas Básicas en el marco de las valoraciones corporeizadas. En el apartado anterior alegamos a favor de la posibilidad de que las ECS sean homólogas seriales respecto de sus contrapartidas Básicas. Lo que la homología serial significa en el caso de las emociones, es que las formas CS evolucionaron sus rasgos funcional-adaptativos respecto de sus versiones Básicas, dejando relativamente intactos el resto de los mecanismos causales —neurales, fisiológico y expresivo-conductuales. Las ECS tienden a diferir de sus contrapartidas Básicas en términos de la especificidad de la situación u objeto formal que las activa, sin que dejen de representar, de algún modo, el mismo objeto formal. Así, podríamos decir que tanto la vergüenza-B y la vergüenza-CS tienen el mismo objeto formal: subordinación o pérdida de estatus en la jerarquía social. Pero cada advocación representaría el mismo tema de formas distintas. La vergüenza-B lo representa en la medida en que se activa en situaciones en las que el organismo se subordina sobre la base de la fuerza y coerción, y la vergüenza-CS en la medida en que se activa en situaciones en las que el organismo sufre un revés en una competencia de prestigio. Esto implicaría que la relación de homología serial entre las ECS y sus contrapartidas Básicas apunta a la existencia de un género emocional —unido por su historia evolutiva, similitud de respuestas neurofisiológicas, expresivo conductuales y un mismo objeto formal— que incluye dos especies que se distinguen por la forma peculiar en la que cada una representan el mismo objeto formal. Ahora bien, es factible considerar que haya otros rasgos, a nivel neurofisiológico y psicológico, que distingan a las ECS de sus advocaciones Básicas. En primer lugar, mientras que las ECS presentan una expansión hacia regiones corticales, las homólogas Básicas ocupan, principalmente, regiones subcorticales. Derivadas de estas diferencias neuronales, habría diferencias psicológicas, como la capacidad de las ECS de ser sensibles —a diferencia de las contrapartidas básicas— a procesos cognitivos más complejos realizados en regiones corticales, como juicios. No obstante la plausibilidad de estas diferencias, tendríamos razones para hablar de un solo género emocional, dado el caso que se ha construido antes a favor de su homología.

Finalmente, podemos ver que la propuesta de Prinz sobre el carácter representacional de las emociones y su descripción de las mismas como *valoraciones corporeizadas*, es compatible con la forma en la que hemos venido tipificando las propiedades neurofisiológicas, expresivo-conductuales y biológicas que constituyen, según nuestra teoría, la estructura explicativa de las emociones.

6. Efectos psicológicos característicos.

Otro rasgo psicológico distintivo de las emociones tiene que ver con la forma en la que afectan otros procesos cognitivos como la atención, la motivación, la memoria, el aprendizaje (Cosmides y Tooby, 2000; 2008) y la elección (Damasio, 1994; 2000; A. Damasio, Tranel, H. Damasio, 1998). Es de esperarse que dichas influencias involucren conexiones neuronales entre las estructuras y circuitos que controlan las emociones y las estructuras en las que se realizan dichos procesos. La identificación de estas correlaciones está bien documentada en las investigaciones de Damasio (1994, 2000, 2005) respecto de las emociones, la motivación y la elección; LeDoux (2000) respecto de las emociones el aprendizaje y la atención; Rolls, (2000, 2005) respecto de las emociones, la motivación, el aprendizaje, la memoria y la atención; Gallagher, (2000) respecto de las emociones y el aprendizaje; Barton y Aggleton (2000), respecto de las emociones el reconocimiento y la memoria. Más adelante, cuando se aborde la cuestión de las propiedades fenomenológicas de las emociones, se propondrá una explicación de cómo estos efectos psicológicos son posibles gracias a la conciencia o fenomenología emocional (II, v.II, V).

IV. PROPIEDADES O RASGOS DE SU ARQUITECTURA.

En cuanto a su arquitectura funcional, las emociones exhiben, no modularidad fodoriana —como se definió antes—, sino lo que Prinz (2006a) denomina *cuasi-modularidad*.

7. Cuasi-modularidad.

Esta propiedad compleja de las emociones involucra los siguientes rasgos:

a. Especialización funcional, lo que significa que los *cuasi-módulos* tienen funciones específicas, más allá de las cuales no les es posible operar. En el caso de las emociones, parece bastante claro que exhiben este rasgo. Como se especifica antes —sobre sus propiedades

biológicas— las emociones están diseñadas como respuestas adaptativas a situaciones evolutivamente relevantes, que desencadenan cambios fisiológicos, expresivo, conductuales y psicológicos que tienen como finalidad enfrentar de forma satisfactoria la situación. También, en tanto que *representaciones* se activan ante ciertos *temas evolutivos* relevantes y, en este sentido, existen precisamente para desempeñar tal función.

b. Las fallas (breakdowns) características en los *quasi-módulos* remiten a disfunciones distintivas, relacionadas con daños en las estructuras cerebrales o neuroquímicas que se asocian con los *quasi-módulos*. Este aspecto de la *quasi-modularidad* es consistente con los datos que se proporcionaron en el apartado I, sobre las estructuras cerebrales especializadas en la producción de emociones: cada emoción, en tanto que respuesta adaptativa, involucra la activación de estructuras y circuitos neuronales específicos. La consistencia se presenta porque es sencillo suponer que las *fallas características* estarán relacionadas con lesiones que comprometen la circuitería neuronal emocional. Así, por ejemplo, Chwalisz y sus colegas (1988) mostraron que la capacidad para experimentar emociones disminuye cuando se dañan los centros cerebrales que permiten que la información de lo que sucede en el cuerpo llegue al sistema nervioso central: a saber, la corteza insular anterior. La llamada ceguera emocional se presenta cuando se lesiona una estructura íntimamente relacionada con las emociones: la amígdala. Finalmente, la alexitimia —condición en la cual, la persona es incapaz de reconocer sus propias emociones, o las confunden con simples dolores o punzadas— parece estar causada por un funcionamiento anormal de una estructura llamada cíngula anterior.

c. El procesamiento automático y semi-automático: característico de la *cuasi-modularidad* debe entenderse de la forma correcta. Los *cuasi-módulos* pueden estar sujetos a cierta forma de control voluntario, en el sentido de que ciertas consideraciones de las que estamos al tanto influyen en la respuesta emocional que desplegaremos⁶³, lo que significa que no son completamente automáticos. Por ejemplo, la percepción visual, que Prinz considera *cuasi-*

⁶³ Como cuando sabemos que la presencia de una persona nos llenará de gozo y propiciamos y planeamos, no sólo el encuentro, sino la pauta de conductas que nuestro cuerpo asumirá y decidimos (de antemano) no tenerlas. Como cuando anticipamos nuestro encuentro con el ser amado y elegimos, de antemano, hacer que tenga lugar el encuentro y, llegado el momento, abrazarlo y besarlo efusivamente, tal y como sabemos que la emoción misma (en tanto que cambios fisiológicos y conductas estereotípicas) nos dispondrá. O, a la inversa, como cuando, anticipadamente, se elige todo lo contrario.

modular, puede estar sujeta a cierto grado de control voluntario. Por ejemplo, cuando buscamos algo, tendemos a descartar propiedades cromáticas o formas que no están asociadas con el objeto buscado. Incluso, el estar en modo de búsqueda visual da lugar a errores en los que ciertos objetos se ‘ven cómo’ aquello que se busca. Por supuesto, en muchos casos, los sistemas *cuasi-modulares* se activan y trabajan de forma automática, sin ningún esfuerzo o intervención, como cuando, en el caso de la percepción visual, se presenta una preferencia automática, pero previamente inducida (por familiaridad) hacia cierto tipo de imágenes (Quinn, 1998). Las emociones presentan las mismas pautas de activación: automática y semi-automática. El segundo caso se presenta cuando, cierta predisposición, hasta cierto punto voluntaria, guía nuestras respuestas emocionales; como cuando anticipamos una confrontación y resolvemos que responder con firme enojo o indignación sería lo más adecuado. Por supuesto, las emociones también se activan automáticamente, como cuando sobreviene el miedo al perder repentinamente el sustento, la tristeza al ver alguien llorando o el enojo cuando alguien nos cierra el paso al conducir. En resumen, el carácter automático implicado en la *cuasi-modularidad* concede espacio conceptual para que la activación de las emociones esté sujeta a cierto grado de control.

d. Un cuasi-módulo está construido a partir de reglas y representaciones innatas, lo que quiere decir, en el caso de las emociones, que existen mecanismos innatos que activan patrones específicos de cambios corporales ante la presencia de cierto tipo de estímulos (*temas evolutivos*). Si bien los patrones de cambios neurofisiológicos que distinguen a cada emoción están programados y cableados en nuestra arquitectura neurofisiológica y no están sujetos a modificaciones voluntarias, los estímulos son más flexibles en este sentido. Siguiendo a Prinz (2004, 2006a), suponemos que nacemos con carpetas que almacenan las representaciones de los estímulos que causan cada emoción. Tales carpetas (*triggering files*) contienen representaciones innatas de estímulos que causan las emociones —p. ej., los ruidos fuertes y los empujones están incluidos como representaciones en las carpetas de activadores de miedo y enojo, respectivamente— pero, también puede agregar o eliminar nuevas representaciones de estímulos. Más adelante hablaremos del sistema de recalibración de estímulos que hace posible esta plasticidad de activación en las emociones, por ahora, debe quedar claro que, sólo en parte, las representaciones de estímulos concretas que activan las emociones son innatas.

d. *Ser estímulo-dependiente*⁶⁴ es el rasgo distintivo de los *cuasi-módulos*. La *dependencia-de-estímulo* significa que los módulos están constreñidos o limitados por sus estímulos característicos; es decir, responderán *general y principalmente* (no *exclusivamente*) a los estímulos característicos que los activan. El constreñimiento implicado en la *dependencia-de-estímulo* no impide que procesos de arriba-hacia-abajo (como juicios) lleguen a influir en el funcionamiento del módulo. Debe notarse que la *dependencia-de-estímulo* resulta ser un rasgo modular más plástico y flexible que el encapsulamiento, ya que acepta la posibilidad de que los procesos internos al módulo puedan verse intervenidos desde otros sistemas.

Prinz (2006b), ha negado que el encapsulamiento sea una propiedad modular. Alega que existe alguna evidencia empírica que favorece la tesis de que los sistemas modulares (perceptuales) se comunican entre sí. En este sentido, menciona el “Efecto McGurk”, el cual muestra que el sistema visual puede verse afectado por el auditivo y viceversa. Fodor postuló el encapsulamiento como una propiedad de los sistemas modulares a partir de casos como la ilusión “Müller-Lyer”, en los que no importa la información que se nos proporciona en relación a la igualdad en longitud de las líneas, seguiremos percibiéndolas como de tamaños diferentes. Pero, de acuerdo con Prinz, una explicación alternativa es que los estímulos (perceptuales) que tengan la dirección abajo-hacia-arriba *vencerán* (*trump over*) a los procesos con dirección arriba-hacia-abajo (juicios o creencias), en caso de que lleguen a entrar en conflicto. He aquí, un caso en el que el módulo exhibe *dependencia-de-estímulo*. Cuando no se presente un conflicto de este tipo, los estímulos con dirección arriba-hacia-abajo podrán afectar el módulo, como cuando se nos señala que podemos ver distintas figuras en una sola imagen, como en el caso de las ilusiones por composición ambigua: p. ej, el famoso ejemplo del ‘Patonejo’.

Las emociones también exhiben *dependencia-de-estímulo*. Si bien responden general y principalmente a los estímulos que, por su diseño, deben detectar y representar, las emociones pueden verse afectadas por procesos de arriba-hacia-abajo. Esto sucede, por

⁶⁴ Este rasgo sustituye al *encapsulamiento informacional* de los módulos fodorianos.

ejemplo, en los casos en los que un juicio o pensamiento antecede o es la causa de la emoción; como cuando sentimos miedo después de juzgar que un asaltante acecha nuestra casa. Sin entrar en la cuestión de si es o no el peligro el que causa, en tanto que estímulo, la emoción en este caso, lo que es evidente es que esta instancia de miedo fue influida por un proceso arriba-hacia-abajo (juicio), a diferencia de muchas otras instancias de miedo en donde la causa es un estímulo apropiado y genera un proceso abajo-hacia-arriba, como cuando vemos una serpiente en nuestra senda. Para ver de forma más clara por qué las emociones exhiben *dependencia-de-estímulo*, debemos comenzar por notar, en primer lugar, que, en los casos en los que las anteceden juicios, son sensibles a procesos y actividad en sistemas ajenos al propio módulo. Mientras esto es de esperarse en los *cuasi-módulos*, es inaceptable e inexplicable en términos del encapsulamiento de los módulos fodorianos. Pero también puede verse que las emociones, en casos en los que los estímulos que deben detectar y representar entran en conflicto con la información que proviene de otros sistemas, los primeros vencerán a la segunda y exhibirán ese aspecto recalcitrante que las caracteriza (como a las ilusiones Müller-Lyer): en pocas palabras, exhibirán *dependencia-de-estímulo*. Casos de este tipo se presentan cuando, a pesar de nuestras creencias y juicios, nos resulta complicado dejar de sentir una emoción: como cuando sentimos miedo cuando un león nos ruge detrás de un cristal reforzado, a pesar de que sabemos que estamos a salvo; o cuando experimentamos enojo con una mesa por habernos golpeado con ella, a pesar de que sabemos que no pudo haber tenido intención de lastimarnos, etc. Los ejemplos parecen apoyar la idea de que, cuando se trata de emociones, si se presenta un conflicto, los estímulos indicados vencerán a nuestro buen juicio. En caso de que no se presente alguna tensión entre estímulo y juicio, el último podrá guiar e influir en la emoción. En esto radica la *dependencia-de-estímulo* y es un rasgo distintivo de la *cuasi-modularidad* de las emociones.

La *cuasi-modularidad* y, principalmente, la *dependencia-de-estímulo*, es consistente con, y ayuda a explicar, un par de hechos que parecían ser incompatibles en el marco de la modularidad fodoriana y su noción de encapsulamiento; a saber, su sensibilidad a procesos superiores como la cognición (juicios), y su carácter *recalcitrante*.

Otro rasgo sumamente relevante de la arquitectura de las emociones, que es útil también para entender su plasticidad en términos de la variedad y diversidad de estímulos

que pueden activarlas, es su *sistema de recalibración de estímulos*. Aunque ya hemos introducido este rasgo de la arquitectura de las emociones, cuando discutimos el carácter representacional de las emociones, ahora es importante entender cómo encaja en el marco general de la *quasi-modularidad* de las emociones.

En la medida en que las emociones, en tanto que *cuasi-módulos*, son *estímulo-dependientes* (y, por tanto, no exhiben encapsulamiento), la mera postulación de un sistema de recalibración de estímulos emocionales es menos polémica. El modelo que proponemos para el sistema de recalibración de estímulos emocionales, de nuevo, retoma las ideas de Prinz (2006a, 2004, 2003a y 2003b).

8. Sistema de recalibración.

Dado el carácter cambiante y complejo de los entornos naturales y sociales en los que se desenvuelven los animales, incluidos los seres humanos, es indispensable contar con mecanismos que les ayuden a enfrentar nuevos retos adaptativos. Las emociones, en tanto que tienen como objetivo responder de forma eficiente a viejos retos evolutivos, que no dejan de presentarse bajo nuevos aspectos, precisan de la capacidad de calibrar y recalibrar constantemente los estímulos cuya detección y representación les ha sido encomendada evolutivamente. La forma en que las emociones responden a este problema evolutivo se entiende de la siguiente forma.

La plasticidad de las emociones en términos de los estímulos que las pueden activar, obedece a que tienen a su disposición *mecanismos representacionales* diversos: en primer lugar y principalmente (a) las *percepciones*, luego (b) las asociaciones adquiridas entre un rasgo del entorno y una experiencia afectiva agradable o desagradable (condicionamiento); o (c) los juicios o valoraciones articuladas en términos proposicionales/conceptuales (Prinz, 2004, pp.74-77). Ahora bien, ¿cómo se deben entender la recalibración de estímulos emocionales a partir de la disposición que tienen las emociones respecto de los mecanismos representacionales mencionados?

Prinz considera que nacemos con una especie de carpeta mental (para cada emoción) que contiene representaciones de los rasgos distintivos del entorno que, al ser percibidos, activan la respuesta emocional. Así, para la emoción del miedo, rasgos como ‘tener la

aparición de un objeto largo y enroscado’, ‘oscuridad’, ‘pérdida súbita de la estabilidad’ o ‘una gran altura’, serían algunos de los elementos archivados (de fábrica o de forma innata) en la carpeta de los *activadores* del miedo que, al ser detectados en el entorno por nuestro sistema perceptual, desencadenarían la emoción. Los miedos compulsivos, como el que se desencadena y experimenta al ver una serpiente, tienen como *activadores* rasgos que se encuentran de forma innata en la carpeta ‘activadores de miedo’, y que son *percibidos* en el entorno. Aquí, el *mecanismo representacional* que causa la emoción —el patrón de cambios neurofisiológico— es (a) una *percepción*. Ahora bien, las carpetas de activadores de cada emoción pueden recalibrarse —agregar nuevos *activadores* y eliminar otros— en la medida en que las emociones, en tanto que *cuasi-módulos*, son sensibles a procesos y mecanismos de representación más complejos. Una forma de aprender a tener miedo de cosas nuevas o de agregar elementos a una carpeta de activadores emocionales, es mediante el condicionamiento o la asociación de un rasgo del entorno con una experiencia afectiva agradable o desagradable. Para seguir con el mismo ejemplo, el miedo a las habitaciones pintadas de color amarillo se pudo haber adquirido después de haber pasado, durante la infancia, penosos días de enfermedad y agonía en una habitación amarilla. En este caso, el activador también es percibido, pero no se trata de un rasgo incluido de nacimiento en la carpeta, se ha agregado mediante (b) una asociación entre dicho rasgo y una experiencia afectiva desagradable⁶⁵. Finalmente, con la adquisición del lenguaje, es posible conceptualizar el objeto formal que ejemplifican los *activadores* de cada emoción. *Peligro*, en el caso del miedo. Esto hace posible que, al considerar situaciones en términos (abstractos) de juicios y creencias —como la de que ‘ganará las elecciones un candidato con tendencias fascistas’— se extraigan algunas implicaciones a las cuales podríamos juzgar como exhibiendo el objeto formal de la emoción del miedo: *peligro*. En estos casos, el mecanismo representa-

⁶⁵ El trabajo de Öhman (1993) ofrece evidencia a favor del enfoque de *activadores* innatos y de la noción de *calibración* de las carpetas que almacenan dichos *activadores*. Öhman muestra que, mientras es muy sencillo lograr que un macaco desarrolle una respuesta de miedo ante la presencia de serpientes, arañas y precipicios, se requiere de un largo condicionamiento para que lleguen a temer a cosas como flores, hongos o figuras geométricas. Inversamente, mientras que unos días eliminan los miedos adquiridos (condicionados) a flores, hongos y figuras geométricas, los miedos a serpientes, arañas y precipicios persisten a lo largo de la vida. Estos hallazgos sugieren, en primer lugar, que es plausible la tesis de que existen elementos innatos almacenados en las carpetas de activadores de emociones (son aquellos a los que se reacciona con mayor facilidad y que es muy complicado eliminar). Por otro lado, abonan a la hipótesis del sistema de calibración (actualización) de la carpeta de activadores emocionales mediante la asociación o condicionamiento.

cional por medio del cual se agrega un elemento a la carpeta de *activadores* de miedo, y que desencadena la emoción es (c) un juicio o valoración en términos proposicionales.⁶⁶

El sistema de recalibración de los estímulos emocionales se explica, entonces, apelando a la comunicación que existe entre las emociones y otros mecanismos representacionales, a partir de los cuales es posible actualizar los archivos en las carpetas de activadores emocionales, y donde las actualizaciones han de respetar el tema evolutivo que importa a cada carpeta o emoción. Este modelo para la arquitectura del sistema de recalibración de estímulos, es consistente con el carácter *cuasi-modular* de las emociones y ayuda a hacer inteligible el hecho de que las emociones parezcan sensibles, en algunas ocasiones, a la experiencia y la razón —por llamar de algún modo al mecanismo cognitivo que es útil para detectar estímulos adecuados para una emoción a partir de inferencias sobre creencias—, lo cual resulta inaceptable a la luz del encapsulamiento fodoriano.

Ahora es importante notar que nuestro modelo de la arquitectura de las emociones es inmune a la objeción (*supra*, II, v.I) que Griffiths (1997) avanza en contra de la tesis de que las Emociones Cognitivas Superiores (ECS) sean emociones en el sentido en que sí lo son las Emociones Básicas (EB). Griffiths afirma que las ECS no exhiben algunos rasgos de la modularidad fodoriana, como el encapsulamiento o impenetrabilidad cognitiva, que sí presentan las EB. Pero, ya que según nuestro modelo de la arquitectura emocional postula la *cuasi-modularidad* y un *sistema de recalibración de estímulos*, no sólo es irrelevante que las ECS no exhiban los rasgos de la modularidad fodoriana; mejor aún, se explica cómo es posible que las ECS y muchas EB, pueden ser sensibles a procesos efectuados desde otros módulos.

9. Duración breve

Decir que las emociones tienen una duración breve es conveniente ambiguo. La brevedad es relativa y, en el caso de las emociones, esto significa que puede durar algunos minutos o,

⁶⁶ Debe notarse que, en muchos casos, la agregación de elementos a las carpetas de activadores de emociones mediante juicios y valoraciones proposicionales, requieren del mecanismo de asociación de rasgos de la situación, con experiencias afectivas agradables o desagradables. Así, para extraer el rasgo de peligrosidad implicado en la creencia ‘ganará las elecciones un candidato con tendencias fascistas’, de antemano, tendríamos que haber asociado a un gobernante con tal característica (fascismo), con la medida en que pone en riesgo valores civiles y políticos como la vida, la libertad, la justicia y la igualdad.

quizás, hasta algunas horas. La brevedad, en este sentido vago, es común en otros procesos o fenómenos biológicos. Piénsese en la copulación entre los mamíferos. En el mismo sentido, se podría decir que se trata de un fenómeno relativamente breve, a pesar de que su duración puede oscilar de algunos segundos o varios minutos, y hasta horas. En el caso de las emociones, los factores de los que depende su duración son, en primer lugar, la presencia y persistencia del estímulo o situación que las desencadenan. Mientras la causa de la emoción siga presente y manifestando el *tema evolutivo* que representa, la emoción —los mecanismos de retroalimentación neurofisiológicos— seguirá experimentándose. Por ejemplo, si detectamos una serpiente en la vereda por la que caminamos y enseguida la vemos huir en la dirección opuesta, el miedo tenderá a desaparecer gradualmente. Sí, por el contrario, la serpiente permanece en su lugar y en posición de ataque, o incluso si avanza hacia nosotros, el miedo seguirá experimentándose. Del mismo modo, el enojo que sobreviene ante un insulto tenderá a desaparecer si nos alejamos de la persona que nos insultó, pero seguirá presente si nos enfrascamos en un intercambio de insultos. El segundo factor que determina la duración de la emoción es su reverberancia fisiológica. El patrón de cambios fisiológicos que caracteriza a cada emoción, y su percepción, presenta cierta inercia que no se extingue abruptamente, como sí lo puede hacer el estímulo. El carácter fisiológico de las emociones explica la resonancia emocional que perdura incluso algún tiempo después de que el estímulo o situación que activó la emoción ha desaparecido o quedado atrás. Un estudio de Bradley y Lang (2000) sugiere que el tiempo de restablecimiento de las funciones y procesos fisiológicos y endócrinos alterados por una emoción puede ir de los dos a los quince minutos; siempre y cuando el estímulo o situación que la activa desaparezca o sea despojado —por medio de una nueva valoración— del *tema evolutivo* con el que se representó originalmente.

Algunos objetarían, en contra de esta tesis acerca de la duración de las emociones, que existen emociones que se caracterizan por prolongarse durante lapsos tan largos como años o la vida misma de una persona. En este sentido, se suele hablar de que *X* ha amado toda su vida a *Y*, o de que *Z* odia desde la infancia a su padre. El problema con estos contraejemplos es que pasan por alto el carácter *disposicional* de las emociones. Básicamente, cuando se habla de emociones de largo plazo, como en los ejemplos que se acaban de mencionar, lo que se quiere decir es que se tiene cierta disposición a tener ciertas emociones

hacia ciertas personas y circunstancias, cuando estas confluyen. Considérese el de una persona adulta, digamos *A*, que declara que *siente* asco por los gusanos desde que tiene memoria. No diríamos que *A*, por ejemplo, después de haber realizado todas sus actividades cotidianas sin haber visto un solo gusano en todo el día, *siente* o *sintió* asco por los gusanos en las pasadas 24 horas. Lo que en realidad se quiere decir en los casos como en los de *A*, es que, si viera o tocara o incluso si imaginara vívidamente gusanos, *sentiría* de hecho asco. *A* tiene la disposición de sentir o tener la emoción de asco ante la presencia (real, potencial o imaginaria) de gusanos. Lo mismo sucede con los supuestos casos de amor y odio de largo plazo. A lo que en realidad nos referimos es a que tenemos la disposición de sentir o experimentar amor u odio hacia ciertas personas o situaciones en ciertas circunstancias características. Lo que significa que, ante estas personas o situaciones —como en el caso del asco— tendremos la disposición a tener ciertos patrones de activación neurofisiológica y expresivo-conductuales —entre otros procesos— característicos. En realidad, otra forma de exponer el mismo punto, cuando *A* dice que ha amado a *X* durante años y años, uno debería pensar que la conducta de *X* ha sido apropiada para causar constantemente *la emoción* de amor en *A*. Los detalles, el apoyo, la comprensión que ha mostrado *X*, así como sus ausencias y problemas, han favorecido un apropiado relieve emocional que han mantenido a *A* unida afectivamente a *X*. Pero no significa que *A* haya amado, efectivamente, a *X* todo ese tiempo. En todo ese tiempo es posible que *A* también haya odiado y despreciado a *X*, que se haya enojado con él e, incluso, que le haya temido. Quizás todo este relieve y altibajos emocionales, en lugar de dar lugar a emociones a largo plazo (esta idea quizás se tenga que cuestionar), propicia ciertas formas de vínculos o dependencia afectiva. Quizá esta es la base de las relaciones a largo plazo (que no de las emociones a largo plazo), la cantidad de emociones breves y discretas que se experimentan con una persona.

En resumen, según la postura que defendemos, las emociones de largo plazo no son más que la disposición de tener episodios emocionales breves y discretos. En realidad, las emociones, entendidas como el conjunto de propiedades homeostáticamente articuladas que estamos caracterizando, no son tan largas como la usanza coloquial, en algunos casos, lo sugiere

V. PROPIEDADES FENOMENOLÓGICAS.

10. Fenomenología de las emociones

El rasgo que, quizás, más claramente se suele asociar con las emociones es la forma en la que se experimentan. La cuestión con dichas propiedades consiste en cómo ubicarlas en el espacio explicativo de una caracterización o teoría general de las emociones. La respuesta que ofrecemos a esta cuestión, no proviene de lo que los modelos circunflejos (Russell, 1980; y Plutchik, 1984) de las emociones dicen sobre la valencia —el tono hedónico o la cualidad placentera o desagradable en la experiencia de una emoción. En lugar de usar la fenomenología de las emociones como un principio explicativo, intentaremos identificar la función que tiene en el marco de los mecanismos causal-homeostáticos en el que se ven implicadas las demás propiedades físicas relevantes —especificadas antes— que constituyen la estructura explicativa de las emociones.

Pero, antes de ofrecer una caracterización positiva de la naturaleza y función de la fenomenología o conciencia emocional, es necesario atender a una objeción recurrente en contra de la idea de que la fenomenología sea un rasgo distintivo o esencial de las emociones. Se suele decir que podemos tener emociones sin fenomenología, como cuando decimos, antes de las elecciones, que tenemos miedo de que gane cierto partido por su historia represiva, o como cuando decimos en una plática que nos indignan y enojan las políticas sociales del gobierno. La misma aclaración respecto del carácter disposicional de las emociones, en el apartado anterior, puede ayudar a entender este tipo de casos en los que, supuestamente, la fenomenología está ausente en ciertos episodios emocionales. La cuestión es que, en los ejemplos mencionados, el lenguaje no reporta propiamente el estado psicológico actual de la persona. Las personas de los ejemplos, de hecho, no tienen miedo o están enojadas. Lo que significa, si es verdad lo que dicen, es que uno tendrá la disposición a sentir miedo si gana cierto partido político y pone en práctica ciertas políticas represoras; mientras que el otro tendrá la disposición a enojarse cuando abra un periódico y lea sobre otro caso más de corrupción. Considero que una gran cantidad de contraejemplos de esta clase, pueden perder peso si se interpretan bajo la luz del carácter disposicional de las emociones. Otro tipo de casos de supuestas emociones sin fenomenología requieren de una estrategia distinta. Por ejemplo, es habitual que las personas digan cosas como las siguientes

sin sentir nada realmente: a un amigo que nos invitó a su fiesta, ‘Lo siento pero no podré asistir a la fiesta’; a una persona que nos pide ayuda económica en la calle, ‘Me apena mucho pero no puedo ayudarle’; a un compañero de trabajo, ‘Me sorprende que le hayan puesto tocino a mi emparedado’. En estas ocasiones, de nuevo, los términos emocionales no se refieren a estados psicológicos que de hecho están ocurriendo. Son sólo expresiones o recursos retóricos que utilizamos para caracterizar situaciones en las que, a partir de ciertos supuestos o normas sociales, deberíamos estar experimentando alguna emoción específica. Deberíamos sentirnos tristes por defraudar a un amigo, culpables por no ayudar a alguien, o sorprendidos si se violan nuestras expectativas. Pero, como una cuestión de hecho, no siempre sucede así. En el primer caso, se puede deber a que sabemos que nuestro amigo es comprensivo; en el segundo a que no conocemos realmente a la persona (no sabemos si nos defrauda en realidad); y en el tercer caso puede deberse simplemente a que no todas nuestras expectativas violadas causan sorpresa. En tales casos, el recurso de atribuirnos emociones no parece ser más que un recurso retórico para aludir a la emoción que, supuestamente, tendríamos que experimentar, aunque de hecho no sea así.

Nos parece que estas dos estrategias sugeridas pueden ser suficientes para hacer frente a los casos tentativos de emociones sin fenomenología. Ahora es turno de ofrecer una caracterización positiva de la fenomenología de las emociones. Es hora de exponer lo que, según el modelo que proponemos, es la fenomenología de las emociones: cuál es su función y cómo encaja en el racimo de propiedades homeostáticas que constituyen la estructura explicativa de las emociones.

La fenomenología de las emociones está vinculada de forma más evidente con las propiedades neurofisiológicas (II, VII.I, I), biológicas (II, VII.I, II), representacionales y otros efectos psicológicos (II, VII.I, II) de las emociones. Antes se ha sostenido que las emociones involucran patrones de cambios fisiológicos que, a su vez, son orquestados desde ciertas estructuras cerebrales específicas (II, VII.I, I). También hemos sostenido que, desde el punto de vista psicológico-funcional, las emociones son *valoraciones corporeizadas* (II, VII.I, III), esto significa que los patrones de respuestas fisiológicas características de cada emoción —que se orquestan desde ciertas estructuras cerebrales— tienen como finalidad *representar* rasgos perdurables del entorno que son de relevancia evolutiva para los

individuos de la especie (II, VII.I, II). Cada patrón de cambios fisiológicos asociado con una emoción característica representa un *tema evolutivo*⁶⁷. Los sentimientos o fenomenología de las emociones no son más que la percepción de los cambios fisiológicos característicos de cada emoción (Prinz, 2004) y, como se ha mencionado antes, el **pons**, la **ínsula**, la **corteza cingulada anterior (regiones dorsal y rostral)**, las **cortezas somatosensoriales primaria y secundaria**, son algunas de las estructuras implicadas en la producción de dichos sentimientos (fenomenología). Pero si los cambios fisiológicos son ya, en sí mismos, representaciones de *temas evolutivos* (en tanto que *evaluaciones corporeizadas*), ¿qué función tiene la fenomenología de las emociones? ¿De qué sirve experimentarlas conscientemente de formas características?

Tanto Prinz (2006a) como Damasio (1994, 2005) consideran, y estamos de acuerdo con ellos, que la fenomenología de las emociones no es una mera floritura psíquica. Para Damasio, los sentimientos implicados en las emociones tienen como finalidad poner al tanto al organismo de su estado y posibilidades en relación con el entorno natural y social. Desde esta perspectiva, la fenomenología sirve como una especie de alarma que favorece procesos de aprendizaje que se nutren y robustecen en la especificidad de la experiencia emocional (Damasio, 1994; Gallagher, 2000). La vieja máxima que reza, “la letra con sangre entra”, y la no tan añeja, pero bien conocida tesis pedagógica que asocia el placer del juego con el aprendizaje, bien pueden reflejar la idea implicada en la hipótesis de la función de la fenomenología de las emociones que proponemos: las experiencias emocionales favorecen y refuerzan los procesos de aprendizaje⁶⁸.

Prinz (2004), por su parte, alega a favor de una función de la fenomenología de las emociones que va más allá del aprendizaje. Para entender qué función le atribuye Prinz a las emociones, es necesario exponer su propuesta sobre cómo surge la fenomenología de las

⁶⁷ Lo que nosotros llamamos ‘tema evolutivo’ es muy parecido a lo que otros llaman ‘objeto formal’ (Kenny, 1963; Lyons, 1980), o ‘tema relacional’ (Lazarus, 1991). Se trata de aquello ‘sobre lo que es’ o representa una emoción. En el caso de los ‘temas evolutivos’ refieren a rasgos generales y recurrentes del entorno (peligro, fracaso, obtención o pérdida de pareja potencial, etc.) que tienen un impacto potencial en el éxito adaptativo de los individuos de una especie. La forma en la que los temas evolutivos son representados, según nuestro modelo, es por medio de cambios fisiológicos en el organismo.

⁶⁸ Más sabiduría popular puede citarse para ejemplificar la idea en cuestión: “Nadie aprende en cabeza ajena”, “Una vez mordido, dos veces precavido” y otros refranes por el estilo, apuntan al hecho, no sólo de que parece necesario haber pasado por una experiencia para aprender algo, sino al hecho adicional de haber tenido una experiencia afectiva importante para reforzar lo que enseña cierto tipo de situación. Ni qué decir de las personas que aprenden a no ser tan crédulos con los demás después de haber sido defraudados o traicionados.

emociones. Prinz considera que la conciencia en general, incluida la conciencia o fenomenología emocional, involucra tres procesos o niveles. En el caso de la conciencia o fenomenología de las emociones, el desarrollo es el siguiente.

En el nivel más bajo, un estímulo o estado o proceso psicológico adecuado activa respuestas auto-regulatorias y el monitoreo de cambios fisiológicos aislados. Las estructuras cerebrales involucradas en este nivel son el **pons**, la **corteza somatosensorial primaria** y la **ínsula**. En el nivel intermedio la información se integra y tienen lugar las representaciones corporales de los patrones de cambios fisiológicos que se desencadenan. Las estructuras cerebrales que mapean estos patrones de cambios son la **corteza somatosensorial secundaria** y, prominentemente, la **corteza cingulada anterior** en su región **dorsal**. Prinz, siguiendo a Damasio (1999), sostiene que es en esta etapa en la que surge la conciencia o fenomenología emocional.⁶⁹ Prinz cita alguna evidencia que sugiere que este es el caso. Hay casos de personas con una condición llamada ‘falla autonómica pura’. Estas personas presentan lesiones en regiones como las **cortezas somatosensoriales** y la **corteza cingulada anterior** y, como consecuencia, ven muy reducidas sus capacidades para detectar cambios fisiológicos y experiencias emocionales⁷⁰. En el nivel superior los patrones de cambios fisiológicos pueden ser *reconocidos* (a pesar de sutiles diferencias) como ejemplares de un tipo de emoción particular. Las regiones cerebrales implicadas en este proceso parecen ser la **corteza prefrontal ventromedial** y la **corteza cingulada anterior** en su región **rostral**. De nuevo, la evidencia para sostener esta hipótesis proviene de casos clínicos. Los pacientes que sufren de alexitimia —una condición en la que una persona es incapaz de *reconocer* o *nombrar* las emociones que experimenta— suelen tener lesiones (entre otras regiones) en la **corteza cingulada anterior**, en la región **rostral**⁷¹.

⁶⁹ Prinz reconoce que cada emoción debería trazar un patrón distintivo dentro de las estructuras cerebrales involucradas en el surgimiento de la fenomenología emocional. Un estudio de Damasio, et. al. (2000) muestra que, en efecto, esto es lo que sucede para el caso de emociones como la tristeza, alegría, enojo y miedo.

⁷⁰ Más evidencia y estudios que abonan a la tesis de que la conciencia o fenomenología emocional se presenta en el segundo nivel, es citada en Prinz (2004). En todos ellos se relatan disociaciones entre la capacidad de experimentar, reportar o reconocer emociones y lesiones en las regiones cerebrales señaladas.

⁷¹ Vale la pena subrayar que el proceso mediante el cual Prinz propone que surge la conciencia apunta a la integración vertical de un sistema neural que comienza en regiones subcorticales y culmina en áreas de la corteza cerebral. Este tipo de aproximaciones (como se ha mencionado en II, VII.I, I) es bastante fructífera, ya que ayuda a entender cómo cierto tipo de capacidades (aprendizaje, emociones, etc.) pueden realizarse de forma casi mecánica (cuando están orquestadas desde estructuras cerebrales más primitivas) o de forma más flexible y compleja (cuando están involucradas regiones corticales).

No obstante la evidencia y viabilidad de las hipótesis propuestas a favor de la idea de que la conciencia o fenomenología emocional surge en el nivel intermedio de procesamiento, Prinz considera que es insuficiente tal como está. Dado que es un hecho, respecto de otras modalidades de conciencia perceptual que se realicen los tres niveles de procesamiento de información sin que ésta tenga lugar —como en los casos de ceguera atencional⁷²—, Prinz sugiere que es necesario añadir un elemento más. El último elemento que se agrega para llegar al surgimiento pleno de la conciencia o fenomenología emocional es la *atención*. Un rápido experimento mental ayudará a mostrar cómo la atención parece ser central para la fenomenología emocional. Imaginemos a una persona en un avión sufriendo un episodio de miedo a volar. Su espalda se mantiene rígida y sus brazos aferrados a los respaldos. En algún momento, el acompañante nota el sobresalto emocional y lo distrae contándole una historia graciosa. Mientras escucha atentamente, el pasajero bien pudo haber dejado de atender momentáneamente a su miedo mientras hacía un esfuerzo por seguir el relato de su acompañante. Una vez finalizado éste, el pasajero se habría encontrado a sí mismo en la misma situación, rígido y aferrado a los respaldos de brazos. No obstante, durante este lapso, podría decirse que dejó de ser consciente de su miedo, gracias a que su *atención* se dirigió al relato de su acompañante.⁷³

Ahora bien, ¿por qué habría de requerirse la atención para el surgimiento pleno de la conciencia de la fenomenología emocional? La respuesta a esta cuestión ayudará a explicar cuál es la función de la conciencia en general, y de la fenomenología de las emociones, en particular. La atención *selecciona* parte de la información perceptual —*filtrando* y *amplificando* ciertos aspectos de la misma— para su procesamiento ulterior. La información seleccionada de esta forma (atencional) es enviada y está a disposición de la memoria a corto plazo, donde puede ser utilizada para procesamientos más complejos como el razonamiento controlado y la solución de problemas. Los procesos controlados son los que hacen posible articular respuestas a situaciones noveles. Este tipo de respuestas también se denominan

⁷² La ceguera atencional —más allá de los curiosos experimentos diseñados para hacerla evidente, como el de la clase que se da en llamar, ‘*Gorilas entre nosotros*’— es bastante común. Tiene lugar cuando una tarea u objeto nos absorbe a tal grado que prácticamente podríamos decir que dejamos de percibir el resto de las cosas que ocurren a nuestro alrededor. En tales casos, es seguro que se sigue procesando la información perceptual en los tres niveles propuestos por Prinz —hay una conciencia periférica—, sólo que no estamos al tanto de ella.

⁷³ Un efecto similar se conseguiría si, durante un ataque de rabia, usted se da a la tarea de resolver manualmente una multiplicación de varias cifras.

respuestas deliberativas. La idea central es que nuestras respuestas controladas dependen de que cierto tipo de información esté disponible en la memoria a corto plazo, para realizar sobre ellas deliberaciones que conduzcan a la elección de conductas adecuadas a la situación y solución de problemas. La información que llega a la memoria a corto plazo para realizar dichas tareas, debe pasar por el filtro de la *conciencia* y la *atención*. Por lo tanto, la función de la conciencia en general debe ser la de enviar información a la memoria a corto plazo para su uso en procesos cognitivos controlados (razonamiento y resolución de problemas). En el caso concreto de la conciencia o fenomenología emocional, está surge también en el nivel intermedio, en el que tiene lugar la representación de patrones de cambios fisiológicos, sólo cuando la atención se dirige a estas representaciones. En tal caso, la información filtrada, amplificada y seleccionada —que no son más que representaciones de patrones de cambio corporales— pasa a la memoria a corto plazo para procesamientos ulteriores. Ahora bien, ya que los patrones de cambios fisiológicos, a su vez, representan *temas evolutivos*, contienen en sí mismos información acerca de la situación del organismo. Por lo tanto, la información que es enviada a la memoria a corto plazo es información relevante sobre la situación del organismo. A partir de esta información se pueden desarrollar procesos controlados —razonamiento y solución de problemas— que dar lugar a respuestas deliberadas. Gracias a esta contribución de la conciencia o fenomenología emocional y la atención, piensa Prinz, es posible llegar a modular nuestras respuestas emocionales. De hecho, si esta teoría de la conciencia o fenomenología de las emociones es correcta, resultará que rasgos importantes de la arquitectura de las emociones (II, V.II, IV) dependerán de ella. Si una de las funciones de la conciencia emocional consiste en enviar información a la memoria a corto plazo para su procesamiento ulterior —lo que significa que esa información está a disposición de procesamiento a nivel cortical— entonces es comprensible que las emociones (en ciertos casos) puedan activarse de forma *semi-automática* (que ciertos pensamientos nos predispongan a tenerlas), que sean *estímulo-dependientes* (que su activación pueda verse afectadas por información de otros módulos) y que sean *recalibrables* los estímulos que las causan (que se agreguen nuevos estímulos a la carpeta de activadores de las emociones y se eliminen otros): un pocas palabras, la conciencia hace posible *la sensibilidad cognitiva* de las emociones. Regresando al caso del viajero con miedo a volar, la información que sobre su estado envían la conciencia emocional y la atención a la memoria a

corto plazo, *puede* ser procesada a luz de más información relevante y, *posiblemente*, dé como resultado un tipo específico de respuesta deliberativa, como respirar y acceder a tomar un tranquilizante o salir corriendo antes de que despegue el avión.

Recapitulando, según el modelo de una teoría general de las emociones que proponemos, la fenomenología de las emociones no es gratuita. Un par de teorías con buen sustento teórico y empírico proponen que la función de la conciencia emocional es, por un lado, favorecer procesos de aprendizaje que abonan al bienestar de los individuos de la especie (Damasio, 1999; 2005). Por otro, enviar información a la memoria a corto plazo para su ulterior uso en procesos controlados como el razonamiento, la solución de problemas y elección de conductas. Esto, a su vez, parece hacer posibles rasgos importantes de la arquitectura de las emociones: rasgos que tienen qué ver con la sensibilidad a procesos cognitivos superiores (II, v.II, IV). También, siguiendo a Prinz (2004), se ha propuesto una hipótesis —basada en un cuerpo interesante de evidencia empírica— sobre cómo surge la conciencia o fenomenología de las emociones. Finalmente, mientras se han atendido los expedientes anteriores, se ha hecho evidente cómo la fenomenología de las emociones encaja en el marco de los mecanismos causales-homeostáticos que involucran al resto de las propiedades físicas relevantes que constituyen lo que, según nuestra teoría de las emociones, es la estructura explicativa de las emociones. Ha quedado claro cómo la fenomenología de las emociones tiene sus bases biológicas en ciertas estructuras cerebrales y se haya íntimamente ligada a los cambios fisiológicos que registran (II, v.II, I). En el mismo sentido, se ha visto que la fenomenología de las emociones está vinculada con las propiedades representacionales y biológicas de las emociones, toda vez que dicha fenomenología *es sobre* los patrones de cambios fisiológicos, los cuales no son más que *valoraciones corporeizadas* (II, v.II, III) que representan temas evolutivos (II, v.II, II). Además, se ha indicado cómo las emociones, por medio de su fenomenología, afectan y se relacionan con otros procesos psicológicos como el aprendizaje, la atención y la memoria (II, v.II, III).

VII.II. ¿QUÉ ES UNA EMOCIÓN?

Ahora contamos con los recursos conceptuales necesarios para proponer una caracterización de lo que es una emoción. Una emoción es una instancia de una categoría psicológica que se considera *clase natural RPH*. Las *clases naturales RPH* son fenómenos cuya estruc-

tura explicativa está constituida por racimos de propiedades físicas homeostática y causalmente vinculadas. Las estructuras explicativas de las *clases naturales RPH*, se caracterizan por ser *indeterminadas*, —lo que significa que algunas propiedades que las constituyen pueden estar ausentes en instancias particulares— y *dinámicas* —lo que significa que dichas estructuras explicativas pueden *evolucionar* con el tiempo (véase, parte I).

En el caso de las emociones, su estructura explicativa está constituida, aproximadamente, por diez propiedades físicas relevantes, de cinco tipos distintos, que se hayan vinculadas homeostáticamente. Para empezar, tienen una función adaptativa, la cual consiste en preparar al organismo para enfrentar situaciones recurrentes con relevancia evolutiva; es decir, situaciones que deben afrontarse de forma satisfactoria para la adecuación (*fitness*) de la especie (II, VII.I, II). Cada *tipo* de emoción pretende ser la mejor forma que encontró la evolución para enfrentar un *tipo* de situación recurrente de relevancia evolutiva. La forma en la que la evolución dio lugar a las emociones involucró la articulación gradual de propiedades de distintos tipos. En primer lugar mencionaremos a las *propiedades neurofisiológicas y expresivo-conductuales*. Las emociones pueden disponer de los recursos neuronales de un (hipotético) sistema anatómico-funcional integrado verticalmente en el cerebro. El sistema incluye estructuras del tallo cerebral, del sistema límbico y de la corteza cerebral. Cada emoción dispone de estos recursos para orquestar patrones característicos de cambios fisiológicos que, a su vez, preparan al organismo para enfrentar la situación por medio de expresiones y conductas estereotípicas (II, VII.I, I).

Las emociones también involucran *propiedades psicológicas* relacionadas con las anteriores. Cada tipo de emoción representa un *tema evolutivo*, es decir, una propiedad recurrente del entorno que, para los organismos de la especie, es importante detectar y sortear. La forma en que las emociones representan estos *temas evolutivos* es por medio del patrón de cambios fisiológicos que es característico de cada emoción. Es por esto que también se designan a las emociones como *valoraciones corporeizadas*. La otra *propiedad psicológica* que exhibe una emoción tiene que ver con sus efectos, precisamente, *psicológicos*. Lo que esto significa es que una emoción puede reorientar y activar o desactivar otros procesos cognitivos. Se sabe que una emoción puede incidir en la atención, la memoria, el aprendizaje, la toma de decisiones, etc. Esta propiedad es posible gracias a la integración

vertical de las estructuras cerebrales que se ven involucradas en la producción de una emoción (II, VII.I, III).

Las emociones también tienen *propiedades o rasgos de arquitectura cognitiva* que se hayan vinculadas con las propiedades anteriores. Cada emoción exhibe cuasi-modularidad, lo que significa que, si bien puede activarse algunas veces de forma automática y recalcitrante ante la presencia del estímulo apropiado, otras más puede ser anticipada, modulada o interferida por otros procesos cognitivos. Esta propiedad de las emociones también es posible gracias a la articulación vertical de las estructuras cerebrales que se ven implicadas en la producción de una emoción. El lóbulo prefrontal en la región ventromediana es fundamental para el control de impulsos, el pensamiento basado en normas sociales y la toma de decisiones y, según nuestro modelo, es una de las estructuras que pueden activarse en episodios emocionales. Entonces, es en parte gracias a la *cuasi-modularidad* —y las estructuras cerebrales de las que pueda depender— que es posible pensar en que los procesos que inician una emoción, y que comienzan de abajo hacia arriba, pueden ser afectados por procesos de arriba hacia abajo, de suerte que tienen lugar una retro-alimentación y actualización continuas de la situación del organismo y de su entorno. Esto es lo que sucede cuando alguna deliberación da lugar al control de las conductas emocionales o la desaparición gradual de la misma. El otro rasgo de la arquitectura de las emociones es el sistema de recalibración, gracias al cual es posible agregar nuevos estímulos como activadores de una emoción, o desechar otros. Este sistema explica cómo ciertos tipos de estímulos pueden, con el tiempo, dejar de activar emociones, y otros, que antes no lo hacían, llegar a causarlas. También hay vínculos entre este rasgo de la arquitectura emocional y la circuitería emocional. Es probable que el hipocampo sea una estructura crucial para la recalibración, en muchos casos. Además incide directamente en el aspecto representacional de las emociones: la recalibración actualiza a lo largo de la vida el conjunto de estímulos que contienen el *objeto formal* que activa a cada emoción⁷⁴. Una última propiedad de la arquitectura emocional es la de *duración breve*. Una emoción durará mientras el estímulo que la causa esté presente y exhibiendo el *tema evolutivo* que causa la emoción. Cuando el estímulo des-

⁷⁴ Estos dos rasgos de la arquitectura emocional explican la sensibilidad cognitiva que, algunas ocasiones, exhiben las emociones. La modularidad fodoriana, por el encapsulamiento que implica, vuelve inexplicable este aspecto de las emociones.

aparece o deja de exhibir el *tema evolutivo*, la emoción comenzará a desaparecer. Esta propiedad también está vinculada con otras que hemos mencionado. La duración de la emoción es directamente proporcional al tiempo durante el cual las estructuras cerebrales, el patrón de cambios fisiológicos y las expresiones y conductas características de la emoción estén activos; lo cual se espera que suceda mientras el estímulo activador esté presente (de algún modo... incluso en la imaginación). Obviamente, durante el tiempo que esté activa una emoción, pueden tener lugar los efectos psicológicos antes mencionados y darse la ocasión para recibir información —cuasi-modularidad— que incida en el desarrollo de la emoción (II, VII.I, IV).

Finalmente, están las *propiedades fenomenológicas*. La fenomenología de una emoción —qué se siente experimentarla— puede ser un factor valioso —junto con otros más— para individuar una emoción. Sin embargo, nuestra propuesta se ha preocupado más por identificar su valor funcional. En este sentido, la fenomenología o conciencia emocional tiene dos funciones principales. En primer lugar, favorecer procesos de aprendizaje. Por otro lado, la conciencia emocional también tiene la función de enviar información a la memoria a corta plazo para su ulterior utilización en formas de procesamiento controlado de información. En este sentido, la conciencia emocional está vinculada con los efectos psicológicos característicos de las emociones y con la sensibilidad cognitiva que exhibe la cuasi-modularidad. La conciencia emocional, por su parte, está ligada a las propiedades neurofisiológicas de las emociones. Lo que se siente durante una emoción, son los cambios fisiológicos que ocurren en el cuerpo. Los cambios fisiológicos son lo que se experimenta y de los que es la conciencia emocional. Esta conciencia está asociada con la actividad de estructuras previamente identificadas: la ínsula y la corteza cingulada anterior (II, VII.I, V).

Una emoción, entonces, es la vinculación homeostática de las diez —quizás algunas menos— propiedades referidas. Donde, por vinculación homeostática nos referimos a las interacciones causales —de procesamiento, retroalimentación y actualización— que tienen lugar entre dichas propiedades, las cuales dan lugar a un estado representacional y conductual que favorece la adecuación (*fitness*) del organismo.

VIII. OBSERVACIONES FINALES SOBRE LA TEORÍA

Ahora es necesario probar la viabilidad del modelo de la teoría general de las emociones que hemos propuesto en el capítulo anterior. La forma más directa de hacerlo es mostrar que cualquier emoción que elijamos, ya sea básica o cognitiva superior, es susceptible de ser explicada en los términos que estipula el modelo propuesto, lo cual comportaría que compartan las mismas propiedades y, por lo tanto, podrían considerarse como pertenecientes al mismo dominio ontológico y explicativo: podrían considerarse como una sola clase natural *RPH*. Recuérdese que, los objetores a una teoría general y unificada de las emociones (Griffiths, 1997; Solomon, 1995; Rorty, 1978, entre otros) estarían en contra de estas posibilidades. También, para finalizar esta investigación, valdrá la pena hacer algunas observaciones en dos direcciones. Por un lado, sobre la *expansión* ontológica de las emociones implicada por nuestro modelo. Es decir, sobre la franca posibilidad de que nuestro modelo dé lugar a distinciones de estados emocionales que no están contemplados en nuestro léxico emocional ordinario. Por otro lado, diremos algo sobre cómo ese mismo léxico emocional parece apuntar a distinciones emocionales donde no las hay. La idea que intentaremos probar es que hay más términos emocionales que emociones. Por lo tanto, en relación al léxico emocional, resultará que la ontología emocional se *contraerá*.

VIII.I. PONIENDO A PRUEBA LA TEORÍA

El modelo de teoría general de las emociones que se propuso en el capítulo anterior pretende ofrecer un marco teórico y conceptual para mantener unidas, en un mismo dominio explicativo y ontológico, a todos o a la mayoría de los fenómenos que ordinariamente categorizamos como ‘emoción’. En este sentido hemos venido argumentando que las emociones constituyen una clase natural de Racimo de Propiedades Homeostáticas (*RPH*), lo que significa que comparten una misma *estructura explicativa*, constituida por un racimo —indeterminado y configurado evolutivamente (históricamente)— de propiedades que están vinculados por medio de mecanismos causal-homeostáticos. Hemos intentado probar que este es el caso proponiendo un racimo de diez propiedades —de cinco tipos o niveles explicativos distintos— que, si estamos en lo correcto, están vinculadas homeostáticamente entre sí, y deben ser instanciadas, aproximadamente, por la mayoría de los fenómenos afectivos que llamamos ‘emoción’. Naturalmente, es inviable probar nuestro modelo con cada

estado psicológico que llamamos ‘emoción’. Lo que sí es posible e indispensable es ofrecer, al menos, algunos ejemplos de cómo puede aplicarse a emociones concretas. En el mismo sentido, parece necesario que se elijan, para esta prueba, un caso paradigmático de lo que se suele llamar Emoción Cognitiva Superior (ECS) y otro de Emoción Básica (EB). Después de todo, la corriente en contra de la cual hemos venido remando sostiene que no pertenecen a la misma categoría psicológica, por lo que no configuran un único dominio explicativo y no conforman una clase natural.

Emociones	
Orgullo	Miedo
I. Propiedades neurofisiológicas y expresivo/conductuales	I. Propiedad neurofisiológicas y expresivo/conductuales
<i>1. Estructuras cerebrales especializadas en su producción:</i> Eje hipotalámico-pituitario-adrenal, corteza prefrontal en la región ventromedial, lóbulo, surco temporal superior posterior y el lóbulo temporal izquierdo ⁷⁵ (Tracy, et. al. 2010).	<i>1. Estructuras cerebrales especializadas en su producción:</i> Amígdala, tálamo, hipocampo, corteza somatosensorial primaria, núcleo basal (LeDoux, 2000, 200a; LeDoux y Phelps, 2000).
<i>2. Patrón de cambios fisiológicos:</i> Excitación física transitoria, no cardíaca, y baja frecuencia en el ritmo cardíaco (Fourie, et. al., 2011) ⁷⁶ .	<i>2. Patrón de cambios fisiológicos:</i> Aumento del ritmo cardíaco y la presión sistólica, lo que da lugar a la producción de adrenalina. También tiene lugar un ajuste término que reditúa en la disminución de la temperatura de la piel, se contrae el tracto intestinal de modo que se suspenden los procesos digestivos. Aumenta la presión sanguínea y se activan las glándulas sudoríparas (Levenson, 2006; Ekman, Levenson y Friesen, 1983).
<i>3. Expresión y conductas características:</i> Cuerpo en postura erguida, expansión de la caja torácica, cabeza inclinada ligeramente hacia atrás, brazos en posi-	<i>3. Expresión y conductas características:</i> Expresión facial característica, repliegue de manos y brazos hacia el pecho, alejamiento del tronco y cabeza del

⁷⁵ Nótese que el orgullo activa una estructura cerebral y un sistema neural-endócrino que figuran en nuestra propuesta de estructuras cerebrales especializadas en la producción de emociones (*supra* II, v.ii, I).

⁷⁶ En este estudio, Fourie y sus colegas mostraron que el orgullo y la culpa muestran distintos patrones de respuestas cardiovasculares. Lamentablemente, no se ha profundizado mucho aún en relación al resto de cambios fisiológicos que involucran el orgullo.

ción de jarras apoyados en las caderas o alzados en forma de ‘v’ con los puños apretados y una sonrisa incipiente en el rostro (Tracy y Robins, 2007, 2004).	estímulo que causa la emoción, temblor, inmovilidad o huída (Ekman, 2003, 1999, 1994, 1992).
II. Propiedades biológicas	II. Propiedades biológicas
4. <i>Función adaptativa</i> : Promover estatus, dominancia y prestigio. Comunicar rango social por medio de sus expresiones conductuales características (Clark, 2012).	4. <i>Función adaptativa</i> : Preservar al individuo identificando situaciones de peligro para su linaje (que lo incluyen a él mismo, sus parientes y, en tercer lugar, a los miembros de su especie).
III. Propiedades psicológico-funcionales	III. Propiedades psicológico-funcionales
5. <i>Carácter representacional</i> : Su tema evolutivo es el éxito (potencial ⁷⁷) en competencias de prestigio social ⁷⁸ (Tracy, et. al., <i>op. cit.</i>).	5. <i>Carácter representacional</i> : Su tema evolutivo es el peligro.
6. <i>Efectos psicológicos característicos</i> : Motivación (da lugar a confianza y promueve conductas competitivas), atención (promueve pensamientos auto-referenciales), reinterpretación de las capacidades del individuo en relación a los demás (Tracy, et. al., <i>op. cit.</i> ; Lazarus, 1991).	6. <i>Efectos psicológicos característicos</i> : Atención (orienta la percepción al estímulo peligroso), reinterpretación del contexto y entorno (para buscar rutas de escape y medios de defensa), memoria y aprendizaje (se recuerdan los estímulos peligrosos —sus características y tipo de situaciones donde se encuentran— y se aprenden formas de evitarlos o enfrentarlos) (LeDoux, 2000; Gallagher, 2000; Cosmides y Tooby, 2008, 2000).
III. Propiedades o rasgos de su arquitectura	III. Propiedades o rasgos de su arquitectura
7. <i>Cuasi-modularidad</i> . <i>a. Especialización funcional</i> : También especificada como su propiedad biológica: <i>función adaptativa</i> . <i>b. Fallas características</i> : Se han comprobado, para	7. <i>Cuasi-modularidad</i> . <i>a. Especialización funcional</i> : También especificada como su propiedad biológica: <i>función adaptativa</i> . <i>b. Fallas características</i> : Lesiones en la amígdala

⁷⁷ Es muy importante el adverbio ‘potencial’. Es común que digamos que nos sentimos orgullosos de algo que hemos hecho a pesar de haberlo hecho a solas. En tal caso, ¿qué competencia de prestigio social se ha ganado aquí? ¿Cómo puede ganarse prestigio social a solas? Bueno, nos solemos sentir orgullosos, a pesar de estar solos, por trabajos o habilidades que pudieran ser valorados socialmente. Por eso es tan común que comuniquemos a los demás sobre las habilidades que desplegamos en solitarios, para mejorar nuestro prestigio social.

⁷⁸ Hay que recordar que la forma en que representan las emociones su *tema evolutivo*, en tanto que *valoraciones corporeizadas*, es por medio de los patrones de cambios fisiológicos que las caracterizan (*supra*, II, v.II, III).

<p>el caso de las experiencias emocionales en general (Chwalisz, et. al, 1988), que estas disminuyen cuando se dañan los centros cerebrales que permiten que la información de lo que sucede en el cuerpo llegue al sistema nervioso central: a saber, la ínsula. No hay estudios aún que muestren este tipo de disociación para el caso del orgullo en particular.</p> <p><i>c. Procesamiento automático y semi-automático:</i> El orgullo se activará de forma involuntaria cuando el individuo se percate de haber tenido éxito en alguna competencia de prestigio: como cuando un deportista se percata de que ha ganado una competencia. También se activará bajo cierta predisposición voluntaria: es lo que hace que las personas asistan a premiaciones en las que sus trabajos pueden ser reconocidos. Se anticipa la experiencia de orgullo.</p> <p><i>d. Construcción a partir de reglas y representaciones innatas:</i> Las reglas, en el caso del orgullo, son los mecanismos innatos (neurofisiológicos) que activan patrones específicos de cambios corporales ante la presencia del <i>tema evolutivo</i> del orgullo: éxito en competencias de prestigio⁷⁹. Si bien la neurofisiología de las emociones no está sujeta a control voluntario o modificaciones, el tipo de estímulos que activan la emoción están sujetos a actualizan constante a lo largo de la vida del individuos (véase más abajo, ‘recalibración’).</p> <p><i>e. Dependencia de estímulo:</i> El orgullo regularmente se activará (de no haber otro estímulo acuciante que lo impida) en situaciones que manifiesten el rasgo por el cual evolutivamente (éxito potencial en térmi-</p>	<p>dan lugar a una condición denominada <i>ceguera emocional</i>, cuyo síntoma principal son la disminución de la experiencia y de respuestas de miedo (Phelps, 2005; LeDoux, 2000a, Rolls, 2000).</p> <p><i>c. Procesamiento automático y semi-automático:</i> El miedo se activará de forma automática cuando el individuo perciba un objeto o situación de peligro: como cuando una serpiente sale al paso. También se activará bajo cierta predisposición voluntaria: como cuando caminamos por un vecindario que nos han dicho que es peligroso.</p> <p><i>d. Construcción a partir de reglas y representaciones innatas:</i> Esto refiere a los mecanismos innatos (neurofisiológicos) que activan los patrones específicos de cambios corporales ante el <i>tema evolutivo</i> del miedo: peligro⁸⁰. Si bien la neurofisiología del miedo es innegociable, el tipo de estímulos que lo activan están sujetos a actualizaciones constantes a lo largo de la vida del individuos (véase más abajo, ‘recalibración’).</p> <p><i>e. Dependencia de estímulo:</i> El miedo se activará (de no haber otro estímulo que capte la atención) en situaciones que exhiban peligro —el rasgo por el cual evolucionaron. Pero puede ser interrumpida su activación si se presenta otro estímulo o modificado por procesos cognitivos (como juicios y pensamientos). Por ejemplo, la conducta expresiva de miedo puede verse inhibida por lo que Ekman (2003) llama reglas de despliegue.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

⁷⁹ Este rasgo de la cuasi-modularidad parece implicar la universalidad del orgullo. Tracy y sus colegas (2010) recabaron evidencia a favor de esta tesis. Casi todas las culturas en todas partes del mundo parecen tener un término para el orgullo y mostrar conductas asociadas con éste en situaciones adecuadas.

⁸⁰ La universalidad requerida por este rasgo de la cuasi-modularidad, en el caso del miedo, ha sido constatada en diversos estudios de Paul Ekman (2003, 1999, 1994, 1973, 1972).

<p>nos sociales). Pero puede ser modificado o interrumpido por procesos cognitivos de otros módulos (como juicios o pensamientos). Es decir, puede ser sensible a procesos cognitivos. Por ejemplo, la conducta de orgullo puede verse inhibida parcialmente por consideración sociales.</p>	
<p>8. <i>Sistema de recalibración de estímulos:</i> Las situaciones particulares que activan el orgullo pueden cambiar a lo largo de la vida del individuo, si bien no puede cambiar el <i>tema evolutivo</i> de la emoción: éxito social. Esto explica que de niños experimentemos orgullo al resolver una multiplicación por primera vez, pero desaparezca esa emoción cuando crecemos. Mientras crecemos, son logros y capacidades muy distintas las que provocan nuestro orgullo. Lo único que no cambia, es el tema evolutivo. Cualquiera que sea su activador, tendrá que ver con éxito social, dominancia o prestigio.</p> <p>9. <i>Duración breve:</i> La emoción de orgullo durará mientras el estímulo o situación que la causa persista y/o sea percibido por el individuo experimenta la emoción. La inercia fisiológica puede hacer que la experiencia perdure por un tiempo más (minutos o quizás horas), aunque el estímulo o situación activadora ya no esté presente o se haya valorado de una forma diferente.</p>	<p>8. <i>Sistema de recalibración de estímulos:</i> Los estímulos activadores del miedo pueden cambiar a lo largo de la vida del individuo. No puede cambiar, por otro lado, el <i>tema evolutivo</i> del miedo; a saber, peligro. Es por esto que es posible superar fobias o miedos irracionales, y que lleguemos a temer cosas que antes no habíamos considerado o conocido.</p> <p>9. <i>Duración breve:</i> La emoción de miedo durará mientras el estímulo que la causa persista en el entorno y/o sea percibido por el individuo que experimenta la emoción. La inercia fisiológica puede hacer que la experiencia perdure por un tiempo más (minutos), aunque el estímulo o situación activadora ya no esté presente o se haya valorado de una forma diferente.</p>
<p>V. Propiedades fenomenológicas</p>	<p>V. Propiedades fenomenológicas</p>
<p>10. <i>Fenomenología:</i> El orgullo tiene una fenomenología característica. En términos de valencia, se puede considerar como <i>agradable</i>. Su función (Damasio, 1999, 2005) es favorecer ciertos procesos de aprendizaje. Por ejemplo, se aprenden mejor las conductas que redituaron en prestigio social (Tracy, <i>op. cit.</i>). Otra de sus funciones es enviar información a la memoria a corto plazo para procesamiento ul-</p>	<p>10. <i>Fenomenología:</i> El miedo tiene una fenomenología desagradable distintiva. Su función es favorecer procesos de aprendizaje y enviar información a la memoria de corto plazo para procesamientos ulteriores, si se requieren. Así es posible reconocer y aprender a evitar objetos o tipos de situaciones que involucran peligro a partir de la experiencia. También hace posible analizar la situación de peligro</p>

rior (Prinz, 2004). Por ejemplo, puede dar lugar a recuerdos agradables que motivan para ir en pos de las situaciones que los causan (Tracy, <i>op. cit.</i>).	actual y buscar formas adecuadas para enfrentarla.
<p>Todas las propiedades que figuran en esta tabla se hayan vinculadas por medio de mecanismos causales-homeostáticos que configuran lo que podríamos llamar un circuito emocional. Las propiedades neuronales (1) activan a las propiedades fisiológicas (2) y expresivo-conductuales (3). En las propiedades (1) se implementan y despliegan las propiedades o rasgos de la arquitectura de las emociones: cuasi-modularidad (7), sistema de recalibración (8) y, con ayuda de (2), regulan la duración de las emociones (9). El carácter representacional de las emociones (5) depende directamente de las propiedades (2) y (4). (4) depende de (1). Las propiedades fenomenológicas (10) dependen de (2) y (1) y, a su vez, hacen posibles, junto con (7 c y e), las propiedades psicológicas (6): efectos psicológicos característicos. Ahora bien, en tanto que racimo de propiedades homeostáticas, la información que fluye en el circuito no va en una sola dirección (de abajo hacia arriba). El circuito se retroalimenta y se actualiza constantemente (de arriba hacia abajo). Esto es posible gracias a que ciertos rasgos de la arquitectura como (8) y (7 c y e) y, junto con (19) y (6), pueden retroalimentar el circuito emocional, incidiendo en (5) y en (2) (3), mediante (1).</p>	

Esta tabla comparativa intenta mostrar que en principio es posible describir las propiedades, y explicar aproximadamente la forma en la que están vinculadas causalmente, tanto de una Emoción Básica (EB) como de una Emoción Cognitiva Superior (ECS), con el mismo marco teórico y conceptual. Utilizar el mismo modelo para explicar las emociones en general, atribuyendo las mismas propiedades y mecanismos causales, abona a favor de la tesis que ha sido el eje temático y se ha intentado defender en esta investigación:

Las emociones (ya sean Emociones Básicas o Emociones Cognitivas Superiores), comparten las mismas propiedades y mecanismos causales, lo que sugiere que constituyen una sola clase natural, donde clase natural se entiende como clase natural de Racimo de Propiedades Homeostáticas (RPH).

Hay que insistir en el carácter apenas *paradigmático* de la tabla anterior. Nuestro modelo se verá robustecido sólo en la medida en que el esquema que plantea sea satisfecho por todo aquello o por la mayoría de lo que llamemos emoción, independientemente de que sea una

considerada una EB o una ECS⁸¹. En este sentido, los estudios e investigación disponibles para cotejar nuestro modelo son bastante abundantes para el caso de las EB, de suerte que sería bastante plausible llevarlo a cabo de manera satisfactoria para el resto de las emociones básicas (Ekman, 2003, 1999, 1994). En el caso de las ECS, los datos e investigaciones son más bien incipientes. No obstante, parece haber suficiente información como para comenzar a realizar el mismo ejercicio con otras tantas ECS como la vergüenza (Clark, 2010), los celos de pareja (2010; Panksepp, 2010; Harmon-Jones, *et. al.*, 2009; Mellgren, *et. al.*, 2009) y el amor romántico (Matsumara, *et. al.*, 2011; Esch y Stafano, 2005; Aron, *et. al.*, 2004; Bartels y Semir, 2000). Otras emociones que comienzan a perfilar un cuerpo interesante de investigación e información en torno suyo son la compasión, la culpa y la indignación moral. El panorama para la confirmación del modelo que proponemos es, hasta donde alcanzamos a ver, halagüeño. Aunque no se descarta la eventual necesidad de ajustes y correcciones.

VIII.II. EXPANSIÓN Y CONTRACCIÓN DE LA ONTOLOGÍA EMOCIONAL

El modelo de teoría general de las emociones que hemos propuesto dará lugar, en algunos casos a la *expansión* de la ontología emocional. En otros, el análisis cuidadoso de la terminología emocional, dará lugar a la *contracción* de la misma. La *expansión* puede tener lugar en virtud de la *homología serial*. En II, VII.I, II, vimos que es posible que ECS como el orgullo y la vergüenza tengan contrapartidas Básicas, lo que nos alentaría a distinguir entre una vergüenza-CS y una vergüenza-B, y un orgullo-CS y un orgullo-B: ambos serían advocaciones de una misma emoción con propiedades neuronales, fisiológicas, expresivo-conductuales, etc., ligeramente distintas, pero —ya que comparten la misma historia evolutiva— con el *mismo tema evolutivo*. La posibilidad de que algo similar sea el caso respecto de otras ECS, no puede descartarse. Clark (2012 y 2010) apuesta por la idea de que tanto los celos como la culpa tengan una historia evolutiva como la de la vergüenza y el orgullo, de suerte que instancien homología serial y, por lo tanto, eventualmente tendríamos que hablar de culpa-B y celos-B, por un lado, y culpa-CS y celos-CS, por otro. Otra forma en la que podría tener lugar una expansión de la ontología emocional sería por medio de la *Estrategia de la Ramificación de las Emociones (ERE)* (II, VII, I), si es que llega a requerirse. La

⁸¹ Mantenemos la distinción entre EB y ECS por meros propósitos expositivos, no explicativos, obviamente. La distinción será de utilidad para abordar algunos expedientes pendientes.

estrategia en cuestión parece ser apropiada en casos en los que lo que se considera como una sola emoción instancia, en episodios particulares, patrones de activación neuronal, fisiológica y conductual, distintos. El miedo, por ejemplo, puede dar lugar a respuestas conductuales muy diferentes: quedarse inmóvil, escapar o pelear. Esto comporta que, en cada caso, se activaron mecanismos fisiológicos y neuronales distintos. Puede considerarse un rasgo más en el que posiblemente pueden diferir: alguna característica del estímulo que los causa. En el primer caso —inmovilidad— el peligro que se representa por medio del cuerpo es un peligro *inevitable* y *no-enfrentable*. En el segundo caso —escape— se trata de un peligro *evitable*. Y, en el tercer caso —pelea— de un tipo de peligro *inevitable* y *enfrentable*. Por otro lado, todos estos casos siguen teniendo algunos rasgos en común. Para empezar, solemos referirnos a ellos con el mismo nombre: ‘miedo’. Pueden involucrar la misma expresión facial característica y gritos de auxilio. Además, y lo más importante, todos representan el mismo *tema evolutivo* (peligro) y, por lo tanto, desplegarían la misma función adaptativa —hacer frente al peligro— aunque por distintos medios. Esta sería la razón principal para sugerir que el miedo, en realidad, se ramifica en tres subclases distintas que varían en términos de los patrones de activación neuronal, fisiológica y conductual. La *Estrategia de la Ramificación de las Emociones (ERE)* puede resultar viable para otras emociones. Prinz (2005/2010) considera la posibilidad de que otras emociones se presten para la dicha estrategia. Sugiere, por ejemplo, que la alegría se llegue a ramificar en tres subespecies, en función de las características del estímulo que las causa y las correspondientes diferencias de neuronales, fisiológicas, conductuales, etc., a las que darían lugar. En este sentido, una forma de alegría estaría asociada al placer sensitivo, otra a la satisfacción asociada con el logro, y una más quedaría vinculada con el gozo del juego.

Ya sea que se descubran homólogas seriales para ECS, o sea viable aplicar la *ERE* para alguna emoción, la implicación será la de una expansión de la ontología de las emociones. Cualquiera que sea el caso, no afecta al modelo de una teoría general de las emociones que hemos propuesto. En este sentido, si aceptamos el caso a favor de la homología serial entre el orgullo-B y el orgullo-CS (Clark, 2010a), y la evidencia que parece apoyar la ramificación del miedo en tres subespecies, seguiría siendo posible reelaborar la tabla comparativa anterior de suerte que se respetaran las sutiles diferencias entre las distintas advocaciones de orgullo y miedo. Los dos tipos de orgullo diferían, ligeramente, en relación a su

conducta expresiva y los mecanismos neuronal-fisiológicos (*supra*, II, v.II. I), mientras que seguirían compartiendo, aproximadamente, el resto de sus propiedades biológicas, psicológicas, de arquitectura y fenomenológicas. Algo similar funcionaría para el caso de las distintas subespecies de miedo. La taxonomía resultante de la expansión ontológica propiciada por la *homología serial* y/o la *ERE*, probablemente implicaría que no habría emociones *puras*, por decirlo de algún modo, en el sentido de que la mayoría resultaría, o tener homólogas básicas o tener distintas advocaciones en virtud del estímulo que las causa y las diferencias concomitantes en los patrones de cambios neurofisiológicos y conductuales.

En cuanto a la *contracción* de la ontología de las emociones, ésta tendría lugar en el marco de una revisión de la terminología emocional ordinaria, a la luz de nuestro modelo. En otras palabras, hay muchos términos emocionales que, según nuestra propuesta, no referirían a distinciones psicológicas reales. Por ejemplo, en castellano el término ‘ira’ puede asociarse con, pero también distinguirse de, los siguientes términos:

(1) Irritación. (2) Enfado. (3) Enojo. (4) Indignación. (5) Coraje. (6) Saña. (7) Crueldad. (8) Rabia. (9) Furia. (10) Furor. (11) Venganza. (12) Cólera.

Sin descartar *a priori* la posibilidad de que cada uno de estos términos emocionales pudiera, en efecto, estar asociado con pautas de activación específicas dentro de la matriz de las diez propiedades que conforman la estructura explicativa de las emociones que proponemos, una inspección más detallada puede sugerir que este no es el caso.

Para empezar, podríamos dudar de que los términos ‘saña’, ‘crueldad’ y ‘venganza’, remitan propiamente a estados emocionales. Más bien, parecen referirse a las características de la conducta de alguien que experimenta ira. El término ‘coraje’ parece ser sólo un sinónimo de ‘enojo’ en el que se pone énfasis en la persona que causó nuestro enojo. No necesariamente nos debería llevar a pensar que el término pretende remitir a una emoción distinta del enojo. El término ‘indignación’, por otra parte, tampoco parece denotar algo distinto del ‘enfado’ o ‘el enojo’, acaso apunta a la violación de algún precepto moral como la causa de nuestro enfado o enojo. Finalmente, sólo nos quedan los términos ‘irritación’, ‘enfado’, ‘enojo’, ‘rabia’, ‘furia’, ‘furor’ y ‘cólera’. Y sobre estos también podríamos dudar si remiten a estados emocionales distintos. Los términos ‘rabia’, ‘furia’ y ‘furor’ quizás no

denoten algo distinto a la cólera, excepto que con el primero parecen destacarse las experiencias fisiológicas, con el segundo la dirección hacia una persona u objeto, y con el tercero la conducta violenta. De ser este el caso, tan sólo nos quedarían los términos ‘irritación’, ‘enfado’, ‘enojo’ y ‘cólera’, junto con el término ‘ira’, como candidatos interesantes para realizar distinciones emocionales reales. Pero tampoco es tan prometedora esta posibilidad. Bien podríamos seguir pensando que todos los términos refieren a un mismo tipo de emoción —*v. gr.* la ira—, sólo que con distintos grados de intensidad: donde el término ‘irritación’ refiere a una forma de ira de baja intensidad, y el término ‘cólera’ a una ira más bien intensa. En tal caso, la *contracción* de la ontología emocional, tomando como marco el vocabulario emocional ordinario, sería radical. Estaríamos hablando, finalmente, de una sola emoción, a saber, la ira, que podría presentarse en distintos niveles de intensidad. No obstante, contamos con recursos para evitar una *contracción* tan radical de la ontología emocional. La *Estrategia de la Ramificación de las Emociones* podría advertirnos sobre la posibilidad de encontrarnos frente a un posible género de emoción que se ramifica en subespecies. La diferencia de intensidades entre la irritación, el enfado, el enojo y la cólera, dan lugar a respuestas conductuales distintas, las cuales deben obedecer a patrones de activación neural y fisiológica distintas. Esto sugiere que la ira es una emoción con subespecies, quizás cuatro, quizás más o quizás menos. Cómo sea, cualquiera que sea la estrategia que se elija, debe ser respaldada por investigación y evidencia empírica. Lo que es importante subrayar aquí es que un análisis similar puede llevarse a cabo para el caso de otras familias de términos emocionales⁸². Es posible descartar términos emocionales, como referentes de estados emocionales específicos, en virtud de que pueden referir, más bien, a la conducta asociada con tener un episodio emocional, o porque ponen énfasis en ciertos rasgo de la situación, o de lo que siente quien tiene la emoción o de a qué se dirige o de la forma en que se expresa, sin pretender aludir a una diferencia sustancial real. En pocas palabras, el nutrido léxico emocional, en muchos casos, no está hecho para reflejar distinciones psicológicas, sino para resaltar rasgos específicos de la propia emoción y del contexto⁸³. Esto

⁸² Por ejemplo, la familia de términos emocionales asociados con ‘tristeza’: pesar, amargura, congoja, miseria, depresión, etc.

⁸³ En comunicación personal, José Luis Díaz nos ha puesto al tanto de un conjunto de estudios, realizados por él mismo y otros investigadores (Díaz, 2007 y 2005; Díaz y Flores, 2001; Flores y Díaz, 2009; Mercadillo, Díaz y Barrios, 2007), que desde hace tiempo han avanzado en la dirección que aquí se sugiere. En dichos estudios, entre otras cosas, se identifican familias de términos emocionales en el castellano y se reconoce que,

significará, necesariamente, que no existirán tantas emociones como lo sugiere el vocabulario emocional ordinario. La revisión del mismo en los términos que acabamos de sugerir, implicará una *contracción* de la ontología emocional. Consideramos que, estar al tanto de esto, es crucial para llevar por buen camino una teoría general de las emociones.

más que apuntar a una distinción afectiva real, cada término parece apuntar a diferencias en términos de intensidad y función lingüística.

CONCLUSIONES

A lo largo de esta investigación hemos intentado argumentar, desde los flancos que hemos considerado pertinentes, a favor de la tesis (iii) que afirma que las emociones constituyen una clase natural. ¿Podemos concluir que hemos mostrado que es verdadera esta tesis? Difícilmente. Acaso podríamos conceder que, después de toda la discusión precedente, se ha hecho un poco más plausible. Y, ¿qué se ha hecho para aspirar a esta modesta conclusión?

En la primera parte, **Clases Naturales: Revisando la Tradición**, después de un recorrido histórico en torno a las distintas acepciones de ‘clase natural’, de deslindes de advocaciones nominalistas anti-realistas y esencialistas demasiado estrictas, discutimos y ofrecemos argumentos para justificar la adopción de una noción de ‘clase natural’ más liberal. De una noción que, al contrario de las demás, ofreciera un marco conceptual más idóneo para defender la tesis (iii). En este sentido, logramos identificar cómo ciertas características de una noción de ‘clase natural’, asociada con el *esencialismo metafísico* anidado en la semántica de ascendencia kripkeana, juegan un papel crucial para negar que las emociones —y los estados mentales en general, junto con otros fenómenos biológicos como las especies— pudieran constituir una clase natural. En concreto, notamos que dicha noción de ‘clase natural’ implica la idea de que las propiedades que constituyen la esencia de una clase natural, conforman un conjunto *bien delimitado y estático*. Estas características impiden que un fenómeno que exhiba indeterminación y cambios —a lo largo de la historia— en las propiedades que lo caracterizan cuente como una clase natural. En un marco conceptual tan estricto, naturalmente, a cosas como las emociones y las especies se les negará el estatus de clase natural. Nuestra oposición a esta noción de clase natural se basó en la idea de que tal noción está más bien fundada en escrúpulos definicionales que en la forma en la que la realidad está organizada. Consideramos que, si bien la realidad puede cortar fino cuando se trata de fenómenos y entidades de niveles más básicos —físico y químico—, adecuándose a ciertos escrúpulos definicionales, en otros niveles —biológico y psicológico— puede presentar indeterminación y cambio. Por esta razón, es necesario reformular la noción de ‘clase natural’, de suerte que logre captar dichos aspectos de la realidad. En esta dirección, optamos por adoptar la noción, más liberal y

ontológicamente adecuada, de clase natural de Racimo de Propiedades Homeostáticas (*RPH*) que propone Robert Boyd (1979/2007, 1991, 1999). Las clases naturales *RPH*, tienen una estructura explicativa constituida por un racimo —posiblemente indeterminado y dinámico— de propiedades físicas que se hayan vinculadas por mecanismos homeostático-causales. Esto significa que la presencia de unas favorecerá la presencia de otras y, en conjunto, definirán la matriz causal y explicativa de la clase natural. Esta concepción de clase natural proporciona un marco conceptual más *ad hoc* para entender y explicar fenómenos cuyas propiedades características presentan cierta vaguedad y cambios en el tiempo, como las emociones, las especies, las enfermedades, etc. Esta transición, de una noción esencialista estricta de clase natural a una más liberal y ontológicamente respetable, fue un paso importante hacia la defensa de la tesis (iii).

Los siguientes pasos para hacer plausible dicha tesis, la cual afirma que las emociones conforman una clase natural, se dieron en la segunda parte de este trabajo: **Emociones y Clases Naturales**. Comenzamos analizando el caso a favor de la tesis (ii) más limitada que sostiene que —sólo— las Emociones Básicas (EB) constituyen una clase natural —la cual implica que (iii) es falsa. Siguiendo a Griffiths (1997) encontramos buenas razones y argumentos para sustentar dicha tesis. Notamos que hay evidencia interesante para sostener que todas las emociones básicas instancian, aproximadamente, las mismas propiedades: neurofisiológicas, expresivo-conductuales y de arquitectura —modularidad fodoriana. Donde no seguimos a Griffiths, es en la conclusión que deriva de dicho escenario. Según Griffiths, las Emociones Cognitivas Superiores no pertenecen a la misma clase natural que las EB, ya que, según él, ninguna exhibe las propiedades que distinguen y constituyen la estructura explicativa —entendida en términos de clase natural *RPH*— que comparten las últimas. Lo que concluye Griffiths es que las EB y las ECS no conforman un único dominio explicativo y ontológico. Las emociones, EB y ECS, no conforman una misma categoría psicológica. No conforman una clase natural. Insatisfechos, obviamente, con tal veredicto, nos dimos a la tarea de ofrecer argumentos y evidencia para replicar dicha conclusión. Nos embarcamos en la empresa de proponer una teoría de las emociones que mostrara, al menos, que es viable defender la tesis (iii): que las emociones —ya sea EB o ECS— sí conforman una clase natural. En este sentido, de entrada consideramos que hay más propiedades asociadas con las emociones de las que se reconocen para el caso de las

EB, por lo que resultaría necesario ampliar el racimo de propiedades físicas que constituyen la estructura explicativa de las emociones. De tal suerte, a partir de nuestra teoría, se identifican diez propiedades —de distintos niveles explicativos— físicas relevantes como constitutivas de la estructura explicativa de las emociones:

- (1) Propiedades neuronales.
- (2) Propiedades fisiológicas.
- (3) Propiedades expresivo-conductuales.
- (4) Propiedades adaptativas.
- (5) Propiedades representacionales.
- (6) Efectos psicológicos característicos.
- (7) Propiedades o rasgos de arquitectura —cuasi-modularidad.
- (8) Sistema de recalibración.
- (9) Duración breve.
- (10) Fenomenología.

En tanto que propiedades de una clase natural *RPH*, intentamos mostrar que todas están vinculadas homeostáticamente, esto es, que la presencia de algunas favorece y hace posible la presencia de las demás. También advertimos que, ya que es posible que, en casos particulares, alguna de ellas no se instancie —dado el carácter indeterminado de las clases naturales *RPH*— esto no debería considerarse como un contraejemplo a nuestra teoría.

Mientras caracterizamos y argumentamos a favor de la inclusión de cada una de las propiedades mencionadas en la estructura explicativa de las emociones, ofrecíamos evidencia de que diversas emociones, tanto EB como ECS, instanciaban de hecho varias o la mayoría de dichas propiedades. En el caso concreto del orgullo —una emoción que se ha considerado ECS— y el miedo —considerada una de las seis EB—, construimos un cuadro comparativo que mostraba que era posible identificar en ambas las diez propiedades que propusimos como constitutivas de la estructura explicativa de las emociones. Si la evidencia, argumentos y la aplicación concreta de la teoría a emociones específicas no son del todo inadecuados, podríamos decir que, en efecto, hemos abonado lo suficiente como para hacer plausible la tesis (iii): las emociones conforman una clase natural *RPH*, constituyen un solo dominio explicativo y ontológico, toda vez que parecen compartir una misma estructura explicativa.

Finalmente, aunque la teoría de las emociones que proponemos podría conllevar una eventual revisión y expansión de la ontología emocional —dada la posibilidad que abre de hallar homólogas Básicas de varias ECS y de ramificación en subespecies de emociones conocidas— esto no malogra en absoluto nuestra defensa de la tesis (iii). Como se ha visto,

las homólogas básicas de ECS y las subespecies de emociones conocidas parecen instanciar las propiedades que constituyen la estructura explicativa de las emociones.

Este trabajo tiene como título, *Emociones: Hacia el Descubrimiento de una Clase Natural*, y pudo haberse desarrollado con argumentos mucho mejores, más destreza expositiva y mejor estructura. No obstante, nos parece que hemos ofrecido algunas buenas razones para pensar que —si partimos de bases conceptuales y teóricas adecuadas, e interpretamos y organizamos la evidencia disponible de forma apropiada— la investigación empírica de las emociones en los últimos años apunta al descubrimiento de una clase natural.

REFERENCIAS

- Adolphs, Ralph y Tranel, Daniel** (2000), 'Emotion Recognition and the Human Amygdala', en Aggleton (2000).
- Aggleton, John** (ed.) (2000), *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*, O.U.P., New York.
- Aggleton, John y Young, Andrew** (2000), 'The Enigma of the Amygdala: On its Contributions to Human Emotion', en Lane y Nadel (2000).
- Aron, Arthur; Fisher, Helen; Mashek, Debra; Strong, Greg; Li, Haifang y Brown, Lucy** (2005), 'Reward, Motivation and Emotion Systems Associated with Early-Stage Intense Romantic Love', *Journal of Neurophysiology*, DOI: 10.1152/jn.00838.2004
- Ax, Albert F** (1953), 'The Physiological Differentiation between Fear and Anger in Humans' en *Psychosomatic Medicine*, Vol. 15, No. 5, pp. 433-442.
- Barret, Lisa-Feldman** (2006), 'Are Emotions Natural Kinds?', *Perspectives on Psychological Science*, Vol. 1, No. 1, pp. 28-57.
- Barret, Lisa-Feldman** (2005), 'Feeling is Perceiving: Core Affect and Conceptualization in the Experience of Emotion' en Barret, *et. al.* (2005).
- Barret, Lisa-Feldman; Niedenthal, Paula y Winkielman, Piotr** (eds.) (2005), *Emotion and Consciousness*, The Guilford Press, New York.
- Bartels, Andreas y Zeki, Semir** (2000), 'The Neural Basis of Romantic Love' en *NeuroReport*, Vol.11, No. 17, pp. 3829-3834.
- Barton, Robert y Aggleton, John** (2000), 'Primate Evolution and the Amygdala' en Aggleton (2000).
- Bechtel, William. y Mundale, Jennifer.** (1999), 'Multiple Realizability Revisited: Linking Cognitive and Neural States' en *Philosophy of Science*, No. 66, pp. 175-207.
- Ben-Ze'ev, Aaron** (2000), *The Subtlety of Emotion*, MIT Press, Cambridge, Mass.
- Bermudez, José Luis** (2005), *Philosophy of Psychology: a Contemporary Introduction*, Routledge, New York.

- Bickle, John** (2006), 'Multiple Realizability' en *Stanford Encyclopedia of Philosophy*: <http://plato.stanford.edu/entries/multiple-realizability/#NeuSuc>
- Borod, J.** (ed.) (2000), *Handbook of Neuropsychology of Emotion*, O.U.P., New York.
- Boyd, Richard** (1999), 'Homeostasis, Species, and Higher Taxa', en Wilson, R. (1999), *Species: New interdisciplinary essays*, MIT Press, Cambridge Mass.
- Boyd, Richard** (1992), 'Constructivism, Realism, and Philosophical Method', en Earman (1992).
- Boyd, Richard** (1991), 'Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds' en *Philosophical Studies*, No. 61, pp. 127-148.
- Boyd, Richard** (1988/2007)¹, 'How to be a Moral Realist?' en Shafer-Landau y Cuneo (2007), *Foundations of Ethics*, Blackwell, Malden, Mass.
- Boyd, Richard** (1979/1993), 'Metaphor and Theory Change: What is "Metaphor" a Metaphor for?' en Ortony (1979/1993).
- Bradley, M.M. y Lang, P.J.** (2000), 'Measuring Emotion: Behavior, Feeling, and Physiology' en Lane y Nadel (2000), *Cognitive Neuroscience of Emotion*, OUP, New York.
- Cannon, Walter** (1915), 'Cambios Corporales en el Dolor, Hambre, Temor y Cólera' en Solomon y Calhoun (1984).
- Clark, Jason** (2012), 'Integrating Basic and Higher-Cognitive Emotions within a Common Evolutionary Framework: Lessons from the Transformation of Primate Dominance into Human Pride' en *Philosophical Psychology*, pp. 1-24.
- Clark, Jason** (2010), 'Relations of Homology Between Higher Cognitive Emotions and Basic Emotions', en *Biology Philosophy*, Vol. 25, No. 1, pp. 75-94.
- Clark, Jason** (2010a), 'Hubristic and Authentic Pride as Serial Homologues: The Same but Different', en *Emotion Review*, No. 2, pp. 397-8.
- Cosmides, Leda y Tooby, John** (2008), 'The Evolutionary Psychology of the Emotions and their Relationship to Internal Regulatory Variables', en Lewis, *et. al.* (2008).

¹ El de la izquierda es el año original de publicación y el de la derecha es el año de la edición consultada.

Cosmides, Leda y Tooby, John (2000), 'Evolutionary Psychology and the Emotions' en Lewis y Haviland-Jones (2000).

Couch, Mark (2004), 'A Defense of Bechtel and Mundale', en *Philosophy of Science*, Vol. 72, No. 2, pp. 198-204.

Chalmers, David (1996/1999), *La Mente Consciente: en busca de una Teoría Fundamental*, Gedisa, Barcelona.

Charland, Louis (2002), 'The Natural Kind Status of Emotion', en *British Journal of Philosophy of Science*, Vol. 53, pp. 511-537.

Churchland, Paul (1995), *The Engine of Reason, the Seat of the Soul*, MIT Press, Cambridge, Mass.

Chwalisz, Kathleen; Diener, Ed y Gallagher, Denis (1988), 'Autonomic arousal feedback and emotional experience: Evidence from the spinal cord injured', en *Journal of Personality and Social Psychology*, Vol. 54, No. 5, pp. 820-828.

Damasio, Antonio (1994/2006), *El Error de Descartes: La Emoción, la Razón y el Cerebro Humano*, Crítica, Barcelona.

Damasio, Antonio (2000), 'A Second Chance for Emotion', en Lane, Richard y Nadel, Lynn (eds.) (2000).

Damasio, Antonio (2005), *En Busca de Spinoza: Neurobiología de la Emoción y de los Sentimientos*, Crítica, Barcelona.

Damasio, A.; Grabowski, T. J.; Bechara, A.; Damasio, H.; Ponto, L. L. B.; Parvizi, J. y Hichwa, R. D. (2000), 'Subcortical and Cortical Brain Activity During the Feeling of Self-Generated Emotions', en *Nature Neuroscience*, No. 3, Vol. 10, pp. 1049-1056.

Damasio, Antonio; Tranel, Daniel y Damasio, Hana (1998), 'Somatic Markers and the Guidance of Behavior', en Jenkins y Oatley (1998).

de Sousa, Ronald (1987), *The Rationality of Emotion*, MIT Press, Cambridge, Mass.

Díaz, José Luis (2005), 'La identidad del dolor: ¿lesión, congoja, lamento o neuromatriz?' en *Salud Mental*, 24 (4), pp. 25-35..

- Díaz, José Luis** (2007), *La conciencia viviente*, F.C.E., México.
- Díaz, José Luis y Flores Gutiérrez, Enrique** (2001), 'La estructura de la emoción humana: un modelo cromático del sistema afectivo' en *Salud Mental*, 28 (2), pp. 13-32.
- Dixon, Thomas** (2003), *From Passions to Emotions: the Creation of a Secular Psychological Category*, C.U.P. Cambridge, U.K.
- Donellan, Keith** (1983), 'Kripke and Putnam on Natural Kind Terms' en Ginet y Shoemaker (eds.) (1983).
- Dreyfus, Hubert** (1999), *What Computers Still can't Do: a Critique of Artificial Reason*, M.I.T. Press, Cambridge, Mass.
- Duque, Jorge Eduardo; Moscoso, Óscar Hernán y Devia Cubillos, Alfonso** (2004), 'El Lóbulo Insular: un Lóbulo de Procesamiento Cortical Visceral' en *Acta Neurológica Colombiana*, Vol. 20, No. 2, pp. 90-2.
- Dupré, John** (2001), 'Natural Kinds', en Newton-Smith (2001).
- Dupré, John** (1993), *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*, Harvard University Press, Harvard.
- Earman, John** (ed.) (1992), *Inference, Explanation, and other Philosophical Frustrations*, University of California Press, Berkeley.
- Ekman, Paul** (2003), *Emotions Revealed*, Henry Holt Books, New York.
- Ekman, Paul** (1999), 'Basic Emotions', en Dalglish y Power (1999), *The Handbook of Cognition and Emotion*, John Wiley and Sons, New York.
- Ekman, Paul** (1994), 'Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique', *Psychological Bulletin*, Vol. 15, No. 2, pp. 268-287.
- Ekman, Paul** (1992), 'An Argument for Basic Emotion', *Cognition and Emotion*, 6, pp. 169-200.
- Ekman, Paul** (1973), 'Cross-Cultural Studies of Facial Expression', en Ekman (1973), *Darwin and Facial Expression: A Century of Research on Review*, Academic Press, New York.

- Ekman, Paul** (1972), 'Universals and Cultural Differences in Facial Expression of Emotion' en Cole, J. (ed.), *Nebraska Symposium on Motivation*, University of Nebraska Press, Nebraska.
- Ekman, Paul y Friesen, Wallace** (1986), 'A New Pan-Cultural Facial Expression of Emotion', en *Motivation and Emotion*, Vol. 10, pp. 159-168.
- Ekman, Paul y Friesen, Wallace** (1971), 'Constants across Cultures in the Face and Emotion', en *Journal of Personality and Social Psychology*, Vol. 17, pp. 124-129.
- Ekman, Paul; Levenson, Robert y Friesen, Wallace** (1983), 'Autonomic Nervous System Activity Distinguishes among Emotions' en *Science*, Vol. 221, No. 4616, pp. 1008-1010.
- Ellis, Brian** (2001), *Scientific Essentialism*, CUP, Cambridge, Mass.
- Elster, Jon** (1999/2001), *Sobre las Pasiones*, Paidós, Barcelona.
- Esch, Tobias y Stefano, George** (2005), 'The Neurobiology of Love' en *Neuroendocrinology Letters*, No. 3. Vol. 26, pp. 175-192.
- Evans, Dylan** (2001), *Emociones: la Ciencia del Sentimiento*, Taurus, Madrid.
- Faucher, Luc y Tapolet, Christine** (eds.), (2006), *The Modularity of Emotion*, en *Canadian Journal of Philosophy*, Volumen suplementario 32.
- Flores Gutiérrez, Enrique y Díaz, José Luis** (2009), 'La respuesta emocional a la música: atribución de términos de la emoción a segmentos musicales', en *Salud Mental* 32 (1), pp. 21-34.
- Fodor, Jerry** (1983), *The Modularity of Mind: An Essay in Faculty Psychology*, MIT Press, Cambridge, Mass.
- Fourie, M.M.; Rauch, H.G.; Morgan, B.E.; Ellis, G.F.; Jordaan, E.R. y Thomas, K.G.** (2011), 'Glut and Pride are Heartfelt, but not Equally so', en *Psychophysiology*, No. 48, Vol. 7, pp. 889-899.
- Frank, Robert** (1988), *Passions within Reason: The Strategic Role of the Emotions*, Norton, New York.

- Gallagher, Michela** (2000), 'The Amygdala and Associative Learning', en Aggleton (2000).
- Ghiselin, Michael** (1978), 'A Radical Solution to the Species Problem', *Systematic Zoology*, 23, pp. 536-544.
- Ginet, Carl y Shoemaker, Sidney** (eds.) (1983), *Knowledge and Mind: Philosophical Essays*, O.U.P., Oxford.
- Goodman, Nelson** (1978), *Ways of Worldmaking*, Hackett, Indianapolis.
- Griffiths, Paul** (1997), *What Emotions Really Are: The Problem of Psychological Categories*, University of Chicago Press, Chicago.
- Griffiths, Paul y Sacarantino, Andrea** (2009), 'Emotions in the Wild: the Situated Perspective on Emotion' en Robbins y Aydede (2009).
- Hacking, Ian** (1991/2007), 'A Tradition of Natural Kinds', en Lange (2007).
- Hacking, Ian** (2007), 'Putnam's Theory of Natural Kinds and their Names is not the same as Kripke's' en *Principia*, Vol. 11, No. 1, pp. 1-24.
- Hansberg, Olbeth** (1990), '¿Por qué los Estados Mentales no son Clases Naturales?' en *Dianoia*, No. 36. pp. 213-27.
- Hansberg, Olbeth** (1996), *La Diversidad de las Emociones*, F.C.E., México.
- Harmon-Jones; Peterson, Carly y Harris, Christine** (2009), 'Jealousy: Novel Methods and Neural Correlates' en *American Psychological Association*, No. 1, Vol 9, pp. 113-117.
- Hart, Sibyl y Legerstee, María** (eds.) (2010), *Handbook of Jealousy: Research and Multidisciplinary Approaches*, Blackwell, Mass.
- Hatzimoysis, Anthony** (ed.) (2003), *Philosophy and the Emotions*, C.U.P., Cambridge.
- Hull, David** (1978), 'A Matter of Individuality', *Philosophy of Science*, No. 45, pp. 335-60.
- Hsu, Ming; Anen, Cédric y Quartz, Steven** (2008), 'The Right and the Good: Distributive Justices and Neural Encoding of Equity and Efficiency' en *Science*, Vol. 320, No. 5879, pp. 1092-1095.

James, William (1884), '¿Qué es una Emoción', en Solomon, Robert & Calhoun, Cheshire (1984).

Jenkins, Jennifer y Oatley, Keith (eds.), (1998), *Human Emotions: a reader*, Blackwell, Mass.

Jones, Karen (2006), 'Quick and Smart? Modularity and the Pro-Emotion Consensus', en Faucher y Tappolet (2006).

Kenny, Anthony (1963), *Action, Emotion and Will*, Routledge and Kegan Paul, London.

Kripke, Saul (1981/1995), *El Nombrar y la Necesidad*, IIF/UNAM, México.

Kukla, André (2000), *Social Constructivism and the Philosophy of Science*, Routledge, Londres.

Lamm, Claus y Singer, Tania (2010), 'The Role of Anterior Insular Cortex in Social Emotions', en *Brain, Structure & Function*, 214, pp. 579-5991.

Lane, Richard y Nadel, Lynn (eds.) (2000), *Cognitive Neuroscience of Emotion*, O.U.P., New York.

Lange, Marc (2007), *Philosophy of Science: An Anthology*, Blackwell, Oxford.

LaPorte, Joseph (2004), *Natural Kinds and Conceptual Change*, CUP, New York.

Larsen, Jeff; Berntson, Gary; Poehlmann, Tiffany y Cacioppo, John (2008), 'The Psychophysiology of Emotion', en Lewis, *et al.*, (2008)

Lazarus, Richard (1991), *Emotion and Adaptation*, OUP, New York.

LeDoux, Joseph (2000), 'Cognitive-Emotional Interactions: Listen to the Brain', en Lane y Nadel (2000).

LeDoux, Joseph (2000a), 'The amygdala and emotion: a view through fear', en Aggleton (2000).

LeDoux, Joseph (1996), *The Emotional Brain: the Mysterious Underpinnings of Emotional Life*, Simon & Schuster, New York.

LeDoux, Joseph y Phelps, Elizabeth (2000), 'Emotional Networks in the Brain' en Lewis, *et al.*, (2000)

- Levenson, Robert** (2003), 'Blood, Sweat and Fears: The Autonomic Architecture of Emotion', en *Annals of the New York Academy of Sciences*, No. 1000, pp. 348-366.
- Levenson, Robert; Ekman, Paul y Friesen, Wallace** (1990), 'Voluntary Facial Action Generates emotion-specific Autnomic Nervous System Activity', en *Psychophysiology*, Vol. 27, No. 4., pp. 363-384.
- Levenson, Robert; Ekman, Paul; Heider, Karl y Friesen, Wallace** (1992), 'Emotion and Autonomic Nervous System Activity in the Minangkabau of West Sumatra', en *Journal of Personality and Social Psychology*, Vol. 62, No. 6, pp. 972-988.
- Lewis, Michael y Haviland-Jones, Jeannette** (eds.) (1993), *The Handbook of Emotions*, Guilford Press, New York.
- Lewis, Michael y Haviland-Jones, Jeannette** (eds.) (2000), *The Handbook of Emotions*, Guilford Press, New York.
- Lewis, Michael; Haviland-Jones, Jeannette y Barrett, Lisa-Feldman** (2008), *Handbook of Emotions*, Guilford Press, New York.
- Loar, Brian** (1990), 'Phenomenal states', en Chalmers (ed.) (2002).
- Locke, John** (1690/1999), *Ensayo sobre el Entendimiento Humano*, F.C.E., México.
- Lyons, William** (1980), *Emotion*, C.U.P., Cambridge, U.K.
- Mackie, Penelope** (2006), *How Things might have been: Individuals, Kinds, and Essential Properties*, OUP, New York.
- Markova, Gabriela; Stieben James y Legerstee, Maria** (2010), 'Neural Structures of Jealousy: Infants' Experiences of Social Exclusion with Caregivers and Peers' en Hart, *et. al.*(2010).
- Matsumara, Kenta; Yamakoshi, Takehiro y Rolfe, Peter** (2011), 'Love Styles and Cardiovascular Responses Types', *International Journal of Psychological Studies*, No. 2, Vol. 3, pp. 21-28.
- Matsumoto, David; Kasri, Fazilet y Kooken, Kristie** (1999), 'American-Japanese cultural differences in judgments of expression intensity and subjective experience' en *Cognition and Emotion*, Vol. 13, No. 2, pp. 201-218.

- Matsumoto, David; Keltner, Dacher; Shiota, Michelle; O'sullivan, Maureen y Frank, Mark** (2008) 'Facial Expressions of Emotion' en Lewis, *et. al.* (2008).
- Mellgren, Roger; Hromatko, Ivana y McArthur, Deborah** (2009), 'A Test of Evolutionary Explanation of Jealousy in the United States and Croatia', en *Journal of General Social Issues*, No. 19, Vol. 6, pp. 915-931.
- Mellor, D. H.** (1977), 'Natural Kinds' en *British Journal for the Philosophy of Science*, Vol. 28, No. 4, pp. 299-312.
- Mercadillo, Roberto., Díaz, José Luis, Barrios, Fernando.** (2007), 'Neurobiología de las emociones morales', en *Salud Mental*, 30 (3), pp. 1-11.
- Nadel, Lynn** (1992), 'Multiple Memory Systems: What and Why', en *Journal of Cognitive Neuroscience*, Vol. 4, No. 3, pp. 179-188.
- Nagel, Thomas** (1986/1998), *Una Visión de Ningún Lugar*, F.C.E, México.
- Newton-Smith, William Herbert** (ed.), (2001), *A Companion to the Philosophy of Science*, Blackwell, Londres.
- Nussbaum, Martha** (2001), *Upheavals of Thought*, C.U.P., Cambridge.
- Neu, Jerome** (2000), *A Tear is an Intellectual Thing: The Meanings of Emotion*, OUP, Oxford.
- Öhman, Arne** (1993), 'Fear and Anxiety as Emotional Phenomena: Clinical Phenomenology, Evolutionary Perspectives, and Information Processing' en Lewis, *et. al.* (1993).
- Öhman, Arne; Fredrickson, Mats; Hughdal, Kenneth y Rimmo, Per-Arne** (1976), 'The Premise of Equipotentiality in Human Classical Conditioning: Conditioned Electrodermal Responses to Potentially Phobic Stimuli' en *Journal of Experimental Psychology*, Vol. 103, No. 4, pp. 313-337.
- Ortony, Andrew.** (ed.), (1979/1993), *Metaphor and Thought*, C.U.P., Cambridge, Mass.
- Panksepp, Jaak** (2010), 'The Evolutionary Sources of Jealousy' en Hart, *et al.* (2010)

Panksepp, Jaak (2008), 'The Affective Brain and Core Consciousness: How does Neural Activity Generates Emotional Feelings?', en Lewis, *et. al.* (2008)

Panksepp, Jaak (2000), 'Emotions as Natural Kinds within the Mammalian Brain', en Lewis y Haviland-Jones (2000).

Panksepp, Jaak (1998), *Affective Neuroscience*, OUP, Oxford.

Papineau, David (2002), *Thinking about consciousness*, Clarendon Press, Oxford.

Penrose, Roger (1996), *La Nueva Mente del Emperador*; Grijalbo, Barcelona.

Phan, Luan; Wager, Tor; Taylor, Stephan y Liberson, Israel (2002), 'Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI', en *NeuroImage*, No. 16, pp. 331-348.

Phelps, Elizabeth (2005), "The Interaction of Emotion and Cognition: Insights from the Studies of the Human Amygdala" en Barret, *et. al.* (2005).

Platts, Mark (1997: 2nd ed.), *Ways of Meaning: Introduction to a Philosophy of Language*, Routledge & Kegan Paul, London.

Plutchik, Robert (2001), 'The Nature of Emotions' en *American Scientist*, Vol. 89, No. 4, pp. 344-350.

Plutchik, Robert (1984), *Emotion: A Psychoevolutionary Synthesis*, Harper and Row, New York.

Prinz, Jesse (2012), *The Conscious Brain: How Attention Engenders Experience*, O.U.P., New York.

Prinz, Jesse (2006a), 'Is Emotion a form of Perception?' en Faucher y Tapolett (2006).

Prinz, Jesse (2006b), 'Is the Mind Really Modular?', en Stainton (2006), *Contemporary Debates in Cognitive Science*, Blackwell, Oxford.

Prinz, Jesse (2005/2010), '¿Cuáles son las emociones básicas?', *Cuadernos de Crítica*, no. 55, UNAM-IIF, México.

Prinz, Jesse (2004), *Gut Reactions: A Perceptual Theory of Emotion*, O.U.P., New York.

Prinz, Jesse (2003a), 'Embodied Emotions', en Solomon (2003).

- Prinz, Jesse** (2003b), 'Emotion, Psychosemantics, and Embodied Appraisals', en Hatzimoysis (2003).
- Putnam, Hilary** (1975/1984), 'El significado de "significado"', *Cuadernos de Crítica* 28, IIF/UNAM, México.
- Putnam, Hilary** (1978/1991), *El Significado y las Ciencias Morales*, IIF-UNAM, México.
- Quine, Wilard Van Orman**, (1969/1975), 'Géneros Naturales', en Quine, W.V.O., *Relatividad Ontológica y otros Ensayos*, Tecnos, Madrid.
- Quinn, Paul** (1998), 'Visual pop-out in young infants: Covergent Evidence and Extension' en *Infant Behavior and Development*, Vol. 21, No. 2, pp. 273-288.
- Rolls, Edmund** (2005), *Emotions Explained*, O.U.P., New York.
- Rolls, Edmund** (2000), 'Neurophysiology and Functions of the Primate Amygdala, and the Neural Basis of Emotion' en Aggleton (2000).
- Robbins, Phillip y Aydede, Murat** (eds.) (2009), *Cambridge Handbook of Situated Cognition*, C.U.P., Cambridge.
- Rorty, Amelie** (1980), *Explaining Emotions*, University of California Press, Berkeley.
- Rorty, Amelie** (1978), 'Explaining Emotions', *The Journal of Philosophy*, Vol. 75, pp. 139-171.
- Russell, Bertrand** (1948/1992), *El Conocimiento Humano: Su Alcance y Límites*, Ed. Planeta-De Agostini, Barcelona.
- Russell, James** (2006), 'Emotions are not Modules', en *Canadian Journal of Philosophy*, Vol. 36, Suplemento No. 32, pp. 53-71.
- Russell, James** (2003), 'Core Affect and the Psychological Construction of Emotion' en *Psychological Review*, Vol. 110, No. 1, pp. 145-172.
- Russell, James** (1994), 'Is there Universal Recognition of Emotion from Facial Expression? A Review of the Cross-Cultural Studies', en *Psychological Bulletin*, Vol. 15, No. 2, pp. 102-141.

- Russell, James** (1980), 'A Circumplex Model of Affect' en *Journal of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161-1178.
- Schachter, Joseph** (1957), 'Pain, fear, and anger in hypertensives and normotensives' en Mendels (1975), *The Psychobiology of Depression*, Spectrum, New York.
- Schachter, Stanley; Singer, Jerome** (1962), 'Determinantes Cognoscitivos, Sociales y Fisiológicos de Estado Emocional', en Solomon y Calhoun (1984).
- Scherer, Klaus; Banse, Rainer; Wallbott, Herald y Goldbeck, Thomas** (1991), 'Vocal Cues in Emotion Encoding and Decoding' en *Motivation and Emotion*, Vol. 15, No. 2, pp. 123-148.
- Scherer, Klaus; Banse, Rainer y Wallbott, Herald** (2001), 'Emotion Inference from Vocal Expression Correlate across Languages and Cultures' en *Journal of Cross-Cultural Psychology*, Vol. 32, No. 1, pp. 76-92.
- Schwartz, Gary; Weinberger, Daniel y Singer, Jefferson** (1981), 'Cardiovascular differentiation of happiness, anger, sadness, and fear following imagining and exercise' en *Psychosomatic Medicine*, Vol. 43, No. 4, pp. 343-64.
- Searle, John** (1980), 'Minds, Brains, and Programs', en *The Behavioral and Brain Sciences*, Vol. 3, No. 3, pp. 417-457.
- Shapiro, Lawrence** (2000), 'Multiple Realizations' en *The Journal of Philosophy*, Vol. 97. No. 12, pp. 635-54.
- Soames, Scott** (2002), *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*, O.U.P., Oxford.
- Soca, Ricardo** (2004), *La Fascinante Historia de las Palabras*, Rey + Naranjo Editores, Bogotá.
- Solomon, Robert** (2007), *Ética Emocional: una Teoría de los Sentimientos*, Paidós, Barcelona.
- Solomon, Robert** (2003), *Thinking about Feeling: Contemporary Philosophers on Emotion*, OUP, Oxford.

- Solomon, Robert** (1995), 'Some Notes on Emotion: "East and West"', *Philosophy East and West*, Vol. 45, No. 2, pp. 171-202.
- Solomon, Robert** (1976), *The Passions*, Hackett Publishing Co., Indianápolis.
- Solomon, Robert y Calhoun, Cheshire** (1984), *¿Qué es una Emoción?: Lecturas Clásicas de Psicología Filosófica*, F.C.E., México
- Stanford, Kyle y Kitcher, Philip** (2000), 'Refining the Causal Theory of Reference for Natural Kind Terms' en *Philosophical Studies*, Vol. 97, No. 1, pp. 97-127.
- Tassinari, Louis y Cacioppo, John** (1999), 'Unobservable Facial Actions and Emotion', en *Psychological Science*, Vol. 3, No. 1, pp. 28-33.
- Teroni, Fabrice** (2007), 'Emotions and Formal Objects' en *Dialectica*, Vol. 61, No. 3, pp. 395-415.
- Tracy, Jessica; Shariff, Azim y Cheng, Joey** (2010), 'A Naturalist's View of Pride' en *Emotion Review*, Vol. 2, No. 2., pp. 163-177.
- Tracy, Jessica y Robins, Richard** (2008), 'The Automaticity of Emotion Recognition', en *Emotion*, Vol. 8, No. 1, pp. 81-95.
- Tracy, Jessica y Robins, Richard** (2007), 'The Prototypical Pride Expression: Development of a Nonverbal Behavioral Coding System' en *Emotion*, Vol. 7, No. 4, pp. 789-801.
- Tracy, Jessica y Robins, Richard** (2004), 'Show your Pride: Evidence for a Discrete Emotion Expression' en *Psychological Science*, 15, pp. 194-197.
- Tucker, Don; Derryberry, Douglas y Luu, Phan** (2000), 'Anatomy and Physiology of Human Emotion: Vertical Integration of Brainstem, Limbic and Cortical Systems', en Borod (2000).
- Tye, Michael** (2000), *Consciousness, color and content*, MIT Press, Cambridge, Mass.
- Vrana, Scott; Spence, Ellen y Lang, Peter J.** (1988), 'The startle probe response: A new measure of emotion?' en *Journal of Abnormal Psychology*, 97, 487-491.

Waller, Bridget; Vick, Sara-Jane; Parr, Lisa; Bard, Kim y Pasqualini, Marcia; et al.
(2006), 'Intramuscular electrical stimulation of facial muscles in humans and chimpanzees:
Duchenne revisited and extended', en *Emotion*, Vol. 6, No. 3, pp. 367-82.