

Universidad de Chile
Facultad de Filosofía y Humanidades
Departamento de Filosofía

Una Alternativa para el Problema Mente-Cuerpo: Emergentismo y Cognición Corporizada y Situada

Tesis para optar al Grado de Licenciado en Filosofía

Alumno:

BERNARDO AGUILERA DREYSE

Profesor Guía: Guido Vallejos Oportot

Santiago, Diciembre de 2008

Epígrafe . .	4
RESUMEN . .	5
INTRODUCCIÓN . .	6
ALGUNOS ASPECTOS BÁSICOS PARA EL ESTUDIO DEL PROBLEMA MENTE CUERPO . .	7
1.1 Hacia una ontología para lo mental . .	7
1.2 Propiedades mentales intencionales y fenomenales . .	7
1.3 Causalidad de lo mental y materialismo . .	9
TEORÍA DE LA IDENTIDAD Y FUNCIONALISMO . .	12
IDENTIDAD DE LOS CASOS Y SUPERVENIENCIA . .	16
EL EMERGENTISMO BRITÁNICO . .	19
EMERGENTISMO DÉBIL Y CIENCIAS DE LA COMPLEJIDAD . .	24
EMERGENTISMO DÉBIL Y COGNICIÓN CORPORIZADA Y SITUADA: UNA ALTERNATIVA PARA EL PROBLEMA MENTE-CUERPO . .	31
CONSIDERACIONES FINALES . .	34
BIBLIOGRAFÍA . .	36

Epígrafe

“La idea cartesiana de una mente separada del cuerpo bien pudo haber sido el origen, a mediados del siglo XX, de la metáfora de la mente como un programa informático. En realidad, si la mente puede separarse del cuerpo quizá se pueda intentar comprenderla sin apelar a la neurobiología, sin necesidad alguna de ser influidos por el conocimiento de la neuroanatomía, la neurofisiología y la neuroquímica. Es interesante, y a la vez paradójico, que muchos científicos cognitivos que creen que pueden investigar la mente sin recurrir a la neurofisiología no se considerarían dualistas.”
(Damasio, 1996, pág. 230)

RESUMEN

En el presente trabajo se exploran alternativas para el problema mente-cuerpo, bajo el supuesto básico de preservar al materialismo y la causalidad de las propiedades mentales. Tras explicar los problemas que presentan las aproximaciones tradicionales a este problema –identidad, funcionalismo, superveniencia- se expone la propuesta del emergentismo británico. Finalmente se defiende una versión “débil” del emergentismo, que de la mano con un enfoque corporizado y situado de la cognición, puede responder algunos de los principales problemas que presentan las otras propuestas y constituir una alternativa para el problema mente-cuerpo.

INTRODUCCIÓN

Algunas de las preguntas filosóficas que más llaman la atención al sentido común son ¿qué es la mente?, ¿dónde está y qué relación tiene con el mundo físico?, ¿sobrevivirá mi mente una vez que mi cuerpo deje de hacerlo?. Una respuesta popular y antes indiscutida era que la mente consistía en algo distinto del cuerpo, incluso capaz de sobrevivirle a él. Sin embargo, conforme ha avanzado la ciencia y nuestra comprensión del mundo se hace más completa, esta respuesta ha perdido valor. Existe consenso dentro del mundo informado del hecho de que el mundo consta sólo de partículas físicas, agregadas con mayor o menor complejidad, y que incluso la mente tiene su origen en la biología de los organismos. De hecho, el estudio actual de la mente humana se lleva a cabo dentro de la ciencia cognitiva, la cual articula una explicación de lo mental a partir de evidencia empírica proveniente de distintas disciplinas. Sin embargo, las preguntas filosóficas persisten: ¿qué es la mente?, ¿cómo es posible que mi mente, tal como la experimento, exista en un mundo formado sólo por partículas físicas?, ¿cómo se explica que mi mente pueda causar mis pensamientos y llevarme a tomar decisiones que afectan al mundo?, etc. Estas preguntas se enmarcan dentro del llamado “problema mente-cuerpo”, uno de los temas centrales para la filosofía de la mente y la ciencia cognitiva.

El propósito de este trabajo será proponer una alternativa de solución para este problema, basada en un enfoque que tiene algo de retrógrado y algo de innovador: el emergentismo. Si bien esta tesis ya fue propuesta hace casi un siglo y muchos la asocian con nociones vitalistas ya obsoletas, ha sido retomada dentro de la ciencia cognitiva durante las últimas décadas, en el contexto de las denominadas ciencias de la complejidad, que asimismo se asocian con perspectivas de la cognición que surgen como alternativas a las teorías tradicionales. Para fundamentar mi propuesta, examinaré las alternativas tradicionales dadas para el problema mente-cuerpo, destacando sus problemas para luego evaluarlos en el contexto del emergentismo. Mi conclusión será que la mejor propuesta para solucionar estos problemas consiste en una síntesis entre una variante del emergentismo clásico y enfoques corporizados y situados de la cognición.

Comenzaré estableciendo los principales aspectos de los ámbitos mental y físico que deben tenerse en cuenta para la discusión posterior. Luego, en las secciones 2 y 3 discutiré las alternativas tradicionales y preponderantes para el problema mente-cuerpo, concluyendo la sección 3 con un listado de requisitos mínimos que debiese cumplir una alternativa a este problema para dar cuenta de sus aspectos más relevantes y conflictivos. En la sección 4 expondré el emergentismo de acuerdo a su formulación clásica, para luego desarrollar mi propuesta en las secciones 5 y 6. Luego presentaré algunas consideraciones finales que a modo de conclusión revisarán el listado de requisitos que presenté antes para así evaluar si constituye mi propuesta una alternativa plausible para el problema mente-cuerpo.

ALGUNOS ASPECTOS BÁSICOS PARA EL ESTUDIO DEL PROBLEMA MENTE CUERPO

1.1 Hacia una ontología para lo mental

Comenzaré refiriéndome a la terminología ontológica que emplearé en este trabajo, para así evitar confusiones derivadas del distinto uso que los autores hacen de estos términos. Esto puede parecer dogmático, en tanto no me detendré a justificar el empleo de tal o cual terminología, pero para las pretensiones de este trabajo bastará con que nos expresemos con claridad. Asumo que el mundo se compone de entidades, las cuales se caracterizan por tener propiedades. Estas propiedades se instancian en las entidades y pueden variar de un tiempo a otro, y esto determina que las entidades cambien o pasen de un estado a otro. Este cambio se produce por la acción de una propiedad o conjunto de propiedades, que causan que otra propiedad o conjunto de propiedades se instancien en una entidad.

Asumo también que existen dos clases de propiedades, las físicas y las mentales. Ejemplos de propiedades físicas: pesar 5 gramos, tener un contorno redondeado, tener un circuito neuronal de tal y tal característica. Ejemplos de propiedades mentales: percibir el color rojo, creer que lloverá mañana, sentir miedo. No discutiré aquí cómo se relacionan ambos tipos de propiedades. Sólo consideraré como un hecho el que, a diferencia de los entes inanimados, algunas especies animales poseen propiedades mentales, las que en conjunto se denominan “mente”. A partir de este dominio de propiedades mentales es posible, tal como lo hace la ciencia con las propiedades físicas, intentar establecer conexiones legaliformes y elaborar una teoría explicativa, que correspondería a una teoría de lo mental o psicología.

1.2 Propiedades mentales intencionales y fenomenales

Los filósofos además suelen distinguir dos clases de propiedades mentales: las propiedades mentales intencionales y las propiedades mentales fenomenales. Las primeras se caracterizan por estar dirigidas o referidas a un objeto, característica que se denomina intencionalidad, y que según la influyente tesis de Brentano suele constituir el aspecto distintivo de lo mental. Las propiedades mentales intencionales involucran una actitud del sujeto hacia su objeto intencional, el cual suele definirse como una proposición (una oración que es valorable como verdadera o falsa). Ejemplos son: *creer* que mañana es sábado, *desear* beber un vaso de agua, etc.

Las propiedades fenomenales, por su parte, se refieren a sensaciones o experiencias cualitativas, como p.ej. percibir un color, sentir miedo, etc. Estas propiedades se refieren

al modo como el sujeto experimenta su propia mentalidad, y por lo tanto para probar su existencia se apela a la introspección y no a datos empíricamente objetivables. Esto explica tanto el que a menudo sean obviadas por la psicología científica, como el que resulte dificultosa su caracterización, hasta el punto de que a la ciencia cognitiva tradicionalmente suele prescindir de ellos, bastándole sólo con los estados mentales intencionales para explicar la conducta humana. Sin embargo, dado lo diverso, evidente y omnipresente de nuestra experiencia consciente, asumiré que este aspecto fenomenal de la mente debe ser incluido dentro de una ontología y una ciencia de lo mental (para más argumentos al respecto, ver Chalmers (1996) y Searle (2004)). Más aún, creo que la división entre propiedades mentales intencionales y fenomenales, si bien explicativamente útil en algunos casos, resulta artificial dada la naturaleza de las propiedades mentales, por razones que expondré a continuación y en la sección 6.

En primer lugar, es posible distinguir el carácter intencional de las propiedades fenomenales. En general, las propiedades fenomenales suelen describirse como percepciones sensoriales que abarcan tanto los estímulos externos (p.ej. experimentar el color rojo de un objeto) como los internos, es decir, sensaciones del propio cuerpo (p.ej. experimentar dolor en un pie). Siguiendo a Crane (2001), es posible sostener que en estos casos, al igual que con las propiedades mentales intencionales, es posible distinguir una actitud del sujeto hacia un objeto intencional. Cada vez que experimentamos un dolor, existe un modo como nos dirigimos a un lugar del cuerpo donde localizamos ese dolor, que consiste en cómo se nos presenta ese dolor (p.ej. como un daño) y un objeto intencional (p.ej. un dedo del pie). Por lo tanto, la tesis de la intencionalidad que sostiene que todo estado mental está dirigido a un objeto, se aplica también para las percepciones sensoriales y las experiencias corpóreas, siendo entonces posible afirmar que todo el contenido mental tiene carácter intencional. En segundo lugar, postularé la tesis más fuerte de que, además de tener carácter intencional, toda propiedad mental es también fenomenal, es decir, el poseer una propiedad mental siempre involucra el experimentar de alguna manera esta propiedad. Esto puede describirse como “lo que se siente estar” en un estado mental, consistente en la instanciación de una o más propiedades mentales. Pienso que la tradicional aproximación a las propiedades mentales como independientes de su dimensión material, tal como lo hace el funcionalismo, incurre en el error de abstraer a estas propiedades de su principal nexo con el cuerpo: su carácter fenomenal (ver sección 6).

¿Qué razones hay para postular que toda propiedad mental tiene un carácter fenomenal?. Creo que pueden esgrimirse argumentos tanto desde la introspección como desde las neurociencias y la filogenia humana para defender esta posición. En la sección 6 me detendré sobre estos últimos. Respecto a los argumentos introspectivos, cabe señalar que la mente, tal como se experimenta en primera persona, constituye un propio campo de investigación que habitualmente se denomina fenomenología. A partir de ella puede afirmarse que la vida mental humana está invariablemente acompañada de una experiencia que es continua, unitaria, y senso-perceptivamente vinculada con la realidad corporal y ambiental que nos toca vivir. Diversas distinciones relativas a la conciencia pueden hacerse a través del método fenomenológico, las cuales escapan de los propósitos de este trabajo. Remito a Varela et al. (1997) para una completa y original exposición de este tema, y a Gallagher (2007) para una revisión de la tradición fenomenológica de la filosofía continental. Me interesa aclarar que al señalar la relevancia del aspecto fenomenal de los estados mentales no intento defender la preponderancia de una aproximación fenomenológica a la mente humana, ni un enfoque idealista y menos deslegitimar los aportes que la ciencia puede hacer al respecto. Mi afirmación es que para estudiar la mente humana considero relevantes los reportes que todo sujeto puede hacer respecto de su propia

mentalidad, gracias al acceso privilegiado que tiene a ésta, así como la preponderancia que la experiencia consciente entendida en un sentido corporizado y situado tiene en todo momento de la vida mental. En la sección 2 expondré otros argumentos de este tipo, a propósito de la irreductibilidad de los estados mentales.

Antes de concluir, consideraré las ventajas que este enfoque de las propiedades mentales puede tener para este trabajo. La primera de ellas es que puede permitir un abordaje para el denominado “problema difícil” de la conciencia (Chalmers, 1996). Según Chalmers y otros filósofos, la ciencia cognitiva usualmente se ha centrado en enfrentar los problemas “fáciles” respecto de la mente, como sería preguntarse por cómo el cerebro procesa o integra información. Sin embargo, un desafío mayor para la ciencia cognitiva sería responder cómo es posible que este procesamiento pueda estar acompañado de una experiencia interna. Creo que una posible aproximación a este “problema difícil” puede ser tanto esta visión de las propiedades mentales como esencialmente arraigadas a experiencias senso-perceptivas, así como una propuesta emergentista, todos ellos temas que desarrollaré más adelante.

La segunda ventaja que creo encontrar en considerar las propiedades mentales como fenomenales está en que esta aproximación puede constituir un modo de vincularlas con su dimensión material, y así contribuir para una alternativa al problema mente-cuerpo. Intentaré explicar esta vinculación a partir de un ejemplo: una bicicleta tiene la propiedad de transportar a un sujeto a su lugar de trabajo. Esta propiedad se basa en las propiedades físicas que tienen los componentes de la bicicleta. Supongamos ahora que cambiamos las llantas muy gastadas que tenían las ruedas de la bicicleta por otras nuevas. La bicicleta ahora sigue teniendo la misma propiedad de transportar al sujeto a su trabajo, y sin embargo tiene sentido que el sujeto encuentre que esta propiedad no es exactamente la misma. Esto lo expresaría el ciclista si dijese: “sí, me sigue transportando al mismo lugar que antes, pero ahora *se siente* distinto”. Así, si volvemos a las propiedades mentales, podríamos decir que una propiedad mental como tener un dolor o desear no mojarse con la lluvia no es la misma si es instanciada en un ser humano, un gorila o un extraterrestre (si es que los hay). Volveremos sobre este tema en otras partes de este trabajo, particularmente cuando tratemos con los problemas de la tesis de la realizabilidad múltiple.

1.3 Causalidad de lo mental y materialismo

La relación entre propiedades mentales y físicas se torna problemática al momento de intentar conciliar dos supuestos: el de la causalidad mental y el materialismo. El primero de ellos consiste en la atribución de eficacia causal a las propiedades mentales. Empleando un ejemplo, esto supone que cuando poseemos un estado mental con la propiedad de desear coger un objeto al frente nuestro, causamos que nuestro cuerpo movilice nuestra extremidad para alcanzar dicho objeto, de acuerdo con nuestro deseo. Esto supone que las propiedades mentales tienen influencia causal sobre las propiedades físicas de nuestro cuerpo.

El materialismo, por su parte, sostiene la tesis ontológica de que todo lo que existe en la realidad está compuesto por materia física. De esto se desprende que todo los entes y estados del mundo tienen propiedades físicas, y que si existen propiedades mentales éstas deben ser reductibles a, o nada más que, propiedades físicas de algún tipo. Asimismo sostiene que toda instanciación de una propiedad en el mundo, debe tener una causa física

capaz de explicarlo (i.e. señalar la o las propiedades que la causaron y la ley que gobierna dicha interacción). Esto suele denominarse la “clausura causal de la realidad”, ya que señala que las propiedades físicas son las únicas capaces de tener influencia causal sobre otras propiedades físicas.

Distinguiremos al materialismo del fisicalismo, al cual definiremos como la tesis de que toda explicación de los fenómenos del mundo puede ser descrita a partir del vocabulario y las leyes de la física fundamental. El materialismo, en cambio, sostiene que la explicación científica puede incorporar conceptos y leyes distintos de los de la física, como los observados en ciencias que poseen sus propios dominios explicativos, como ocurre, p. ej. con la biología. Ahora bien, esta diversidad explicativa no puede ser ilimitada, ya que de otro modo el materialismo resultaría una tesis trivial. Esto porque podrían incorporarse nuevos dominios explicativos para nuevos (y dudosos) fenómenos, apelando a conceptos y leyes que no tuviesen ninguna relación con lo que actualmente sabemos de los componentes elementales del mundo físico. Por lo tanto, el materialismo debe aceptar que la física fundamental debe tener cierta autoridad al momento de validar un dominio explicativo u otro. ¿Significa esto que los conceptos y leyes de los dominios explicativos de las distintas ciencias deben, en último término, ser reductibles a la física, y por lo tanto el materialismo no es más que una forma de fisicalismo?

La respuesta es controversial. Y lo es en gran medida porque tanto los límites de la física y como las posibles modalidades de reducción no está bien definidos. Pudiera ser, por ejemplo, que propiedades que actualmente no caen dentro del dominio de la física fundamental pudieran llegar a ser reductibles tras progresos futuros en el dominio de la física (p.ej. descubrimiento de nuevas entidades y leyes antes desconocidas). Por otra parte, podrían desarrollarse modelos alternativos de reducción, distintos del modelo estándar de reducción interteórica. Dada su relevancia para esta exposición del tema, explicaré este modelo a continuación. Para llevar a cabo la reducción de una teoría A a una teoría B (i.e. reemplazar la explicación de un mismo fenómeno natural dada por A por la dada por B) la reducción estándar sostiene que es necesario identificar los conceptos de ambas teorías que se refieren a un mismo fenómeno, por medio de “leyes puente” que permitan establecer las correspondientes equivalencias entre sus predicados (Nagel, 1998). Un ejemplo típico corresponde a la reducción del término “agua”, empleado por una teoría macroscópica del agua, por el término “H₂O” propio de una teoría fisicoquímica; no obstante “agua” no figura dentro de los ítems teóricos de la fisicoquímica, es posible identificarlo con el concepto de “H₂O” y así reducir los predicados de la primera a la segunda teoría. Si bien este modelo de reducción resulta coherente a primera vista, plantea una incompleta derivación de las propiedades y relaciones nómicas de la teoría reducida a la teoría base, empresa que, más allá de algunos casos paradigmáticos, ha mostrado ser empíricamente irrealizable y ha sido desplazada por modelos alternativos de reducción (Kim, 1998). Efectuar una reducción interteórica resulta especialmente difícil para el caso de las propiedades mentales, en parte dado que no son directamente accesibles a la observación intersubjetiva, y por otras razones que veremos más adelante. Por otra parte, la suposición de que toda reducción puede llevarse a cabo a partir de leyes que se derivan linealmente de una teoría a otra, supone una visión demasiado simplificada de la realidad. Cuando los fenómenos son estudiados en toda su complejidad, muchas veces siguen conductas no-deterministas y deben seguirse derivaciones que pueden ser no-lineales, como ha sucedido con la “reducción” de propiedades químicas y biológicas a través de la mecánica cuántica (McLaughlin, 2008). Más adelante volveremos sobre este tema, en particular cuando abordemos al emergentismo.

Tanto la causación mental como el materialismo conforman un trasfondo teórico usualmente aceptado para el problema mente cuerpo, así como para la ciencia, la filosofía y el sentido común contemporáneos. Sin embargo, al asumir ambos supuestos se presentan un problema. De acuerdo con el materialismo, todo evento físico está causalmente determinado por otros eventos físicos, mientras que de acuerdo con la causación mental, un evento físico puede ser causado por uno mental. Pero, si las propiedades mentales son algo distinto de las propiedades físicas, ¿cómo es posible que puedan tener influencia causal sobre éstas sin romper la clausura causal de la realidad propia del materialismo?. La filosofía de la mente ha intentado resolver este problema a través de distintas formas de plantear la relación entre propiedades mentales y propiedades físicas, las cuales serán el tema de la siguientes secciones.

TEORÍA DE LA IDENTIDAD Y FUNCIONALISMO

Una posible solución al problema de la causalidad mental en un plano materialista está en identificar las propiedades mentales con propiedades físicas, generalmente instanciadas en regiones del cerebro. Esta teoría fue propuesta como un modo de validar el acceso introspectivo a los estados mentales, sin que por ello deba aceptarse que éstos señalen algo distinto del mundo físico (Feigl, 1984). Un estado mental y un estado cerebral pueden describirse como la instanciación de propiedades provenientes de distintos sistemas conceptuales, pero que sin embargo tienen un mismo referente, un estado físico del cerebro. Esta relación de identidad puede también describirse como una de reducción, en tanto una descripción mental de un evento es reducible a una descripción cerebral del mismo. Ahora bien, si un estado mental se caracteriza de acuerdo con su rol causal (Armstrong, 1990), la teoría de la identidad postula que cuando un estado mental causa un estado físico, debe existir un estado físico idéntico al primero y que es responsable del efecto causal. De esta manera, la teoría de la identidad permite mantener la causalidad de lo mental sin transgredir al materialismo.

Esta teoría se vincula con el modo de reducción estándar entre teorías. Como ya señalamos en la sección anterior, conectar las propiedades mentales de modo lineal con las propiedades de otras ciencias como la neurofisiología parece ser una pretensión tanto ambiciosa como irrealizable, en parte por la poca claridad conceptual que existe respecto de una ciencia de lo mental así como de cómo funciona el cerebro. Empero, las objeciones tradicionales a la teoría de la identidad son de dos tipos, que revisaremos a continuación.

La primera y tal vez más decisiva de estas objeciones es de tipo empírico, y se basa en el argumento de la realizabilidad múltiple propuesto por Putnam (1991). Este argumento sostiene que resulta implausible que una propiedad mental se corresponda invariablemente con una propiedad física que sea la misma a lo largo de todas las especies y todos los seres humanos capaces de instanciar aquella propiedad mental, y por tanto no es posible identificar propiedades o tipos mentales con propiedades físicas del cerebro. Las propiedades mentales no corresponderían propiamente a propiedades físicas, sino a propiedades funcionalmente caracterizadas de acuerdo a su rol causal, insertas dentro de una serie de otras propiedades funcionalmente descritas dentro de la mente del organismo. Esta tesis corresponde al funcionalismo, que además constituye el enfoque predominante para abordar el problema mente-cuerpo dentro de la ciencia cognitiva. Antes de pasar al funcionalismo propiamente tal, revisaremos un segundo grupo de objeciones a la teoría de la identidad.

El teórico de la identidad puede insistir que gran parte de las objeciones a su teoría se basan en la ignorancia, y que debemos volcarnos a la neurofisiología si es que queremos encontrar dónde están y qué son los estados mentales. De hecho, la teoría de la identidad desde sus inicios fue considerada una hipótesis científica, a ser completada con el progreso de la ciencia, tal como alguna vez se llegó a identificar el agua con moléculas de H₂O o el rayo con una descarga eléctrica (Smart, 1984; Place, 1984)). Sin embargo, existen argumentos no empíricos en contra de que esta empresa pueda llevarse a cabo para el caso

del problema mente-cuerpo. El más tradicional de ellos apela a la llamada Ley de Leibniz, que sostiene que cuando dos entidades son idénticas, deben entonces ser idénticas en todas sus propiedades. De esto se puede seguir que conforme con su identidad con un estado cerebral, un estado mental puede tener la propiedad de pesar 4 gramos, o que un estado cerebral puede tener la propiedad de ser alegre, todas ellas consecuencias a primera vista absurdas. Una salida que fue propuesta por los teóricos de la identidad fue la de reemplazar la riqueza descriptiva de nuestros informes experienciales por una descripción de tipo “tópico neutral”, que sólo procura informar un estado perceptivo estándar frente a un estímulo determinado, abstrayendo de nuestro informe las propiedades fenomenales que solíamos atribuir al estímulo (Smart, op.cit.). Una variedad de este argumento está en que debemos corregir y perfeccionar nuestros informes introspectivos de acuerdo a lo que realmente son: procesos cerebrales. Esto llevaría a reformular nuestra concepción de mente de modo que se ajuste a lo que la neurofisiología nos enseña, es decir, aceptar que por ejemplo mi sensación de rojo está ubicada en mi corteza parietal, o que hay estados cerebrales con propiedades semánticas (Churchland, 1992).

Sin embargo, esto parece excesivo, en particular para quienes reclaman un trato especial para las propiedades mentales. Siguiendo un famoso argumento de Jackson (1986), podemos imaginar a una persona que ha conocido todos los hechos científicos relacionados con la neurofisiología y la física de los colores. Sin embargo, si siempre ha vivido dentro de una habitación en blanco y negro y no ha tenido la ocasión de presenciar objetos coloreados, debemos aceptar que no posee un completo conocimiento del color. Dicho de otro modo, al salir de la habitación en blanco y negro y observar una manzana roja, habrá adquirido nuevo conocimiento. Un conocimiento de naturaleza diferente del conocimiento científico, uno que sólo puede adquirirse en primera persona, y que no podrá ser reducido a enunciados científicos sin dejar de lado aspectos esenciales de la experiencia. Dicho en palabras de Searle (1992), mientras que al reducir agua a H_2O estamos llevando a cabo una redefinición de un concepto por otro que nos aporta una mayor comprensión del mismo fenómeno, al reducir nuestra experiencia consciente de dolor a procesos neurofisiológicos estamos dejando fuera de la explicación a hechos que no son expresables por medio de la teoría neurofisiológica. El componente subjetivo propio de experimentar un dolor, o percibir un objeto de color rojo, corresponde a una clase especial de hecho que ocurre en la conciencia, distinto de un hecho objetivable a través de la física empírica, y por lo tanto irreducible a ésta. Puede apelarse, por tanto, a la inconmensurabilidad de una teoría psicológica basada en la experiencia en primera persona con respecto a teorías científicas basadas en hechos evidenciables intersubjetivamente. Esto nos llevaría a la aceptación de propiedades mentales que no son reductibles ni pueden ser, ni siquiera en principio, explicables a partir de propiedades físicas, lo que resulta contradictorio con el materialismo. Conciliar la peculiar naturaleza de las propiedades mentales con una explicación neurocientífica es un requisito para cualquier alternativa al problema mente-cuerpo. En este trabajo intentaremos seguir este camino, indagando en modelos alternativos de identidad y reducción en el contexto del emergentismo.

Poco atrás dijimos que la teoría de la identidad podía ser refutada por el argumento de la realizabilidad múltiple, y que la tesis que se desprende de este argumento corresponde al funcionalismo, según el cual el estudio de las propiedades mentales puede llevarse a cabo de manera independiente de la base física donde éstas se instancian. Para el funcionalismo estas propiedades son caracterizadas como estados funcionales (habitualmente estados transicionales de una máquina de Turing, que es un dispositivo computacional abstracto), siendo sólo relevante para explicar la actividad mental el conocer cómo cada estado se relaciona con los inputs, outputs, y otros estados funcionales

internos del organismo. El funcionalismo se asocia con una visión la mente como un conjunto de propiedades intencionales, o más bien, como una serie causal de actitudes proposicionales. Como ya dijimos, esta serie puede imaginarse situada dentro de una máquina de estados, que manipula computacionalmente estos estados de acuerdo con la historia causal de inputs, estados internos y outputs conductuales que tiene el sistema. Esto suele denominarse funcionalismo computacional, o teoría computacional de la mente, y supone que la cognición humana puede ser suficientemente explicada a partir de esta caracterización funcional de las propiedades mentales. Dado que para el funcionalismo estas propiedades se individualizan respecto a su rol causal, y de manera independiente de su dimensión material, la relación de estos estados con la dimensión física del organismo resulta irrelevante para los propósitos de caracterizar una teoría de lo mental. De esto se sigue que una misma propiedad mental puede ser instanciada en múltiples sustratos físicos, conforme al argumento de la realizabilidad múltiple. Dicho de otro modo, una propiedad mental se relaciona con una disyunción (potencialmente infinita) de propiedades de una teoría física, estableciéndose así un dominio de lo mental explicativamente independiente de lo que ocurre en el cerebro (Fodor, 1999).

En este punto es importante señalar algunos aspectos lógicos de la identidad y reducción que deben tenerse en cuenta para no malentender estos argumentos. De acuerdo con la teoría de la identidad que hemos expuesto en esta sección, cuando postulamos una relación de identidad entre una propiedad mental y una propiedad cerebral estamos afirmando que existe una identidad necesaria entre ambas partes. Esto porque, siguiendo a Kripke (1980), las expresiones que empleamos para designar a cada una de estas propiedades (p.ej. percibir rojo y la activación del circuito neuronal R) están invariablemente referidas a ellas. Esto porque, al igual que con la identidad entre agua y H₂O, ambas expresiones son “designadores rígidos”, ya que mantienen esta relación de identidad de manera necesaria, en todos los mundos posibles. En este contexto, el argumento de la realizabilidad múltiple sostiene que la identidad entre propiedades mentales y propiedades cerebrales no puede ser necesaria porque es lógicamente concebible que p.ej. cuando percibamos rojo no se active el circuito neuronal R, sino otro S, o T, etc. Como contraparte, el argumento de la realizabilidad múltiple postula una identidad entre una propiedad mental y una disyunción de propiedades físicas que no puede, por tanto, ser necesaria, sino que es contingente, ya que debe ajustarse a las relaciones lógico/causales que corren para el mundo donde se instancia la propiedad mental. Creo que los argumentos en contra de una identidad necesaria para las propiedades mentales físicas son definitivos, pero sin embargo creo que la alternativa funcionalista no resulta satisfactoria, por las razones que veremos a continuación. La solución que presentaremos en este trabajo se orienta hacia una relación de identidad contingente, pero inmersa en el contexto de las propiedades neuro-corporales de la especie humana y propiedades mentales emergentes.

Otro problema del funcionalismo está en que basa su concepción de mente exclusivamente en las propiedades mentales intencionales, sin lograr dar cuenta del carácter fenomenal que éstas tienen (ver sección 1.2). Dos sujetos con una propiedad mental funcionalmente idéntica (con la misma relación de inputs y outputs con sus estados internos), pueden diferir en el modo por el que experimentan esta propiedad, y por lo tanto el funcionalismo es incapaz de distinguir este aspecto cualitativo de los estados mentales. Nuevamente, creo que esto se produce justamente por la característica funcionalista de individualizar las propiedades mentales de un modo totalmente abstraído de su dimensión física. Como se argumentaremos en varias partes de este trabajo, el aspecto fenomenal de las propiedades mentales es tan constitutivo de las mismas como lo es la base material de la cual dependen.

La principal razón por la que creo que la realizabilidad múltiple resulta insatisfactoria como alternativa para el problema mente-cuerpo radica en que no logra establecer una relación entre mente y cuerpo consistente con la naturaleza causal de las propiedades mentales. Si, por ejemplo, un sujeto retira su mano de una olla caliente, suponemos que sintió un dolor que fue la causa de que realizara dicho movimiento. Ahora bien, la propiedad mental de sentir dolor tuvo que consistir en algún tipo de proceso en el cuerpo del sujeto que derivase en el movimiento de su extremidad, un proceso capaz de movilizar estructuras físicas, de transferir energía y modificar el estado de las cosas en el mundo. Todo esto no puede ser llevado a cabo por una propiedad individuada sólo por su rol funcional, en tanto ella no es más que una formalización o descripción dentro de una transición de estados abstractos que no tienen repercusión sobre el mundo real (Cleland, 1992). Por demás, el argumento de la realizabilidad múltiple no permite vincular una propiedad funcional con sus múltiples realizadores de un modo consistente con sus poderes causales, en tanto se instancia en una disyunción de propiedades físicas, cada una de las cuales tiene regularidades nomológico-causales que son diferentes. Una disyunción de propiedades no basta para explicar efectos causales. Así, no tendríamos por qué esperar que el dolor propio de un extraterrestre (que suponemos tiene una base física muy diferente al nuestro) posea exactamente las mismas regularidades que el dolor en un ser humano, o que constituya un mismo estado funcional, ya que sus propiedades físicas son distintas (Kim, 1999a).

La teoría de la identidad y el funcionalismo conforman dos aproximaciones clásicas al problema mente cuerpo, y pueden ser visualizadas como dos polos dentro de una aproximación materialista al problema. En un extremo fisicalista, la teoría de la identidad intenta definir las propiedades mentales como idénticas a propiedades físicas, reduciéndolas a un dominio que parece incapaz de explicar aspectos distintivos de las propiedades mentales. En tanto, en el otro extremo el funcionalismo plantea un mentalismo que abstrae a tal punto la dimensión física de las propiedades mentales que no logra explicar su efecto causal sobre el mundo real. Más adelante en este trabajo examinaremos si el emergentismo alcanza a aportar una aproximación al problema mente-cuerpo capaz de situarse en un punto intermedio y aceptable entre estos dos polos. A continuación completaremos nuestra revisión de las principales aproximaciones a este problema, preparando la discusión sobre el emergentismo como una alternativa.

IDENTIDAD DE LOS CASOS Y SUPERVENIENCIA

La tesis de la Identidad que revisamos en la sección anterior, al menos en su formulación inicial por parte de Place, Smart y Feigl, es consistente con una identidad de los tipos. Es decir, una identidad entre tipos o propiedades de estados mentales con tipos o propiedades de estados físicos, como ocurre por ejemplo al identificar la propiedad mental de percibir el color rojo con la propiedad cerebral de activar un circuito neuronal R. Se plantea una identidad necesaria entre dos tipos de estados o propiedades, y así de acuerdo con el ejemplo anterior cada vez que en un sujeto se instancia la propiedad de percibir rojo debe existir la instanciación de la propiedad consistente en la activación del circuito neuronal R en su cerebro. Como ya vimos, esta teoría postula una relación de identidad difícil de sustentar, particularmente al confrontarla con el argumento de la realizabilidad múltiple y a la insuficiencia del lenguaje de las neurociencias para abarcar ciertas propiedades mentales. Sin embargo, puede formularse una versión más débil de identidad entre estados mentales y cerebrales, esta vez no apelando a propiedades o tipos, sino a casos particulares de estados mentales y físicos. Esta relación se denomina Identidad de los Casos (o Tokens), y postula simplemente que toda instanciación de una propiedad mental se acompaña de la instanciación de una propiedad física, sin postular que exista una relación sistemática entre propiedades. Dicho de otro modo, se trata de una identidad de estados o eventos, no de propiedades. De esta manera, y volviendo al ejemplo anterior, basta con que en un caso donde se instancia la propiedad de percibir el color rojo pueda también instanciarse una propiedad física, sin importar si ésta corresponde a la activación de un circuito neuronal R, M o cualquier otro. Del mismo modo, para la identidad de los casos es posible que una misma propiedad física sea instanciada junto con distintas propiedades mentales, haciendo posible que dos individuos físicamente idénticos tengan distintas propiedades mentales. Dado que esta clase de identidad no aporta ningún tipo de relación definida entre los dominios físico y mental, resulta por tanto una alternativa al problema mente cuerpo demasiado imprecisa para ser consistente con el materialismo. Así como la identidad de los tipos conforma una tesis materialista demasiado fuerte y ambiciosa, la identidad de los casos conforma una demasiado débil (Kim, 2006).

Una manera de conciliar a la identidad de los casos con el materialismo está en la tesis de la superveniencia. Ésta establece la restricción de que dos entidades que son idénticas en sus propiedades físicas no pueden diferir en sus propiedades mentales. Dicho de otro modo, los eventos mentales covarían con los eventos físicos de modo tal que no puede haber un cambio en lo mental sin un cambio correlativo en el ámbito físico. Si bien la superveniencia aporta los requerimientos mínimos para que una identidad de casos no se aparte del materialismo, sigue siendo, no obstante, una tesis débil. Al igual que el funcionalismo, permite que una propiedad mental pueda identificarse con la instanciación de propiedades físicas diversas, quitando relevancia al ámbito físico para explicar una teoría de lo mental. Volviendo a nuestro ejemplo, de acuerdo con la superveniencia si un individuo tiene la propiedad de percibir el color rojo, entonces tiene que poseer una propiedad física tal que cada vez que esta propiedad física está presente, tiene (necesariamente) que estar presente la propiedad de percibir rojo. Sin embargo, resulta compatible con la

superveniencia que dicha propiedad física consista en una disyunción de propiedades tales como la activación del circuito neuronal R, S, T o cualquier clase de propiedad física capaz de instanciar la propiedad mental en cuestión. Por lo tanto, la superveniencia establece sólo una relación de dependencia ontológica de lo mental hacia lo físico, sin aportarnos una explicación de la propiedad mental en términos de las propiedades físicas con las cuales se identifica. Este vacío explicativo tiene además el inconveniente de aceptar que puede existir causalidad desde un dominio de propiedades (mentales) distinto del físico (causalidad descendente) sin clarificar cómo esto puede ser posible sin contradecir al materialismo (Kim, 1998).

Recapitulando, a lo largo de esta sección y la anterior, hemos revisado algunas de las principales aproximaciones al problema mente cuerpo que deben tenerse en cuenta para una discusión sobre el tema. La teoría de la identidad postula una reducción de propiedades (tipos de estados) mentales a propiedades físicas, que al menos de acuerdo al modelo tradicional de reducción (que emplea leyes puente que conectan ambos ámbitos conceptuales) parece imposible de realizar, no obstante más adelante volveremos a tratar sobre la identidad de la mano con otros modelos de reducción y explicación. El funcionalismo, por su parte, postula una relación de identidad entre tipos mentales con casos físicos, que adscribe a la tesis de la superveniencia como requisito mínimo para no apartarse de materialismo. El funcionalismo así entendido constituye la aproximación actualmente más aceptada para el problema mente cuerpo, no obstante presenta problemas, varios de los cuales hemos apuntado en estas secciones. Estos se refieren principalmente a lo insuficiente de caracterizar funcionalmente las propiedades mentales sin apelar a su dimensión material y causalmente relevante, así como la imposibilidad de abarcar a las propiedades mentales fenomenales. Si bien al considerarlo junto con la tesis de la superveniencia se establece una relación de dependencia ontológica de las propiedades mentales con las físicas, esta tesis no alcanza a explicar qué aspectos del ámbito físico son relevantes para explicar el carácter peculiar que tienen las propiedades mentales (especialmente su carácter fenomenal), así como su eficacia causal.

A modo de marco general para la discusión posterior, a continuación enumeraremos los requisitos que una aproximación al mente cuerpo debe cumplir para sobreponerse a los problemas que presentan la teoría de la identidad, el funcionalismo y la superveniencia.

1. *Materialismo*: Debe ser compatible con el materialismo. Con esto nos referimos a la tesis ontológica de que en el mundo sólo existen entidades físicas, tesis que debe distinguirse del fisicalismo en cuanto a que el materialismo no sostiene que toda explicación científica deba ser reductible al vocabulario de la física fundamental.

2. *Superveniencia*: Debe ser compatible con la tesis de la superveniencia. Esta tesis está de todos modos implícita en el materialismo, en cuanto a que la totalidad de las propiedades mentales de una entidad dependen completamente de, o están determinadas por, sus propiedades físicas.

3. *Propiedades mentales*: Debe contemplar una ontología para propiedades (o tipos, o clases de estados) mentales. Las propiedades mentales constituyen un requisito para conformar una teoría explicativa de lo mental, como explicaremos al final de esta sección.

4. *Aspecto fenomenal*: Las propiedades mentales no pueden ser desprovistas de su dimensión experiencial, tal como vimos en la sección 1.2.

5. *Propiedades físicas*: Debe contemplar una ontología para propiedades (o tipos, o clases de estados) físicos. Este requisito intenta evitar los problemas del funcionalismo al identificar las propiedades mentales con múltiples realizadores físicos sin definir qué

propiedades de éstos son explicativamente relevantes para lo mental. Dicho de otro modo, para constituir una explicación de las propiedades mentales en un contexto materialista es necesario contar con alguna sistematización de las propiedades físicas que participan y cómo se relacionan con lo mental.

6. *Causalidad descendente*: Dar cuenta de propiedades mentales causalmente eficaces sobre otras propiedades mentales o físicas. Esto debe lograrse sin contradecir al materialismo rompiendo la clausura causal de la realidad, es decir, sin sostener que puede haber causalidad sobre propiedades físicas desde un dominio no-físico. El único modo de hacer esto es preservando la dependencia ontológica y causal de las propiedades mentales con las propiedades físicas, ya que de otra manera se cae en un dualismo de propiedades.

7. *Explicación de la interfaz*: Debe aportar una solución al problema de la interfaz. Esto se refiere a conectar o describir cómo se articulan los distintos niveles explicativos de la cognición humana, en este caso el referente a propiedades mentales con el que trata con propiedades físicas del cerebro y el cuerpo (Bermúdez, 2005). De esta manera, no basta con simplemente afirmar que una propiedad mental se identifica con una propiedad física, sino que hay que agregar una explicación acerca de cómo se identifican estas propiedades y cómo es posible que a través de ambas podamos referirnos a un mismo fenómeno.

8. *Psicología*: Proveer a la psicología de un vocabulario para una ciencia de lo mental. Este último requisito se logra a partir de los anteriores, es decir, habiendo fundamentado la existencia de propiedades mentales causalmente eficaces y establecido su nexo con las propiedades físicas a las que supervienen. El proyecto de una psicología científica se vería enriquecido y delimitado por estos avances, p.ej. a través de una adecuación de su vocabulario o el establecimiento de nuevos nexos causales entre propiedades mentales.

Para justificar estos requisitos, es importante enfatizar que la posibilidad de lograr una teoría explicativa de lo mental pasa por contar con un dominio de propiedades mentales que permita explicar causalmente la instanciación de otras propiedades, ya sean éstas mentales o físicas. De esta manera, la propiedad de sentir dolor puede postularse como la causa de que un sujeto sienta temor, o de que retire su extremidad de un objeto caliente, etc. Lo que interesa es que el hecho de estar con dolor (el recaer en una propiedad mental) es lo que explica el que movamos nuestra extremidad. Si sólo podemos relacionar causalmente eventos (casos particulares) y no propiedades, no podemos alcanzar una explicación de mi movimiento en términos de mi dolor, que es lo que nos parece relevante. Un ejemplo de identidad de casos como ésta corresponde a la tesis del Monismo Anómalo de Davidson (1991), la cual considera que cuando hablamos de propiedades mentales sólo estamos haciendo una interpretación de estados o eventos mentales, que ontológicamente sólo considera como particulares. Por consiguiente, cuando realizamos una explicación psicológica tenemos que renunciar a un nivel ontológico y conformarnos con una interpretación (entre otras posibles) de los eventos mentales, apelando a criterios normativos de racionalidad que son externos a la descripción misma de los eventos (Bermúdez, 2005). De acuerdo con el monismo anómalo, el único nivel donde podemos encontrar leyes que conecten propiedades es el físico, y por lo tanto el único nivel donde podemos alcanzar una explicación científica. Pero dado que pretendemos explicar el dominio de lo mental en un contexto materialista tendríamos que intentar identificar los eventos mentales que observamos con propiedades físicas, lo que resulta imposible en tanto no existen relaciones causales entre propiedades mentales y propiedades físicas. De esto se sigue que una teoría psicológica basada en propiedades mentales es causalmente inerte, e inconmensurable con relación al verdadero nivel causal, el físico, y por lo tanto el dominio de lo mental queda como un epifenómeno explicativamente irrelevante (Kim, 2006).

EL EMERGENTISMO BRITÁNICO

Supongamos que un grupo de científicos se reúne para estudiar un curioso fenómeno. Un gran número de bacterias se encuentran aglomeradas dentro de una piscina llena de agua.

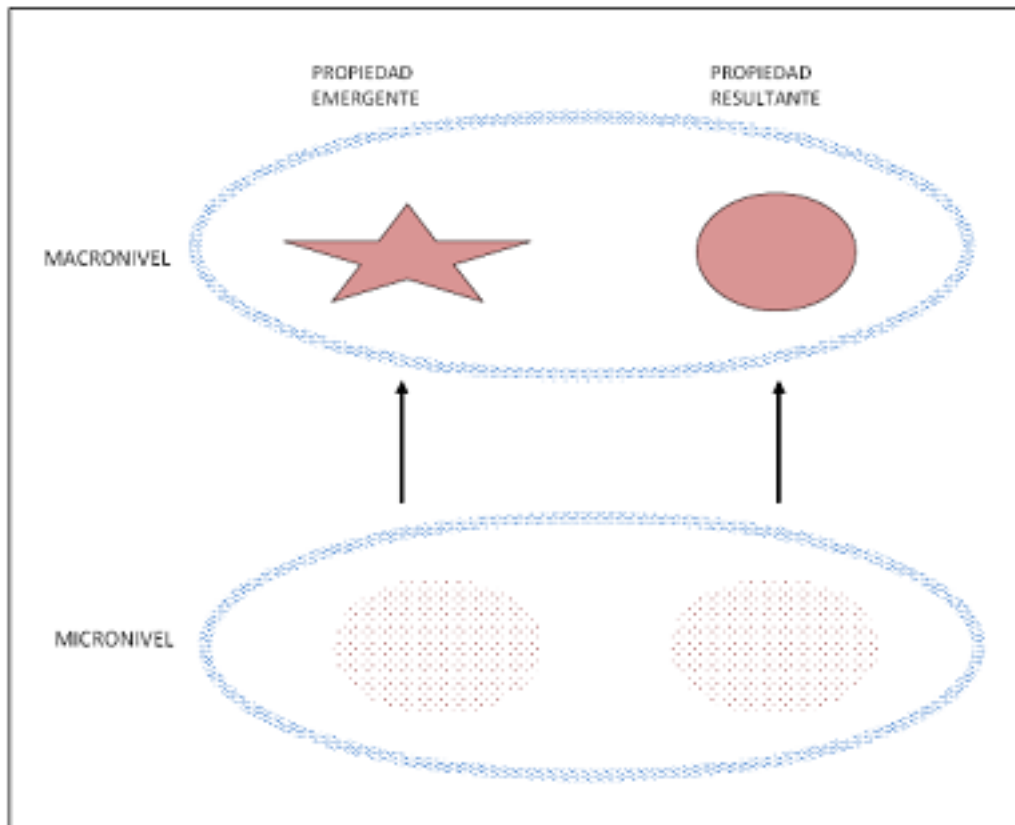


FIGURA 1. La forma oval corresponde a una piscina que contiene agua y bacterias. En parte inferior se ilustra el micro nivel, con las bacterias dispersas de modo aleatorio. En la parte superior, el macro nivel se presenta junto con dos macro propiedades, la resultante y la emergente.

Las bacterias alcanzan a ser tan numerosas que pueden ser visibles claramente como una mancha oscura sobre la superficie, la que presenta una conducta peculiar: en condiciones normales adquiere una forma circular, sin embargo, cuando la temperatura del agua se eleva sobre los treinta grados esta superficie oscura cambia su aspecto circular por uno estrellado (Figura 1).

Tras una intensa investigación de este fenómeno, los científicos concluyen que la forma circular resulta de la acción de las propiedades de las bacterias, y que concuerda con las leyes de la física que gobiernan las relaciones causales entre estas propiedades. Sin embargo, con cierta perplejidad deben reconocer que de acuerdo con las propiedades conocidas de las bacterias y sus relaciones no son capaces de explicar cómo llega a formarse la mancha estrellada. Luego de discutir largamente estos resultados en un intenso debate, los científicos se dividen en dos grupos de acuerdo a la interpretación que

hacen de este curioso fenómeno. El primer grupo sostiene que la forma estrellada resulta impredecible de acuerdo a su conocimiento actual de las propiedades de las bacterias y las leyes de la física que participan de este proceso, pero que esto es producto de la falta de investigación sobre este tipo de fenómenos, y que con toda seguridad en el futuro, con los avances en ciencia e investigación, se conocerán las propiedades y las leyes precisas que dieron lugar a esta mancha estrellada. En tanto, el segundo grupo de científicos no está de acuerdo con la fe de los primeros en que algún día la física logrará explicar y predecir fenómenos como éste. Este grupo llega incluso a afirmar que esto nunca sucederá, porque simplemente esta clase de fenómenos no pueden deducirse a partir del estudio de los constituyentes básicos de donde se originan, ya que consisten en nuevas macro propiedades que emergen del nivel basal cuando éste alcanza cierta configuración y complejidad.

En este supuesto escenario, los científicos del primer grupo pueden considerarse teóricos de la identidad. De acuerdo con el análisis de las propiedades individuales de las bacterias, fueron capaces de deducir y estipular conexiones legaliformes entre este nivel basal y la macropropiedad de conformar una mancha circular. Esta identificación entre ambos niveles de propiedades la extienden a la macro propiedad de conformar una mancha estrellada. El nivel de macro propiedades es, por lo tanto, reductible linealmente al dominio de propiedades y leyes de la física fundamental, y explicable completamente a través de ellas. De esta manera, la posibilidad de reducir en sentido estricto el nivel de macro propiedades hace que la teoría de la identidad sea compatible con el fisicalismo.

En cuanto a los científicos del segundo grupo, consistirían en partidarios del emergentismo. Rechazan que sea posible identificar macro propiedades como la mancha estrellada con propiedades del nivel basal, así como que puedan ser reducidas a éstas. Esto implica que las nuevas propiedades y sus conexiones legaliformes no son identificables con propiedades y leyes de la física fundamental, razón por la cual el emergentismo resulta incompatible con el fisicalismo. Sin embargo, las propiedades emergentes se suponen dependientes del mundo físico, y por lo tanto compatibles tanto con una ontología materialista como con el emergentismo, como veremos más adelante.

El llamado Emergentismo Británico constituye una formulación del emergentismo que podríamos catalogar como clásica. Tiene su origen un conjunto de filósofos que durante los años veinte del siglo pasado elaboraron una teoría emergentista aplicable a las ciencias biológicas, siendo Mill la fuente más citada por ellos como la fuente del concepto de emergencia. De acuerdo con el emergentismo británico, las propiedades emergentes se caracterizan por ser nuevas, impredecibles e irreductibles al nivel base del cual se originan (Morgan, 1923). Dentro de los ejemplos más usados, estaba el de las propiedades emergentes de los compuestos químicos, como por ejemplo las propiedades físico químicas del agua (H_2O), las cuales no resultan de la adición de las propiedades del hidrógeno y el oxígeno tomadas separadas. Estas propiedades emergentes deben distinguirse de las propiedades que son simplemente “resultantes” de las interacciones del hidrógeno y el oxígeno, las cuales corresponden a la simple suma vectorial de las leyes elementales que gobiernan a cada elemento tomado en forma aislada. Como señala Broad (1925), no habría una razón *a priori* para suponer que toda aparición de propiedades deba satisfacer esta condición de ser resultantes de interacciones puramente “mecanicistas”. Más aún, la ciencia contaría con numerosos ejemplos de propiedades que sólo pueden conocerse una vez que emergen de un compuesto cuando este alcanza ciertas proporciones y relaciones, siendo teóricamente imposible deducirlos a partir de sus constituyentes. Sin embargo, el emergentismo británico data del primer tercio del siglo pasado y suponía una concepción

newtoniana de reducción mecanicista, la que en la actualidad ha sido ampliada por la mecánica cuántica, capaz de realizar explicaciones reduccionistas sin apelar a un directo mecanicismo (McLaughlin, 2008). A través de estos desarrollos es posible aplicar leyes no-lineales para “reducir”, por ejemplo, la química a la física. Pese a la no vigencia de algunos de los antiguos ejemplos de propiedades emergentes, la noción de emergentismo sigue teniendo vigencia en el contexto del problema mente-cuerpo, particularmente en relación con la conciencia (o, empleando nuestra terminología, el aspecto fenomenal de las propiedades mentales). Pero antes de evaluar al emergentismo como una alternativa al problema mente-cuerpo, precisaremos los aspectos fundamentales que éste tiene de acuerdo a su definición clásica aportada por Mill, Morgan y Broad, entre otros.

Primero, es necesario dejar claro que los emergentistas británicos no intentaban renegar del materialismo. Para ellos las propiedades emergentes debían ser aceptadas dentro del dominio de las propiedades físicas del mundo, e incorporadas dentro del conocimiento científico. En segundo lugar, el emergentismo sostiene que las propiedades emergentes son nuevas propiedades, que no estaban antes. Notemos que estas dos ideas pueden ser compatibles con una explicación no emergentista, de acuerdo con el ejemplo que planteamos al principio de esta sección: la mancha circular, que era reductible a la microestructura del sistema, puede ser considerada como un fenómeno nuevo, una macropropiedad que no estaba antes y que hace su aparición bajo ciertas condiciones. Esta mancha puede tener otras propiedades reductibles, como poseer una superficie cóncava, las que en ningún caso pueden ser encontradas en las bacterias o en las moléculas de agua tomadas por separado.

Un tercer aspecto del emergentismo británico es el que marca la diferencia en este caso: las propiedades emergentes son además inexplicables, impredecibles, y por tanto irreductibles, a partir de cómo se relacionan los constituyentes del nivel en que se originan. Y esta impredecibilidad es considerada definitiva, como un principio teórico que no depende del grado de avance de la ciencia. Por lo tanto, las propiedades emergentes son sólo congnosibles *a posteriori*, y de este modo incorporables a nuestro conocimiento del mundo. Incluso podrían luego ser relacionadas con determinadas configuraciones de la microestructura de base a través de leyes que traspasan de un nivel emergente a otro, pero siendo estas leyes distintas de las que gobiernan cada nivel por separado (Broad, 1925). Lo principal es que estas nuevas propiedades no pueden ser ni explicadas ni predecidas a partir de sus condiciones basales, hecho que debe ser aceptado sin más, o en palabras de Morgan, con “piedad natural” (Morgan, 1923). Un cuarto aspecto importante del emergentismo británico, y tal vez el más controversial, está en atribuir poderes causales a las propiedades emergentes. Por las dificultades metafísicas que se derivan de este aspecto del emergentismo, lo trataremos con mayor detalle a continuación. Antes señalaremos que un último aspecto típico del emergentismo está en que propone niveles agregativos, jerarquizados según la complejidad que alcanzan, cada uno de ellos fundado en propiedades emergentes originadas en el nivel inferior. Ejemplos típicos de niveles agregativos son la física, la química, la biología, la psicología y la sociología (Emmeche, K ppe, & Stjernfelt, 1997). Cuando las propiedades emergentes causan cambios sobre un nivel inferior (habitualmente llamado “nivel base”), dicho tipo de causalidad suele denominarse “causalidad descendente”. De esta manera, puede decirse que una propiedad emergente puede tener efectos causales tanto sobre otras propiedades emergentes (de su mismo nivel) como sobre el nivel base (causalidad descendente).

Para el tema de la causalidad es pertinente volver a nuestro ejemplo: a partir de la interacción entre las micro propiedades de la piscina pueden surgir dos tipos de manchas

oscuras en la superficie del agua, una circular y otra estrellada. Supongamos que ambas manchas son capaces de causar la migración de algunas bacterias hacia la periferia de la piscina. Entonces, podemos describir las manchas como macro propiedades que son capaces de causar que causar un efecto migratorio sobre parte de la microestructura de la piscina (las bacterias). Asimismo, resulta natural decir que ambas manchas son propiedades del nivel macroscópico que no pueden hallarse en los microcomponentes de la piscina por separado. De esta manera, puede apreciarse que la causalidad descendente no resulta en principio contraintuitiva, ya sea ésta producto de una propiedad resultante o una emergente. Sin embargo, de acuerdo a su formulación clásica, los poderes causales de las propiedades emergentes sobre el nivel base eran un rasgo definitorio del emergentismo. Se suponía que cuando las entidades de un nivel base alcanzan cierto grado de complejidad y organización, emerge de él un nuevo nivel de propiedades, con su propia organización y leyes causales. Estas leyes pueden su vez actuar sobre el nivel base, interaccionando con las leyes ya existentes en él. Ahora bien, volviendo a nuestro ejemplo, notemos que la diferencia entre la mancha circular y la estrellada está en que sólo la segunda, en tanto emergente, podría considerarse una propiedad genuinamente nueva, y por tanto como una nueva ley que se adiciona al sistema, y no el resultado de la suma de las leyes ya presentes en el nivel base.

La noción de causalidad descendente se torna, sin embargo, problemática, a la luz de la superveniencia. Recordemos que la superveniencia constituye una premisa básica del materialismo, que establece la dependencia ontológica del nivel superior con el nivel de base, relación que puede incluirse en el emergentismo británico. Entonces, siguiendo a Kim (1999b) podemos desarrollar la siguiente observación empleando nuestro ejemplo de la piscina. Llamemos a la mancha estrellada “E”, propiedad que emerge del agregado de micro propiedades “M1” del nivel base. Como ya señalamos, esta propiedad emergente E es capaz, por causalidad descendente, de causar la migración de bacterias hacia la periferia, lo que describiremos como la instanciación de la propiedad “M2” en el nivel base. Entonces, puede hacerse el siguiente cuestionamiento: si la propiedad emergente E emerge de la condición basal M1, y causa el efecto M2, ¿por qué entonces no puede M1 desplazar a E como la causa de M2?, y ¿por qué M1 no puede ser suficiente para explicar la instanciación de M2?. Dado que las propiedades emergentes supervienen a su nivel base, sus efectos causales pueden ser atribuidos a este nivel de base, y ser éste nomológicamente suficiente para explicar todo ellos. De este modo, la propiedad emergente parece prescindible para la explicación, en tanto podemos perfectamente explicar M2 en términos de M1, que constituye su nivel fundamental, y por lo consiguiente la propiedad emergente pasa a ser un epifenómeno causalmente inerte.

Si se intenta salvar la causalidad descendente afirmando que las propiedades emergentes son nuevas y autónomas con relación al nivel base, y por tanto causalmente independientes del nivel base, se contradice la tesis de la superveniencia y por lo tanto se aparta del materialismo. No parece extraño entonces que el emergentismo suela despertar sospechas de estar defendiendo formas de dualismo o vitalismo. De hecho, a principios del siglo pasado varios emergentistas postularon que la emergencia originaba sustancias inmateriales, como la “entelequia” de Driesch o el “impulso vital” de Bergson (Sawyer, 2002). El emergentismo de tradición británica, en cambio, adhería al materialismo, así como sostenía que la naturaleza en último término se componía de partículas atómicas materiales y consideraba que las leyes emergentes podían ser incorporadas a la explicación científica de la realidad, con la salvedad de que estas leyes sólo podían ser conocidas a posteriori (McLaughlin, 2008). Esta pretensión materialista, como hemos podido ver, no es sustentable si se quieren defender propiedades emergentes capaces de causalidad

descendente. Y si no tenemos causalidad de estas propiedades, no nos servirán como alternativas para explicar las propiedades mentales. Para dejar esto claro, contrastaremos al emergentismo británico con los requisitos que debe cumplir una alternativa para el problema mente-cuerpo, según vimos en la sección 3.

1. *Materialismo*: está claramente establecido en el emergentismo británico.

2. *Superveniencia*: se da por sentada al suponer la dependencia última de las propiedades emergentes con las propiedades del nivel base.

3. *Propiedades mentales*: son caracterizables como emergentes, y en principio objeto de relaciones nómicas con las propiedades físicas de base.

4. *Aspecto fenomenal*: no es abordado como problema por el emergentismo británico.

5. *Propiedades físicas*: pueden caracterizarse como micro propiedades.

6. *Causalidad descendente*: no se logra justificar la descripción de poderes causales que sean distintos de los que pueden describirse a partir del nivel de base. No se explica de qué manera pueden atribuirse genuinos poderes causales a las propiedades mentales sin contradecir la tesis de la superveniencia.

7. *Explicación de interfaz*: el emergentismo británico no aporta, y ni siquiera considera posible suministrar, una explicación respecto a cómo se conectan las propiedades (emergentes) mentales con las propiedades físicas (de base). Aceptar esto sin más o con “piedad natural” resulta a todas luces insatisfactorio.

8. *Psicología*: no es posible por el incumplimiento de 6 y 7.

Esta formulación del emergentismo, tal como lo hemos presentado aquí, ha sido también denominado “emergentismo fuerte” para distinguirlo de otras formas de emergentismo que no se comprometen del mismo modo con la irreductibilidad e inexplicabilidad de sus propiedades y efectos causales. Una de ellas se denomina “emergentismo débil” (Bedau, 2007), y se asocia con la renovada atención que el emergentismo ha tenido en el último tiempo en relación con llamadas “ciencias de la complejidad”, las cuales abarcan la teoría de sistemas dinámicos, redes neurales artificiales, robótica situada, entre otras aproximaciones a sistemas biológicos y psicológicos complejos. En la siguiente sección intentaré caracterizar esta noción de emergentismo débil como una manera de plantear una alternativa emergentista al problema mente-cuerpo.

EMERGENTISMO DÉBIL Y CIENCIAS DE LA COMPLEJIDAD

Como hemos podido ver, el emergentismo británico intentaba aportar una nueva perspectiva para explicar ciertas macro propiedades que consideraban problemáticas para la ciencia de acuerdo al desarrollo que tenían en ese momento. Esta idea de emergencia se planteaba como un concepto excepcional a ser incluido dentro de una concepción materialista y reduccionista de hacer ciencia. Hemos revisado ya las inconsistencias del emergentismo británico, así como la obsolescencia de sus ejemplos de propiedades emergentes irreductibles de acuerdo con los avances actuales en mecánica cuántica y otras ciencias (McLaughlin, 2008). Sin embargo, el emergentismo ha tenido también otros desarrollos, que en general se caracterizan por tener una visión más bien holista de la realidad, es decir, una donde se rechaza una visión reduccionista y atomista de hacer ciencia, y se reemplaza por otra donde ciertos fenómenos complejos dependen y deben ser estudiados en relación con la totalidad del sistema al cual pertenecen (Sawyer, 2002). Ejemplos de esta tradición emergentista están en la psicología de la Gestalt, el organicismo sociológico alemán y el pragmatismo americano. Todos ellos tenían en común la resistencia a dejar de ver a los organismos como un todo complejo, cuyo estudio debía incorporar la participación de macro propiedades emergentes de la globalidad del sistema y por lo tanto irreductibles a sus micro propiedades. Creo que esta tradición holista encuentra asidero en perspectivas actuales en la ciencia cognitiva, que enfatizan la naturaleza corporizada y situada de la mente, temas sobre los que volveré antes de concluir esta sección. En esta sección continuaré desarrollando una forma de emergentismo que creo permite conciliar esta tradición holista de la emergencia con el emergentismo británico, así como encuentra respaldo empírico en las llamadas ciencias de la complejidad y en las perspectivas corporizada y situada de la cognición.

El emergentismo británico aceptaba la dependencia ontológica y causal de las propiedades emergentes a las propiedades del micro nivel de base. Sin embargo, postulaba un emergentismo “fuerte” donde los emergentes debían tener poderes causales que de algún modo se agregaban a la causalidad del nivel base, los cuales le conferían una controversial novedad ontológica al nivel emergente. Si en cambio se argumentaba que las propiedades emergentes dependen causalmente de las del nivel base, no había argumentos para defender la prevalencia de la explicación emergentista por sobre la basada en el nivel base, salvo apelando a razones epistémicas (Kim, 1999b). El emergentismo “débil” (Bedau, 2007), constituye una versión más elaborada que la británica clásica, capaz de lidiar con éxito estas controversias. Define las propiedades emergentes como ontológica y causalmente reductibles, de manera compatible con el materialismo y la superveniencia. Esto equivale a decir, de acuerdo con la noción elemental de reduccionismo, que no existen entidades ni poderes causales *por sobre ni más allá* el nivel base; el macro nivel es totalmente dependiente del micro nivel. Sin embargo, lo característico de las propiedades (débilmente) emergentes está en que son explicativamente irreductibles, en el sentido de que su explicación no puede ser reemplazada por una basada en el nivel base sin que esto implique trivializar la explicación.

Para comprender esta idea volvamos a nuestro ejemplo de la piscina con bacterias. La macro propiedad estrellada es una propiedad emergente, en tanto que la macro propiedad circular es una resultante. Recordemos también que tras la aparición de la mancha estrellada se agrupaban bacterias en torno a sus vértices. De acuerdo con la noción débil de emergencia, ambas propiedades son ontológicamente reductibles y sus efectos causales son dependientes de la causalidad de las bacterias y moléculas que están en el charco. La diferencia entre ambas está en la posibilidad de explicarlas a partir del micro nivel. La macropropiedad circular se denomina resultante porque su forma y repercusión sobre el resto del charco puede ser completamente deducida a partir de la agregación de los estados y las relaciones espaciales de las bacterias y moléculas del charco. De este modo, la forma redondeada podría ser calculada por la sumatoria de las áreas de las bacterias aglomeradas en torno a un centro común, etc. La macropropiedad estrellada, en cambio, no es derivable desde una extrapolación lineal de los estados y relaciones entre las bacterias y las moléculas aisladas. Esto debido a que la conducta de los componentes del micro nivel alcanza tal grado de complejidad que su evolución en el tiempo se torna impredecible. Esto ocurre en gran medida dado que el charco es un sistema abierto, que permite el flujo de microcomponentes hacia dentro y fuera de éste, así como la posibilidad de eventos accidentales. Si bien el punto de partida para estudiar este fenómeno consiste en microelementos cuyas interacciones son perfectamente conocidas y deterministas, su conducta dejará de ser pronosticable tras cierto lapso de tiempo, aún cuando se cuente con toda la información necesaria. Este tipo de fenómenos, que comprenden procesos de la física así como de varios niveles y subniveles sobre ella, son estudiados dentro de la llamada teoría del caos, siendo una de las principales consecuencias de esta teoría el divorcio del antiguo matrimonio entre determinismo y predictibilidad (Emmeche, Kørpe, & Stjernfelt, 1997).

Ahora bien, pese a que estos fenómenos son en estricto rigor impredecibles, pueden ser objeto de simulaciones computacionales que permitan reproducir la agregación espacial de las causas locales de los micro elementos, y mostrar como esta interacción se propaga en el tiempo dando lugar al macro fenómeno en cuestión. Notemos que la macro propiedad (resultante) circular también puede ser objeto de simulación; el punto está en que ésta mostrará una evolución lineal y determinista, tal como la que podía haber sido pronosticada por derivación simple (p.ej. a través de una fórmula matemática simple). La macro propiedad (emergente) estrellada, en cambio, sólo puede ser derivable a través de simulación. Es importante aclarar que esto no se debe a dificultades epistémicas. Aunque contásemos con un genio laplaciano o una calculadora de altísima capacidad, éstos no serán capaces de pronosticar las propiedades emergentes del macro nivel, ya que la forma lógica de sus cálculos sólo reproduciría la interacción local entre los micro elementos. Sólo se puede conseguir una derivación de propiedades emergentes apelando a simulaciones que reproduzcan cómo evoluciona la complejidad del sistema a través del tiempo. El que la derivación de las macro propiedades sea sólo posible a través de simulaciones computacionales constituye el rasgo más definitorio del emergentismo débil (Bedau, 2007).

Un ejemplo de este tipo de simulaciones son los autómatas celulares, los cuales corresponden a un entramado bidimensional de células que tienen dos estados posibles: estar vivas o muertas (o estado 1 ó 0, etc.). Estas simulaciones evolucionan en el tiempo, aplicando reglas simples y constantes a los microelementos a cada unidad temporal. Un ejemplo de estas reglas son: una célula muerta pasa a estar viva en un tiempo determinado si y sólo si al menos uno de sus vecinos estaban vivos en el momento previo; cuando la célula vive, permanece viva a través del tiempo. Calcular la configuración que tendrá un número dado inicial de células en el entramado a través del tiempo

resulta fácil en un principio, sin embargo, tras cierto tiempo la evolución de las células se torna impredecible. Aparecen además conformaciones de células vivas que evolucionan exhibiendo patrones de conducta uniformes y complejos, que pueden ser interpretados como seres en movimiento que interactúan entre sí. Por ejemplo, grupos de células que se mueven a través del entramado se denominan “planeadores”, otras configuraciones que aniquilan cualquier planeador que se les aproxima se llaman “comedores”, otras que generan periódicamente nuevos planeadores se llaman “productores de planeadores”, etc. Estas agrupaciones de células tiene una evolución que no es trivial, ya que describe ciertos patrones de conducta complejos que pueden ser sólo determinados con precisión a través de simulaciones. Para comprender cómo son estas simulaciones, presentaremos algunos ejemplos de conductas emergentes. En la figura 2 se muestra un ejemplo de simulación donde un comedor “devora” a un planeador:

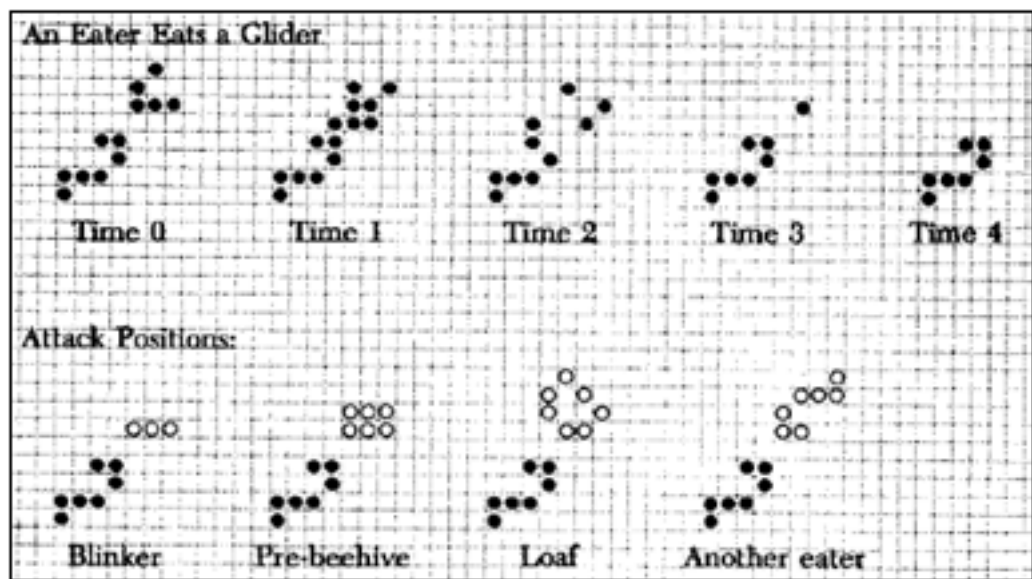


FIGURA 2. (Extraída de Dennett, 2007)

Ahora un ejemplo de un productor de planeadores que emite planeadores:

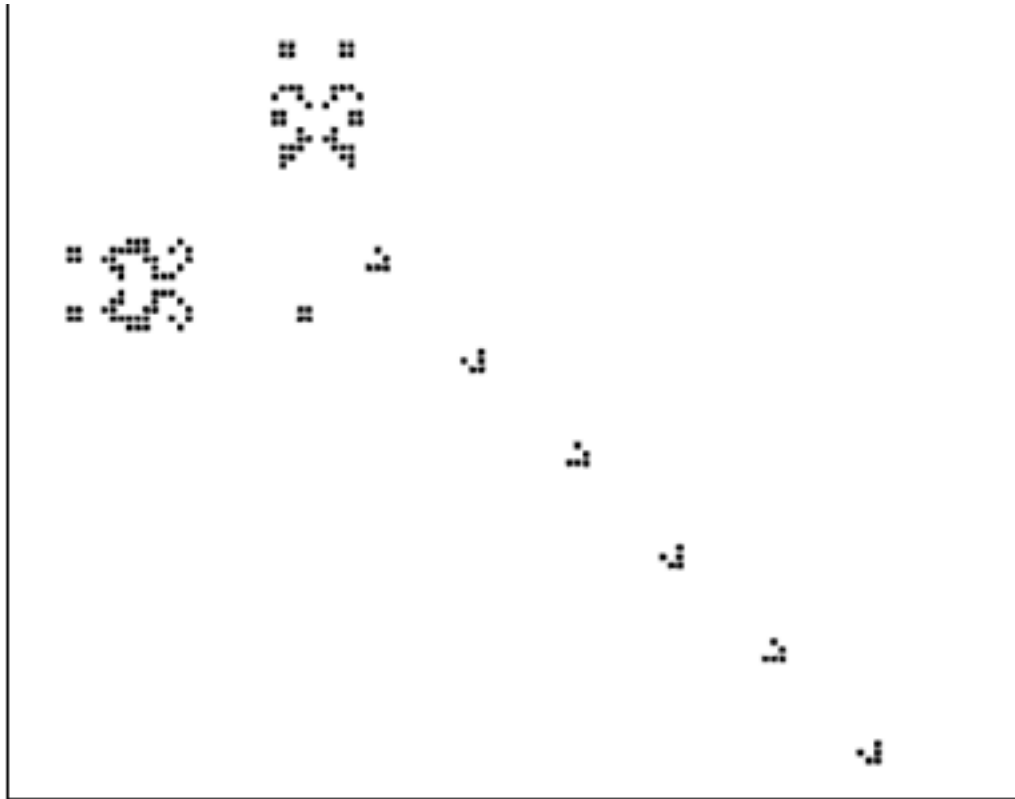


FIGURA 3. (Extraída de Bedau, 2007)

Este productor de planeadores ha emitido ya seis planeadores que avanzan en diagonal hacia abajo. Notemos la célula del extremo izquierdo del planeador más cercano al productor, que dada su ubicación (llamémosla célula 17), pasa periódicamente de estar viva a muerta según avanzan los nuevos planeadores. Este ejemplo es importante ya que muestra una forma de causación descendente. Una macro propiedad emergente, el productor de planeadores, es capaz de causar cambios en una célula individual (del micro nivel), la número 17.

Este ejemplo de causalidad descendente por parte de propiedades débilmente emergentes nos permite evaluar la relevancia que puede tener una explicación causal basada en propiedades emergentes. Como señala Kim (1999b), esta explicación no se justifica si disponemos de una explicación causal del mismo fenómeno basada en los componentes del micro nivel, el cual corresponde al verdaderamente causal y fundamental. De acuerdo con el emergentismo débil, tenemos nuevamente dos posibles explicaciones: una basada en el macro nivel y otra en el micro nivel. ¿Por qué optar por la primera, si ésta es además causal y ontológicamente reducible a la segunda?.

En primer lugar, existe una ventaja epistémica. Si optamos por reemplazar la explicación basada en el macro nivel (el productor de planeadores emite series de planeadores a una cierta velocidad, etc.) por una del micro nivel, nos encontraremos con una compleja trama de datos y relaciones entre los micro componentes que resultarían inconvenientes y menos útiles para una explicación del mismo fenómeno. Si ahora comparamos una explicación de la conducta humana basada en propiedades mentales (macro nivel) con una basada en propiedades físicas (micro nivel), veremos esto con mayor claridad. Supongamos que queremos explicar porqué un individuo salió de su casa a las 8 de la mañana de ayer. Probablemente encontraríamos una explicación bastante simple

y predictivamente útil basada en propiedades mentales: p.ej. que el sujeto tenía el deseo de llegar a su trabajo a las 8:30, y la creencia de que demorará media hora en llegar. Si, en cambio, intentáramos encontrar una explicación neurobiológica basada en neuronas, interacciones fisicoquímicas, neuropéptidos, etc., nos encontraríamos con un conjunto de información irrelevante, compleja y pobremente explicativa respecto de varios aspectos relevantes para la situación, como son el que haya salido temprano, el que quiera llegar puntualmente al trabajo, etc. Una crítica de este tipo ha sido formulada por Putnam (1973), quien sostiene que ciertos sistemas (como el cerebro humano) pueden tener conductas respecto de las cuales una explicación basada en su microestructura es ampliamente irrelevante. Del hecho de que la conducta pueda deducirse de su explicación neurocientífica no se sigue que pueda explicarse desde tal descripción, y esto queda más de manifiesto en la medida que ampliamos nuestro nivel de análisis, desde cerebros a personas, de personas a sociedades, etc.

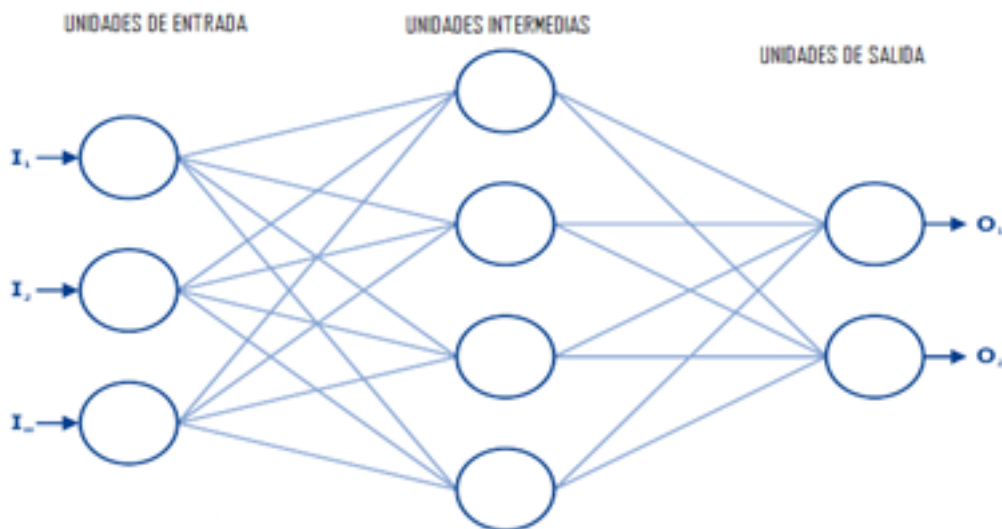
Pero la autonomía explicativa del emergentismo no es sólo epistémica. Las propiedades que son distinguibles en el macro nivel no corresponden a simple azar o “ruido” provenientes del micro nivel, sino que muestran regularidades que son descriptibles de manera objetiva y sistemática, y que además permiten predecir la futura conducta del sistema. El patrón es real, en tanto hay una descripción de la macro fenómeno que es más eficiente que una lectura aleatoria de éste, que sería trivial. Los patrones y regularidades que podemos distinguir en un sistema complejo indudablemente son influenciados por nuestras capacidades senso-perceptivas y por cómo interpretamos lo que vemos, sin embargo esto no quita que esos patrones y regularidades estén ahí, independientes de nosotros (Dennett, 2007). Una aplicación de esta idea aplicada a las propiedades mentales se encuentra en el mismo Dennett (1998), quien sostiene que existen sistemas cuya conducta puede mejor explicada y predecida a partir de una estrategia intencional, basada en atribuirles propiedades mentales intencionales. Esto queda en evidencia al contrastar esta explicación con la que resulta de una estrategia física, llena de detalles explicativamente irrelevantes y predictivamente compleja. Ahora bien, nuevamente el autor es claro en señalar que los patrones de comportamiento intencional están presentes de manera real y objetiva, aunque debemos reconocer que el autor se sitúa en un punto intermedio entre una postura realista e interpretacionista que puede ser criticable.

Estos intentos por escapar del instrumentalismo pueden resultar insatisfactorios. Sin embargo, una visión ontológica del emergentismo débil se ve reforzada al complementarla con el argumento de la realizabilidad múltiple (Bedau, 2007): un mismo patrón emergente en el macro nivel puede instanciarse en distintos micro niveles. Esto puede observarse en simulaciones con autómatas celulares como las que vimos antes en esta sección, donde planeadores y productores de planeadores aparecen a partir de configuraciones heterogéneas de los componentes del nivel base. En palabras del autor (pág. 182):

La autonomía explicativa no indica ninguna macroestructura distintiva en la realidad. Pero los fenómenos de la emergencia débil que pueden ser realizados en una indefinida variedad de diferentes micro contingencias pueden instanciar robustas macro regularidades que pueden ser descritas y explicadas sólo en el macro nivel. El punto no es sólo que la macro explicación y descripción es irreductible, sino que esta irreductibilidad indica la existencia de una macro estructura objetiva. Este tipo robusto de emergencia débil revela algo sobre la realidad, y no sólo respecto a cómo la describimos o explicamos. Por tanto la autonomía de esta robusta emergencia débil es ontológica, no meramente epistémica.

Como ya hemos anticipado, las propiedades emergentes no se restringen sólo a modelos teóricos o simulaciones simples como los autómatas celulares. Durante las últimas décadas se ha venido desarrollando una amplia gama de aproximaciones a sistemas complejos, particularmente dentro del ámbito de las ciencias cognitivas, donde modelos basados en agentes, redes neurales artificiales, robótica situada, etc. han intentado con éxito simular conducta inteligente. A continuación nos referiremos brevemente a las redes neurales artificiales, ya que ellas constituyen un buen ejemplo de propiedades emergentes que podemos analogar con las que atribuimos a la mente humana.

Una red neural artificial tienen una arquitectura simple, compuesta por múltiples nodos conectados entre sí. Estos nodos pueden estar en dos estados posibles: encendidos o apagados (o estado 1 ó 0, etc.). El que estén en un estado u otro depende de las señales que reciben a través de las conexiones con los demás nodos, siguiendo una regla simple de activación. Estos nodos se distribuyen además básicamente en tres capas interconectadas, de manera que la primera de ellas recibe estímulos del exterior (inputs), la segunda sólo consta de conexiones entre nodos, y la tercera capa que representa la conducta del sistema de acuerdo con el patrón de activación que tengan sus nodos (outputs). Estas redes pueden ser “entrenadas” mediante la estimulación sistemática de sus inputs con datos externos que se desea que la red “aprenda”. Para mayores detalles de cómo se llevan a cabo estos experimentos remito a Elman (1998). Para efectos de este trabajo, quisiera señalar que tras adecuadas fases de entrenamiento, una red de este tipo puede mostrar propiedades emergentes que pueden ser detectadas a partir de sus outputs. Un ejemplo de red neural artificial aparece la figura que se presenta a continuación.



*FIGURA 4: Ejemplo de red neural artificial.
Se ilustran los tres niveles que la caracterizan.*

Si esta red es expuesta a inputs correspondientes a patrones de algún tipo (p.ej. imágenes de rostros, letras, palabras, etc.) es capaz de modificar el peso de sus conexiones dentro del sistema, incrementando o disminuyendo el potencial activador de estas. Estos cambios internos consisten en la “memoria” del sistema, ya que logra capturar características de los patrones de modo que, tras sucesivas fases de entrenamiento, la red es capaz de discernir entre un patrón y otro cuando es expuesta a inputs consistentes por ellos. Cuando la red identifica un patrón que ya tiene almacenado en su memoria, emite outputs que dan cuenta de este reconocimiento. Incluso redes de este

tipo pueden subsecuentemente generalizar entre los estímulos aprendidos e identificar patrones presentes en inputs a los que nunca antes había sido expuesta (Rumelhart, 1989). De esta manera, puede decirse que la red puede ejecutar labores de tomar una decisión entre varias alternativas, formular hipótesis inductivas, etc., todas ellas tareas para las cuales asumimos la presencia de propiedades mentales intencionales (que para este caso equivalen a reglas o algoritmos). El punto está en que estas propiedades nos han sido insertadas en el sistema como tales, ni tampoco pueden inferirse a partir de las conexiones que conectan localmente a sus nodos. Estas propiedades emergen de la activación global de la red, y el modo como ésta procesa la información sigue una función no-lineal, de manera que el valor numérico de los outputs no es directamente proporcional a los inputs. Estas características son propias de sistemas complejos como los que hemos revisados en esta sección, los cuales son débilmente emergentes en cuanto a que sus macro propiedades sólo pueden ser predecidas mediante simulación. Asimismo, nuevamente nos encontramos con la pregunta ontológica: ¿tiene realmente propiedades mentales o reglas esta red?. Si buscamos estas propiedades dentro de la red sólo encontraremos nodos y conexiones. El único modo de conocer estas propiedades están en observar la conducta del sistema, la cual muestra regularidades que permiten identificarlas. Por lo tanto, podemos nuevamente hablar de macro propiedades reales, que dependen causal y ontológicamente del micro nivel, pero que son explicativamente irreducibles a éste. Y estos son los rasgos definitorios del emergentismo débil.

Las características de las redes neurales artificiales se enmarcan dentro las ciencias de la complejidad, que estudian cómo conductas inteligentes tales como reconocimiento de patrones o toma de decisiones pueden ser llevadas a cabo por diversos sistemas. Esto gracias a sus capacidades de auto organización, procesamiento paralelo y propiedades emergentes, las cuales sólo son susceptibles de ser explicadas a un macro nivel. El micro nivel pasa a ser explicativamente irrelevante, ya que estas propiedades emergentes pueden ser realizadas en distintos sustratos, ya sean computadoras, redes neurales o cerebros (Haken, 2003). Esta perspectiva se asocia por lo tanto a la tesis de la realizabilidad múltiple que vimos en la sección 2 a propósito del funcionalismo, y esto nos obliga a reconsiderar los problemas que vimos ahí. Al igual que el funcionalismo, el emergentismo aquí expuesto nos ha llevado a definir las propiedades mentales de un modo abstraído de su instanciación material, lo que si bien permite, en principio, elaborar generalizaciones nómicas para una teoría de lo mental, no nos aporta una solución al problema mente cuerpo en su conjunto. Como ya hemos señalado, las propiedades mentales tomadas en un sentido abstracto (ya sea una función matemática o una descripción funcional) e independiente de su base física de instanciación, no pueden ser dotadas de poderes causales que tengan un efecto en el mundo real. Distintos realizadores tienen diferentes regularidades nomológico-causales, y por lo tanto no permiten individuar a una propiedad. Para afrontar estos problemas, y proponer al emergentismo como una alternativa para el problema mente-cuerpo, en la siguiente sección vincularemos la tesis del emergentismo débil con la perspectiva de cognición corporizada y situada.

EMERGENTISMO DÉBIL Y COGNICIÓN CORPORIZADA Y SITUADA: UNA ALTERNATIVA PARA EL PROBLEMA MENTE-CUERPO

La perspectiva denominada cognición corporizada y situada ha tenido un desarrollo paralelo junto con las ciencias de la complejidad, en el marco de alternativas para el enfoque computacional tradicional de la ciencia cognitiva. Sus principios básicos son que para comprender la cognición humana en un contexto biológicamente real debemos prestar atención al rol que juegan el cuerpo y el mundo circundante, y que para comprender esta compleja y sincrónica interacción entre cerebro, cuerpo y mundo se requieren nuevos conceptos, herramientas y métodos, aquellos relacionados con el estudio de fenómenos emergentes, descentralizados y auto organizativos (Clark, 1998; Varela, Thompson, & Rosch, 1997). Esto nos permite plantear un nexo entre la mente, concebida de manera abstracta, con las propiedades físicas propias de los seres biológicos. Asimismo, el aspecto situado de la cognición revela que si se desea comprender a la mente humana tal como ella es en la realidad, debemos considerar que ella se instancia en un cerebro, cuerpo y ambiente que están en continuo cambio e influencia recíproca. Esto implica que la mente emerge de un sistema abierto, donde no hay una derivación lineal a partir un conjunto aislado de variables y por lo tanto no hay evolución de tipo determinista. Sin embargo, como ya hemos adelantado, esto no significa que la conducta de estos sistemas siga un curso aleatorio, ya que gracias a procesos auto organizativos de estos sistemas emergen conductas complejas y regulares que están lejos de ser triviales. Desde una perspectiva dinámica y aplicando principios de la termodinámica estas conductas pueden ser entendidas como estados de equilibrio y reducción de gradientes energéticos, que el sistema lleva a cabo para conservar su estabilidad interna (Tschacher, Dauwalder, & Haken, 2003). Las propiedades emergentes que emergen de esta interacción pueden entonces ser visualizadas como funciones que lleva a cabo un organismo para lidiar de manera eficiente con un ambiente variable y así conservar su integridad, y por lo tanto como respuestas dotadas de intencionalidad. De esta manera, las propiedades emergentes pueden visualizarse en un contexto adaptativo propio de los organismos biológicos, como respuestas auto organizativas que no pueden dejar de ser caracterizadas a partir de su dimensión corpórea y ambiental. La caracterización de propiedades mentales de manera absolutamente independiente de su base física de realización pierde valor al confrontarla con la realidad biológica de la cognición.

Continuaré con una breve revisión respecto de las características neuroanatómicas del cerebro humano, que son relevantes para fundamentar que la mentalidad propia de el hombre y algunos animales no puede ser comprendida sin considerar su estrecha vinculación con estados senso-perceptivos relativos a su cuerpo y ambiente. Siguiendo el influyente enfoque neo asociacionista sintetizado por Geschwind (Catani & H. ffytche, 2005), se considera que las funciones cognitivas del cerebro primitivo de los mamíferos

se basa principalmente en conexiones directas entre estructuras límbicas (relacionado con respuestas instintivas y emociones) y la corteza cerebral sensorial y motora primaria. Durante la evolución humana, estas conexiones fueron haciéndose más numerosas hasta dar lugar a amplias regiones de asociación, permitiendo el desarrollo de funciones cognitivas más complejas y menos dependientes del sistema límbico. Uno de los resultados de estas regiones de asociación multimodal sería la capacidad humana de tener lenguaje y pensamientos de orden superior. Esta perspectiva de la filogenia del cerebro humano no dista mucho de lo que podríamos haber supuesto, ya que es concordante con la máxima darwiniana de que la evolución emplea viejas estructuras y mecanismos para desarrollar nuevos mecanismos y funciones adaptativas. Lo que resulta más interesante es notar que estas antiguas estructuras siguen jugando un papel relevante para la cognición humana. La mentalidad de los mamíferos primitivos se basa en representaciones de modalidad senso-perceptiva relativas al mundo circundante y al propio cuerpo, estrechamente ligadas con regiones subcorticales que generan estados emocionales y detonan respuestas instintivas. En el caso humano, estas distintas señales son integradas y procesadas especialmente en la corteza prefrontal, donde además se llevan a cabo procesos de razonamiento y toma de decisiones. Estos datos sugieren que los impulsos biológicos, estados corporales y emociones constituyen el fundamento sobre el cual se llevan a cabo los pensamientos de orden superior, hipótesis que encuentra respaldo empírico en el estudio de pacientes con daño cerebral (Damasio, 1996).

Volvamos entonces a la distinción entre propiedades mentales intencionales y fenomenales que hicimos en la sección 1.2. Ahí señalamos que esta división resulta artificial dado que las propiedades fenoménicas pueden ser caracterizadas como intencionales, así también porque todo estado mental supone una dimensión material que es propia de quien experimenta ese estado, lo que puede llamarse el *cómo se siente estar* en ese estado. Ahora tenemos más elementos para desarrollar esta tesis de que toda propiedad mental tiene un aspecto fenomenal. Los mamíferos no dotados de lenguaje poseen una mentalidad que sin problemas podríamos catalogar de fenoménica, dado que su cerebro no posee regiones corticales que permitan integrar información más allá de sus impulsos biológicos, estados senso-perceptivos y emociones. Esto conformaría la conciencia animal, que acompañaría a todo momento a los animales suficientemente complejos para tener propiedades mentales (Panksepp, 2007). Gracias a una progresiva encefalización, los mamíferos superiores logran un mayor grado de complejidad e independencia de reacciones instintivas, pudiendo integrar información y generar respuestas más complejas. Los seres humanos, por su parte, alcanzan un mayor control sobre sus estados internos y alcanzan pensamiento computacional gracias al aprendizaje de un lenguaje natural (Esta visión de la evolución del lenguaje es controversial, y la he defendido en otra parte (Aguilera, 2007)). El punto es que las propiedades mentales fenomenales (estados senso-perceptivos y emociones) constituyen una plataforma corporizada y situada donde se desenvuelve la cognición humana, y por lo tanto siempre existe un trasfondo experiencial que acompaña las propiedades mentales, llámese esto conciencia o simplemente vida mental.

Pero, ¿qué relación tiene todo esto con el emergentismo?. El propósito de este trabajo es defender una posición emergentista de los estados mentales. Ahora bien, tomando en cuenta nuestro análisis de las propiedades mentales como encastradas en la conciencia fenomenal, podemos postular que las propiedades mentales corresponden a macro fenómenos originados en un cerebro humano corporizado y situado, de acuerdo a la tesis de la emergencia débil. Si consideramos al cerebro y sus estructuras circundantes como un vasto micro nivel que alcanza un alto grado de interacciones complejas, que típicamente evoluciona de modo dinámico y no- determinista, podemos postular que las

propiedades mentales corresponden a fenómenos emergentes que sólo son cognoscibles a un macro nivel. La actividad basal del cerebro probablemente es de tipo holista o análoga (Panksepp, 2007), que emerge como un trasfondo sobre el cual posteriormente se agregan procesos conscientes de tipo digital o computacional. Ahora podemos ver con mayor claridad cómo esta perspectiva puede lidiar con los problemas de la realizabilidad múltiple. Las propiedades mentales sólo pueden ser concebidas en el contexto corporizado y situado propio del organismo que queremos investigar. Esto, por una parte, asegura que las propiedades mentales tienen una dimensión material, evidenciada por su carácter fenomenal, la cual les confiere la eficacia causal que les faltaba si eran concebidas sólo como abstracciones funcionales o matemáticas. Ahora bien, esto nos restringiría a considerar a las propiedades mentales dentro de ámbitos más bien restringidos, de acuerdo principalmente con las diferencias neuroanatómicas entre las especies. Propiedades mentales comunes tienen que tener estructuras neurales comunes, ya que su aspecto fenomenal sería distinto, y por tanto serían propiedades mentales distintas. Si por ejemplo, si se argumenta que un robot logra instanciar la misma propiedad mental funcional que un perro (p.ej. el deseo de ir a buscar un hueso al patio), puede contra argumentarse que no es así; el perro experimenta o *siente* otra propiedad mental, distinta que la que experimenta el robot. De hecho, no parece razonable que dos individuos con una base física tan distinta, y por lo tanto con una mentalidad cimentada en torno a relaciones nomológico-causales completamente diferentes entre sus componentes, puedan compartir los mismos estados emergentes. Por otra parte, la robustez de los fenómenos emergentes frente leves variaciones en cuanto a su micro nivel de implementación, harían posible la equivalencia entre estados mentales de animales de una misma especie, o entre especies filogenéticamente semejantes. Posiblemente haya que construir psicologías para cada especie, o grupos de especies similares, consecuencia que ya había observado Kim (1999a) que se seguía de la tesis de la realizabilidad múltiple.

Las propiedades mentales serían entonces fenómenos emergentes que no pueden ser explicados a partir un conocimiento acabado de la base física que los origina, no obstante son ontológicamente y causalmente reductibles a ésta. Su aparición y evolución no puede por tanto ser determinada a través de ningún cálculo, más aún considerando que la naturaleza situada de la cognición hace de ésta un sistema abierto. Sin embargo, estas propiedades pueden ser explicadas a partir de simulaciones computacionales (redes neurales artificiales, etc.) que reproducen su evolución y permitirían realizar pronósticos probabilísticos respecto de ellas. Ahora bien, como todo fenómeno emergente, las propiedades mentales sólo son cognoscibles a un macro nivel. Pero este macro nivel tiene aspectos epistémicos sin precedentes, y constituyen un caso único de macro nivel: sólo puede ser conocido en primera persona. Dado que en este trabajo hemos argumentado que este macro nivel va más allá de la conveniencia explicativa y puede definirse propiamente como ontológico, tendríamos aquí una manera de argumentar a favor de la autonomía causal y experiencial de la mente humana. Sin embargo, un desarrollo más acabado de lo que implica esta particular forma de emergencia, sólo constatable de manera subjetiva, escapa de las posibilidades de este trabajo.

CONSIDERACIONES FINALES

A modo de conclusión, confrontaremos nuestra propuesta del emergentismo débil junto a cognición corporizada y situada con los requerimientos que señalamos en la sección 3 como fundamentales para considerar aceptable una propuesta al problema mente-cuerpo.

1. *Materialismo*: el emergentismo débil acepta la reducción ontológica de sus macro propiedades, y por tanto es compatible con él. Como vimos en la última sección, el materialismo puede ser reforzado por el realismo biológico de la visión corporizada y situada de la cognición, haciendo más atractiva nuestra propuesta.

2. *Supervenencia*: junto con la reducción ontológica, la dependencia de las macro propiedades emergentes a las del micro nivel está asegurada. Para el emergentismo no existe un macro nivel autónomo.

3. *Propiedades mentales*: Los fenómenos emergentes al macro nivel pueden ser concebidos como propiedades, en tanto se presentan de modo sistemático y no trivial. Asimismo, vimos cómo de acuerdo con la preservación de su integridad frente al medio ambiente las propiedades emergentes de los organismos pueden ser catalogadas como intencionales, así como las propiedades físicas determinan el carácter fenomenal que estas propiedades tienen en el caso de los animales con vida mental (ver siguiente punto). Notemos que, de acuerdo con el emergentismo débil, si quisiésemos reemplazar la explicación mentalista por una basada en propiedades físicas, no podríamos distinguir los fenómenos característicos de la mentalidad y sus regularidades. En este reemplazo la explicación física tendría que apelar a casos (instancias particulares de propiedades) y no a tipos, y en consecuencia no sería satisfactoria para la conducta.

4. *Aspecto fenomenal*: la dimensión experiencial de las propiedades mentales hace que ellas deban distinguirse no sólo por su rol causal, ya que una propiedad mental intencionalmente idéntica es distinta si es llevada a cabo por organismos con diferencias sustantivas en sus propiedades físicas. Como hemos insistido, toda propiedad mental se *siente* de un modo que es característico de acuerdo a donde es instanciada. Así, un ratón al que se le extrae parte del sistema límbico (si es que sobrevive) podrá desarrollar la misma acción que otro con su cerebro indemne (p.ej. mover una palanca), pero la propiedad mental que explica la acción en cada caso será distinta porque tendrá distinto aspecto fenomenal.

5. *Propiedades físicas*: están presentes en el micro nivel, que puede corresponder a grupos de neuronas, circuitos neuronales, regiones del cuerpo, etc. Las propiedades físicas fundamentan las propiedades emergentes desde un punto de vista ontológico y causal, no obstante no es posible identificar unas con otras sino de manera contingente e indirecta. Dado nuestro contexto de cognición situada y corporizada, el nivel físico deja de ser explicativa irrelevante, como lo era para el funcionalismo y, hasta cierto punto, para el emergentismo. Para explicar porqué unos organismos tienen ciertas propiedades mentales y no otras debemos apelar a una explicación emergentista que distingue entre la micro propiedades de sus constituyentes. Si bien algunos de ellos puede resultar irrelevantes (porque pueden cambiar sin alterar la robustez del sistema), otros pueden jugar un rol determinante en la conducta de estas propiedades a través del tiempo (en los sistemas complejos pequeños cambios en las condiciones iniciales pueden –a veces– generar cambios drásticos en la evolución de sistema). Un ejemplo de la relevancia que

pueden tener las propiedades físicas para distinguir entre propiedades emergentes está en el aspecto fenomenal que vimos en el punto anterior.

6. *Causalidad descendente*: gracias a la reducción causal propia de las propiedades débilmente emergentes, queda claro desde un principio el que no se apela a causalidad *por sobre y más allá* del micro nivel, y por lo tanto las leyes causales del macro nivel no violan las del micro nivel. Para explicar la causalidad del nivel emergente hay siempre una explicación alternativa situada en el micro nivel, la cual sin embargo resulta irrelevante porque no logra representar los fenómenos que nos interesa explicar. De todos modos, si describimos una ley causal entre una propiedad emergente que tiene efectos sobre los elementos del nivel de base, esta relación deberá ser de causalidad diacrónica (no simultánea) para no caer en el absurdo de que el efecto causal del macro nivel modifique la propia base donde esta causa se instancia.

7. *Explicación de interfaz*: gracias a las simulaciones computacionales, así como funciones matemáticas no-deterministas, es posible llenar el vacío explicativo que habitualmente queda en las propuestas no reduccionistas al problema mente-cuerpo. Si esta explicación entre niveles resulta satisfactoria o no, es un tema debatible.

8. *Psicología*: De acuerdo a lo señalado en los puntos anteriores, contamos con un vocabulario para una ciencia de lo mental: propiedades mentales, causalidad mental y una plausible explicación materialista para todos aquellos términos. Además, en nuestro caso la ciencia psicológica tendría la ventaja de ser parte de un proyecto de desarrollo conjunto con las neurociencias y las ciencias de la complejidad, las cuales aportarían una explicación y nuevos conceptos para dar cuenta de los fenómenos mentales tal como se postulan para explicar la conducta y se presentan a la introspección. Aprovecho de mencionar que ya existen líneas de investigación que ya desarrollan un enfoque similar, como la neurofenomenología, no obstante referirme a ellos escapa de las posibilidades de este trabajo. Un último aspecto a considerar está en que, dada la evolución no determinista de las propiedades emergentes respecto de su base física de instanciación, podría suponerse de que estamos frente a un anomalismo de lo mental: no puede describirse leyes que gobiernen la evolución de estas propiedades mentales. Sin embargo esto no es así, ya que este indeterminismo radica sólo en la interfaz entre ambos niveles y no en el macro nivel mental. Como hemos explicado ya, las macro propiedades precisamente pueden calificarse como tales gracias a que es posible distinguir su comportamiento de una mera conducta aleatoria. Este comportamiento es susceptible de generalizaciones nomológicas que permitan explicar las interacciones entre estas propiedades de acuerdo con leyes causales. Pero estas conexiones legaliformes no están aseguradas, en tanto sólo podemos describir a posteriori estas macro propiedades, y describir sus regularidades de acuerdo con criterios estadísticos. ¿Supone esto que la psicología será una mala ciencia?. La respuesta es que no será peor que las demás ciencias inductivas, es decir, que todas las ciencias empíricas conocidas hasta ahora.

BIBLIOGRAFÍA

- Aguilera, B. (2007). Origen y evolución de la facultad del lenguaje desde una perspectiva dinámica. (Tesis de Magíster, Universidad de Chile, 2007). En línea: http://www.cybertesis.cl/tesis/uchile/2007/aguilera_b/html/index-frames.html.
- Armstrong, D. M. (1990). The Causal Theory of Mind. En W. Lycan, *Mind and Cognition: a Reader*. Oxford: Blackwell.
- Bedau, M. (2007). Downward Causation and Autonomy in Weak Emergence. En M. Bedau, & P. Humphreys, *Emergence: Contemporary Readings in Philosophy and Science* (págs. 155-188). London : MIT Press.
- Bermúdez, J. L. (2005). *Philosophy of Psychology: A Contemporary Introduction*. New York: Routledge.
- Broad, C. D. (1925). Mechanism and emergentism. En J. K. Sosa, *Metaphysics: An Anthology* (1999) (págs. 487-498). Oxford: Blackwell Publishers.
- Catani, M., & H. ffytche, D. (2005). The rises and falls of disconnection syndromes. *Brain* (128), 2224-2239.
- Chalmers, D. (1996). *The Conscious Mind*. Oxford : Oxford Univ. Press.
- Churchland, P. (1992). *Materia y Conciencia: Introducción contemporánea a la filosofía de la mente*. Barcelona: Editorial Gedisa.
- Clark, A. (1998). Embodied, Situated and Distributed Cognition. En W. Bechtel, & G. Graham, *A companion to cognitive science* (págs. 506-517). Oxford: Blackwell.
- Cleland, C. (1992). Is the Church-Turing Thesis True? *Minds and Machines* (3), 283-312.
- Crane, T. (2001). *Elements of Mind. An Introduction to the Philosophy of Mind*. Oxford: Oxford University Press.
- Damasio, A. (1996). *El error de Descartes*. Barcelona: Crítica.
- Davidson, D. (1991). Mental Events. En D. Rosenthal, *The Nature of Mind*. New York: Oxford University Press.
- Dennett, D. (1998). *La actitud intencional*. Barcelona: Ed. Gedisa.
- Dennett, D. (2007). Real Patterns. En M. Bedau, & P. Humphreys, *Emergence: Contemporary Readings in Philosophy and Science* (págs. 189-206). London : MIT Press.
- Elman, J. (1998). Connectionism, artificial life and dynamical systems. En W. Bechtel, & G. Graham, *A companion to cognitive science* (págs. 488-517). Oxford: Blackwell.
- Emmeche, C., Køppe, S., & Stjernfelt, F. (1997). Emergence: Towards an Ontology of Levels. *Journal for general philosophy of science* , 83-119.
- Feigl. (1984). Mind-body, not a pseudo-problem. En C. Borst, *The Mind/Brain Identity Theory*. New York: St. Martin's Press.

-
- Fodor, J. (1999). Special sciences. En J. Kim, & E. Sosa, *Metaphysics: An Anthology*. (págs. 504-514). Oxford: Blackwell Publishers.
- Gallagher, S. (2007). Phenomenological Approaches to Consciousness. En M. Bedau, & P. Humphreys, *Emergence: Contemporary Readings in Philosophy and Science* (págs. 686-696). Cambridge: MIT Press.
- Haken, H. (2003). Intelligent Behavior: A Synergetic View. En W. Tschacher, & J.-P. Dauwalder, *The Dynamical Systems Approach to Cognition* (págs. 3-16). Singapore: World Scientific Publishing Company Co. Pte. Ltd.
- Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy* , 83 (5), 291-295.
- Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body problem and Mental Causation*. Cambridge: The MIT Press.
- Kim, J. (1999a). Múltiple realization and the metaphysics of reduction. En J. Kim, & S. E., *Metaphysics: An Anthology* (págs. 515-530). Oxford: Blackwell Publishers.
- Kim, J. (1999b). Making Sense of Emergence. *Philosophical Studies* (95), 3-36.
- Kim, J. (2006). *Philosophy of Mind*. Boulder: Westview Press.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge: Harvard University Press.
- McLaughlin, B. P. (2008). The Rise and Fall of British Emergentism. En M. Bedau, & P. Humphreys, *Emergence: Contemporary Readings in Philosophy and Science* (págs. 19-60). Cambridge: MIT Press.
- Morgan, C. L. (1923). *Emergent Evolution*. London: Williams and Norgate.
- Nagel, E. (1998). Issues in the Logic of Reductive Explanations. En M. Curd, & C. J. A., *Philosophy of Science: The Central Issues* (págs. 905-921). New York: W.V. Norton & Company.
- Panksepp, J. (2007). Affective Consciousness. En M. Bedau, & H. Humphreys, *Emergence: Contemporary Readings in Philosophy and Science* (págs. 114-129). London : MIT Press.
- Place, U. T. (1984). Is Consciousness a Brain Process? En C. Borst, *The Mind Brain Identity* (págs. 42-51). New York: St. Martin's Press.
- Putnam, H. (1973). Reductionism and the Nature of Psychology. *Cognition* , 2, 131-146.
- Putnam, H. (1991). The Nature of Mental States. En D. M. Rosenthal, *The Nature of Mind* . Oxford Univ. Press.
- Rumelhart, D. (1989). The architecture of mind: A connectionist approach. En M. Posner, *Foundations of cognitive science* (págs. 133-159). Cambridge: MIT Press.
- Sawyer, R. K. (2002). Emergence in psychology: Lessons from the. history of non-reductionist science. *Human Development* , 2-28.
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Searle, J. (2004). *Mind: A Brief Introduction*. Oxford: Oxford University Press.
- Smart, J. J. (1984). Sensations and brain processes. En B. C. V., *The Mind Brain Identity* (págs. 52-66). New York: St. Martin's Press.

Tschacher, W., Dauwalder, J. -P., & Haken, H. (2003). Self-organizing Systems Show Apparent Intentionality. En W. Tschacher, & J. P. Dauwalder, *The Dynamical Systems Approach to Cognition* (págs. 183-199). Singapore: World Scientific Publishing Co. Pte. Ltd.

Varela, F., Thompson, E., & Rosch, E. (1997). *De cuerpo presente. Las ciencias cognitivas y la experiencia humana*. Barcelona: Gedisa.