

El enactivismo y la naturalización de la mente*

Ezequiel A. Di Paolo

Ikerbasque, Basque Science Foundation

IAS-Research, Centre for Life, Mind and Society, University of the Basque Country

Centre for Computational Neuroscience and Robotics, University of Sussex

Introducción.

La cognición no sólo es lo que sucede cuando nos detenemos a contemplar el mundo y a intentar comprenderlo. El pensar y el entender son sub-géneros del vivir, del actuar y del relacionarse con la realidad que nos rodea y de la que formamos parte. No se piensa al mundo como condición previa para una acción efectiva. No es necesario ni suficiente que la reflexión preceda a la acción para que esta no pierda su sintonía con el entorno. Por el contrario, nos encontramos de entrada y siempre en una acción significativa e involucrados con nuestros objetos de interés y conocimiento. El mundo es lo que nos *importa* y hacia lo cual nuestro cuerpo tiende y que de algún modo también ya entiende. El hecho de que algo nos importe, que no de lo mismo que algo suceda o no suceda, es el dilema fundamental del estudio científico de la cognición y de la filosofía de la mente. Sólo en ocasiones particulares nuestra experiencia toma la forma de un retiro o una reflexión, interiores sólo en apariencia, pero que no son sino una modalidad más de la acción y, en particular, una forma de ser de nuestro cuerpo en relación con el entorno, con otros y con sí mismo. Conocer es participar. Nada más, y nada menos, que esta idea es lo que motiva al enactivismo, esa nueva forma de plantear el estudio riguroso de la vida, la mente y lo social.

El objetivo del presente trabajo es esbozar el núcleo de ideas del enactivismo para demostrar que se trata de una alternativa viable a otros paradigmas en las ciencias cognitivas, en particular aquellos que caen bajo el manto amplio del funcionalismo. La visión computacionalista/representacionalista de la mente (el cognitivismo) según la cual la cognición se reduce en su esencia al

* Versión final de un trabajo a aparecer en la colección *Nueva ciencia cognitiva: Hacia una teoría integral de la mente*, coordinada por D. P. Chico and M. G. Bedia , 2013, Madrid: Plaza y Valdes Editores.

procesamiento de información en el cerebro, ha dominado la disciplina desde sus comienzos a mediados del siglo XX. Sin embargo, esta visión ha sido duramente criticada durante las dos últimas décadas desde diferentes áreas (robótica autónoma, fenomenología, lingüística cognitiva e inteligencia artificial situada, para nombrar sólo algunos ejemplos). El contenido común de estas críticas ha sido que la observación de sistemas cognitivos “en su estado natural” nos provee a menudo una imagen bastante distinta a la que ofrece el cognitivismo, una imagen en la que procesos que involucran múltiples causas y que abarcan al cerebro, al cuerpo y al entorno, se auto-organizan de forma oportunista para producir la performance cognitiva apropiada en condiciones de demandas temporales severas. Estas observaciones no encajan bien con la imagen computacionalista/representacionalista y requieren un entendimiento más profundo del sistema cognitivo en su autonomía y su capacidad corporizada de crear y buscar sentido.

Una perspectiva que reúne estas y otras posiciones críticas es el enactivismo (Varela, Thompson, Rosch, 1991, Di Paolo, Rohde, De Jaegher, 2010, Di Paolo, 2009, Froese y Di Paolo, 2012, Thompson, 2007). El enactivismo considera a la cognición como una actividad continua moldeada por procesos auto-organizados de participación activa en el mundo y por la experiencia y auto-afección del cuerpo animado. El cuerpo vivo crea un mundo de significados en su ser y su accionar (en inglés este es el significado del verbo *to enact*) y no recibe pasivamente información neutra de un entorno a la cual luego tiene que “sumarle” un significado (es interesante notar que las únicas personas que a veces se comportan de este último modo son los que sufren de patologías como la esquizofrenia y el autismo).

El enfoque enactivo se basa en conceptos que se apoyan mutuamente para formar un núcleo teórico: la autonomía, la búsqueda de sentido, la corporización, la emergencia y la experiencia, (Di Paolo, et al. 2010). Según el enactivismo, las propiedades de los sistemas vivientes y los sistemas cognitivos forman parte de un continuo y se influyen mutuamente. Aunque a veces puede entenderse el enactivismo como presentando un orden ontológico provisorio (desde la vida hacia la mente), este orden se supera a sí mismo a través de relaciones complejas que permiten a la mente y a lo social cobrar una vida propia y re-inscribirse en el metabolismo que le da origen (como veremos más adelante).

El espacio de este trabajo no permite brindar una descripción completa de la evolución de estas ideas, y sus relaciones y contrastes con nociones y desarrollos teóricos y empíricos en las ciencias cognitivas corporizadas. Pero sí tenemos la oportunidad de ver en cierto detalle cuál es la relación conceptual que, según el enactivismo, existe entre la vida, la mente y lo social y cómo tal relación se despliega para contestar los interrogantes básicos de las ciencias cognitivas.

El problema que yace implícito en la noción misma de “ciencias de la mente” es el de cómo naturalizar lo mental. Si la mente es un objeto de estudio científico, deberíamos estar en posesión de un marco teórico y un aparato técnico-científico que de manera explícita nos conduzca desde el estudio de procesos naturales a un entendimiento de procesos mentales entendidos en su forma más amplia: desde los específicos como la toma de una decisión o la formulación de un plan hasta los más genéricos como el poder evaluar una situación en términos de las oportunidades que ofrece para nuestra acción. Debemos poder formular lo esencial de la mente, sin entenderla como un fenómeno exclusivamente humano, o de seres capaces de razonar, o de expresarse a través del lenguaje. Por el contrario, todo proyecto de naturalización de la mente debe evitar la trampa de elevar aquello que es específico (y requiere explicarse como tal) al orden de lo genérico y que sirve como base para comprender lo particular. El enactivismo se plantea nada menos que especificar con precisión qué es lo que caracteriza a *todo* lo mental, todo lo que denominamos con nombres como cognición, pensamiento, emoción, intención, acción, percepción, interacción, atención, expresión, etc. La cuestión que nos planteamos es si allí donde el funcionalismo ha fallado (ya sea el representacionalismo computacional clásico, como en sus versiones modernas como la mente extendida) – es decir en la tarea de proveer una noción clara, precisa, operacional y no antropomórfica de la cognición – el enfoque enactivo puede proponer una solución viable (aunque no necesariamente definitiva y final) e inaugurar una nueva teoría científica de la mente.

Para llevar a cabo este propósito, el enfoque enactivo toma como punto de partida la autonomía del sistema cognitivo (Di Paolo, et al., 2010, Di Paolo, 2009, Thompson, 2007). El punto ciego de la ciencia cognitiva tradicional es la desaparición del sistema cognitivo como agente autónomo. No existen definiciones claras, en estas perspectivas, de qué es un agente, ni de en qué consiste su autonomía. Como decimos, este es literalmente un punto ciego

de las teorías cognitivas tradicionales tanto en el sentido de que el problema no aparece como tal desde el punto de vista de estas teorías, como también en el sentido más sutil que indica que es precisamente el ignorar estas cuestiones lo que permite la formulación de nociones clave como el procesamiento de información, las representaciones mentales y las metáforas computacionales. No sólo se pierde en el funcionalismo la posibilidad de preguntarnos sobre el significado de lo mental, sino que es precisamente esa pérdida la que alimenta este enfoque y le permite avanzar sobre aquello que no involucre este cuestionamiento fundamental (lo que explica el éxito del funcionalismo como paradigma científico).

Si bien es cierto que a medida que una ciencia progresa sus términos clave sufren cambios de significado (considérese la diferencia conceptual entre la velocidad aristotélica y la velocidad galileana), es de esperar que al menos exista un acuerdo suficiente en un momento histórico dado sobre la interpretación pragmática de tales términos por un lado y su aplicación dentro de un marco teórico y empírico por el otro. En las ciencias cognitivas tradicionales, esto no ha sido el caso durante más de medio siglo. No existe consenso siquiera práctico para consolidar discursos sobre nociones de agencialidad, acción, normatividad o autonomía. De hecho, el paradigma tradicional que equipara a la cognición con el procesamiento de información nos presenta con un sistema pasivo que responde o bien a las demandas del entorno, o bien a módulos internos y sub-personales que determinan su estado de motivación. El agente autónomo como un todo definido no existe en esta visión de lo “mental”. A modo marginal, y para evaluar lo extendido que está la noción funcionalista en nuestra cultura, es interesante notar que esta misma pérdida de agencialidad es visible también en las concepciones populares sobre el carácter de lo humano que surgen de lecturas demasiado veloces de los avances en biología o en neurociencia – el ser humano es esclavo de sus genes, sus neuronas, sus imperativos evolutivos, pero rara vez surge de estas nociones una concepción de su autonomía; una autonomía seguramente limitada por la historia, la fisiología y la finitud, pero autonomía al fin.

El primer paso que debemos afrontar para desbloquear a las ciencias de la mente es la formulación de una noción científica, práctica e informada a través de nuestra experiencia cotidiana, de la autonomía y sus ideas cercanas como la agencialidad, la identidad, etc. En esta insistencia sobre la autonomía como noción fundadora de la cognición, el enfoque enactivo se

asemeja, de manera poco sorprendente, a la teoría de la autopoiesis sobre la cual basa varios de sus conceptos fundamentales (Maturana y Varela, 1980). Esta teoría tiene como una de sus motivaciones principales el contrarrestar la preocupante evaporación de la noción de organismo en la biología contemporánea. Por debajo del nivel del organismo (en procesos bioquímicos y redes genéticas de regulación) o por encima de él (en la dinámica evolutiva de poblaciones y ecosistemas), la ciencia respetable trabaja sin cuestionamientos. Al nivel propio del organismo el discurso científico se torna a menudo confuso, a veces hasta místico. Por tal motivo el concepto de organismo tiende a ser evitado (aun en disciplinas en la que el organismo como unidad supuestamente constituye el nivel natural de investigación, como es el caso de la embriología o el estudio del comportamiento animal). La psicología y las ciencias cognitivas, salvo excepciones notables, se han visto en una situación idéntica con respecto al concepto del agente cognitivo autónomo.

Lo interesante de este foco sobre el noción de autonomía es que permite al enactivismo a desarrollar una teoría de la búsqueda de sentido fundamentada en la organización de los seres vivos, (Varela 1991, 1997; Weber y Varela 2002; Di Paolo 2005; 2009; Di Paolo et al. 2010, Thompson, 2007). La búsqueda o creación de sentido (“*sense-making*” en inglés) es la interacción significativa y sujeta a normas entre el agente y su mundo. Es la idea que describe por qué las cosas nos importan, porque no nos son indiferentes. Gracias a la noción de autonomía tal idea puede ser naturalizada. Este desarrollo le permite al enactivismo pasar de un punto ciego del cognitivismo funcionalista, el concepto de autonomía, a otro: la definición misma de la cognición. Las siguientes secciones describen este pasaje en detalle — se trata de no menos que de una naturalización de la esencia de *todo* lo mental.

De la autonomía a la búsqueda del sentido: preliminar

Un punto importante que conviene aclarar de manera temprana es que aun basándose en la organización de los seres vivos para elaborar una teoría de la cognición, el enactivismo no implica un enfoque internalista según el cual lo mental ha de encontrarse en los procesos interiores al individuo viviente, por ejemplo en el cerebro (como se ha interpretado, a veces no sin falta de razón, a la teoría de la autopoiesis). Francisco Varela (uno de los fundadores del enactivismo) ha enfatizado, particularmente en sus últimos trabajos, una

concepción de la cognición como un encuentro en el que un mundo de significados emerge a partir de la actividad coherente del ser vivo en relación a su entorno. Es decir una visión lo mental como relacional. Aun en el caso de organismos simples este encuentro involucra la estructuración del entorno inmediato y la consecuente formación de regularidades que retroalimentan la actividad interna del organismo. Es decir que los procesos cognitivos, por necesidad, involucran múltiples bucles que atraviesan la frontera orgánica del ser vivo.

Estos aspectos relacionales no se niegan en la teoría de la autopoiesis y a menudo se los considera en forma implícita en varios de sus puntos más importantes. Sin embargo, se ha sospechado con frecuencia que esta teoría implica una relación de co-extensión entre el sistema viviente y el sistema cognitivo (ver por ejemplo, Wheeler, 2010). De hecho, una de las fórmulas utilizadas con frecuencia en la literatura sobre la autopoiesis ha sido: “vida = cognición” (Stewart 1995; Boden 1999; Bourguine y Stewart 2004). No es exactamente así como el enactivismo se propone entender la relación entre vida y mente.

A esto le podemos agregar varias declaraciones en la literatura primaria que dejan un sesgo internalista, en particular un énfasis sobre la frontera física del organismo (por ejemplo la membrana de una célula) cuando lo que importa conceptualmente es su frontera en términos de organización, una insistencia en la “no intersección” del dominio constructivo de producción molecular y el dominio interactivo del agente como un todo cuando lo que está lógicamente implicada es la “no reducibilidad” entre tales dominios, y finalmente una insistencia en la organización cerrada tanto de la vida como de otras formas de identidad autónoma, por ejemplo la organización del sistema nervioso (implicando de esta forma una auto-determinación que podría confundirse con auto-suficiencia). Estas y otras declaraciones que revisaremos en detalle en este trabajo admiten una interpretación *no* internalista. Sin embargo, se prestan fácilmente a confusión. En lugar de presentar caso por caso un argumento para aclarar el significado de tales propuestas para así comprender mejor la relación entre vida y mente, es preferible pararse de lleno en el campo enactivista y contar la historia de cuál es específicamente el rol de la autonomía y de la vida en la definición de la cognición; una definición que no esta sesgada por el antropomorfismo ni centrada en cerebro como procesador de información.

Vida a secas

La teoría de la autopoiesis se nos presenta como una biología de la cognición. El enfoque enactivo utiliza cierta terminología de esta teoría, en particular conceptos de autonomía y de acople estructural. Pero inmediatamente se advierte que la enacción es parte de una teoría en desarrollo y que la teoría de autopoiesis en sí misma debe evaluarse críticamente y, si fuera necesario, se debe re-interpretar o extender en consecuencia. Es un error considerar a la teoría de autopoiesis, como fue formulada originalmente, como una teoría terminada (una trampa en la que es fácil caer debido a las formulaciones herméticas que abundan en dicha teoría). Varios investigadores han llegado a la conclusión que tanto como teoría de la vida como teoría de la cognición, la autopoiesis deja muchos interrogantes sin resolver (ver Moreno y Ruiz-Mirazo 1999; Di Paolo 2005; Thompson 2007). En particular, nociones esenciales que podrían servir para comprender los vínculos entre la vida y la mente, nociones como la teleología, la normatividad, la agencialidad, etc., reciben un tratamiento pobre o nulo en la literatura primaria. Y cuestiones biológicas de importancia, como la enfermedad y la salud, la fatiga, el esfuerzo, ni siquiera se plantean.

En la formulación original, el término *autopoiesis* se refiere a una propiedad de la organización de los seres vivos: *su auto-producción y auto-distinción como sistemas físicos*. Un sistema autopoietico es una red de procesos de transformación molecular que se auto-regenera debido a que, entre otras consecuencias, la operación de estos procesos resulta en un mantenimiento de las condiciones que le dan origen a los mismos. Esta forma de auto-producirse se propone como la característica definitoria de la vida, en el sentido de que si pudiéramos determinar que un sistema se auto-produce, o sea que es autopoietico, entonces se trata un ser vivo y viceversa. En principio un observador debería ser capaz de determinar si un sistema es o no es autopoietico (para esto se han ofrecido criterios científicos explícitos, por ejemplo por Fleischaker, 1988). Algo que debe notarse es que una organización auto-producida o autopoietica determina una clase de identidad. Un sistema dado o bien pertenece o no pertenece a esta clase, sin términos medios (de modo similar al que una línea en un plano o bien se cruza o no se cruza a sí misma). Esto se relaciona con una característica curiosa: Si bien el concepto de autopoiesis se refiere a procesos de

construcción y transformación entre moléculas físicas, o sea a procesos dinámicos, el concepto en sí no hace explícita ninguna dimensión temporal. La única forma en la que la temporalidad entra en la teoría es a través de la noción de conservación (o sea la noción de algo invariante en el tiempo). Un sistema autopoiético sufre cambios estructurales y permanece como miembro de la clase de seres vivos siempre y cuando conserve su condición de autopoiesis. De no ser así el “sistema se desintegra” (una frase común en la literatura primaria). La razón por la cual esta tautología tiene interés alguno es que la identidad misma del sistema esta dada a través del efecto de auto-distinción que se sigue de la lógica de la autopoiesis. Pero esta relación necesita de una interpretación adecuada (sino corre el riesgo de ser trivial y reducirse a una identidad del tipo $A = A$). Este es el punto en el que la teoría hace un fuerte uso de imágenes intuitivas.

Dado que la definición de autopoiesis no especifica la manera en que se logra la conservación de autopoiesis, la teoría ha recurrido al uso de metáforas para evitar una interpretación trivial ($A = A$). Se dice que un sistema autopoiético es como un sistema homeostático (un sistema que regula y mantiene una variable dentro de límites determinados), donde la variable regulada es la propia organización del sistema (ver por ejemplo, Maturana y Varela, 1980, p. 9). Un examen detallado de la definición de autopoiesis (ibid., pp. 78-79) deja en claro que semejante analogía no está garantizada. Ni siquiera es muy cercana desde un punto de vista intuitivo. Un sistema homeostático implica al menos alguna noción de monitoreo implícito y activo de las variables esenciales y de reacción a las perturbaciones que dichas variables pudieran sufrir (dentro un rango de funcionamiento). Esta estructura adicional no se deduce de la conservación de una organización de auto-producción, que bien podría existir (en un ambiente favorable) sin necesidad de llevar a cabo ninguna reacción homeostática activa. Por otro lado, en sistemas homeostáticos no se prescribe la posibilidad de que las variables controladas efectivamente varíen y hasta crucen fronteras críticas (la homeostasis de hecho sólo se manifiesta como tal cuando una respuesta adecuada devuelve la variable a los valores permitidos, de otro modo no puede distinguirse de la mera estabilidad). Esto bien podría pasar en sistemas autopoiéticos con algunas de sus variables, pero, ¡ciertamente no con su organización! – La noción (aun como metáfora) de la organización del sistema como variable es poco sostenible, amén de confusa en el contexto de un sistema que se caracteriza por mantener esta organización como *invariante*. Si tal “variable”

cruzara una frontera crítica significaría que el sistema simplemente pasaría a estar muerto sin dar tiempo a ninguna compensación adecuada. La definición de autopoiesis, además, no elimina la posibilidad de una conservación de autopoiesis fortuita y no adaptiva, como sería el caso de un sistema que, sin ningún mecanismo compensatorio, simplemente se encontrara en una situación tal que su auto-producción no se viera afectada por perturbaciones. Quizá tal sistema consiga vivir sólo durante un tiempo muy breve, pero no por ello deja de ser autopoietico durante ese tiempo.

Este problema (que he analizado a fondo en otro trabajo, Di Paolo, 2005) es precisamente el problema de la falta de implicación lógica entre la noción de la conservación de autopoiesis que corresponde a la teoría de clases (atemporal) y la intuición dinámica (extendida en el tiempo) de la vida como homeostasis. Esta falta de implicación es la base de muchos inconvenientes que surgen cuando intentamos entender conceptos interactivos y relacionales como la teleología, la temporalidad, la agencialidad, las relaciones sociales y la normatividad (de hecho, los aspectos más importantes de la cognición) en relación con el concepto de vida entendido como autopoiesis. Tales nociones, a la que podemos sumar la corporización y la experiencia, han constituido el interés fundamental del enactivismo desde su concepción en los trabajos últimos de Varela. En parte, estos trabajos pueden leerse como intentos de reconciliar la lógica de teoría de clases la autopoiesis con la gradualidad y temporalidad de conceptos como el significado, las normas y los valores (ver, por ejemplo, Varela, 1997). Una de las últimas propuestas de Varela era que un sistema autopoietico, por el hecho mismo de ser autopoietico, ve al mundo y se relaciona con él a través de una perspectiva de significado, o sea confrontando las cosas que le son relevantes. El sistema vivo reduce de forma masiva la multi-dimensionalidad del acople físico con el entorno a unas pocas dimensiones que le permiten distinguir lo que es bueno de lo que es malo para su autopoiesis (Weber y Varela, 2002) e ignorar aquello que no es relevante. Dicho de otro modo, existe una relación íntima entre el ser viviente y el ser cognitivo. No es esto lo mismo que decir que existe una relación de equivalencia entre la vida y la cognición, sino decir que un sistema que está vivo es un sistema capaz de interactuar de manera cognitiva, de ver al mundo en términos del sentido que éste tiene para la vida.

Creo que este análisis se aplica a los sistemas vivos reales con los que nos encontramos, sólo que no en virtud del mero hecho de que son sistemas

autopoiéticos. El ejemplo favorito de Varela es una bacteria que nada en una solución acuosa trepando un gradiente químico de azúcar. Es la organización misma de la bacteria (y su comportamiento) la que nos señala a nosotros, los observadores, el hecho de que el azúcar tiene un significado especial para este organismo, mientras que otros compuestos químicos son neutrales o nocivos. También nos parece claro que las concentraciones altas de azúcar son preferible a las bajas. Esto no es un hecho neutro como la descripción de una reacción química, sino un hecho que es relativo a una red de intereses. Es un hecho relacional que sería imposible de apreciar si no estuviéramos en presencia de un organismo para el cual las consecuencias del encuentro entre el compuesto químico y los procesos de auto-construcción y auto-distinción tienen sentidos en términos de ser algo bueno o algo malo. Sin la bacteria el gradiente de azúcar es tan neutro como otros cientos de correlaciones físico-químicas que podríamos observar en el mismo espacio y en el mismo tiempo.

Esta idea, esta conexión íntima entre organización de la vida y búsqueda de sentido, significa nada menos que una naturalización de valores que conduce finalmente a una naturalización de la intencionalidad. Se trataría de un logro radical que ha eludido al representacionalismo, al cognitivismo en general e incluso a cualquier enfoque funcionalista (tradicional o extendido). Sin embargo, la propuesta no puede reconciliarse con la literatura primaria sobre autopoiesis tal como pretendía Varela. Para que funcione, se debe recurrir a nociones intuitivas como “quiebres” en autopoiesis, los cuales pueden ser de orden mayor o menor (Varela, 1991), o sea situaciones en las que el organismo está en mayor o menor riesgo, o funciona mejor o peor. Esto va en contra de la doctrina de conservación de la autopoiesis. Aunque suenen intuitivas, tales nociones no tienen sentido dado que la autopoiesis es un todo o nada. ¿Qué significa estar un 80% vivo? La autopoiesis no se presenta en grados (Maturana y Varela, 1980, p. 94), aún cuando externamente podamos distinguir entre un organismo sano y uno enfermo, por ejemplo. Si no fuera así, no tendría sentido hablar de conservación sistémica de autopoiesis.

Es simplemente imposible derivar a partir de la definición de autopoiesis el hecho de que un ser vivo va a intentar mejorar su situación cuando ésta muestre una tendencia negativa para su conservación (por ejemplo, la bacteria que sube el gradiente de azúcar para compensar la falta de nutrientes). El mejorar las condiciones de auto-producción no es parte de la definición de autopoiesis, y no puede ser deducido a partir de esta definición,

la cual simplemente establece que el sistema se auto-produce, nada más. Por lo tanto, un sistema autopoietico que es capaz de operar de acuerdo a estados potenciales no actualizados (evaluando y evitando, por ejemplo, una situación de riesgo) no puede lograr esto simplemente por el hecho de ser autopoietico. Necesita algo más. Y dado que los sistemas cognitivos a menudo se nos aparecen como capaces de lograr precisamente este tipo de regulación orientada hacia estados potenciales, parecería que la autopoiesis “clásica” y la cognición no pueden seguir un mismo camino a partir de este punto. Un sistema meramente vivo, autopoietico “a secas”, no puede ser un sistema cognitivo.

Más allá de la autopoiesis

Sería prematuro simplemente abandonar la doctrina de la conservación de la teoría de la autopoiesis (la noción de la autopoiesis como un todo o nada), aun cuando pareciera tratarse del principal obstáculo para la naturalización de la cognición sospechada por Varela. El análisis en términos de teorías de clases y basado en propiedades organizacionales es una de las contribuciones más importantes de la idea de autopoiesis. Eliminar este análisis haciendo de la autopoiesis un concepto más “blando” (es decir haciéndolo algo relativo y capaz de sufrir quiebres parciales) significa revertir la situación a una perspectiva menos clara sobre los sistemas vivos definiéndolos a partir de una lista de propiedades (crecimiento, reproducción, sensibilidad, etc.), es decir, sería volver a la misma perspectiva que la teoría intenta superar. Además, semejante jugada no le haría ningún favor a una teoría naturalizada de la cognición dado que las diferencias en capacidades y propiedades cognitivas serían demasiado fácilmente atribuibles a diferencias metabólicas y el resultado sería una teoría bastante débil. El dominio de lo mental no estaría enraizado en la vida, sino reducido a ella. Tal reducción no nos permitiría explicar cómo es posible que a menudo se encuentre significado en situaciones que no tienen un efecto directo sobre el metabolismo. En otras palabras, la mente no poseería una autonomía con respecto al de la vida.

Es necesario aceptar los términos de la cuestión e intentar una solución que provea la propiedad requerida para poder entender las nociones intuitivas que pertenecen a la cognición, y al hacerlo, poder formular una definición operacional de este concepto sin apelar a la intervención regulativa de lo interno, tal como el caso del principio de paridad en la hipótesis de la mente

extendida (Clark y Chalmers, 1993, Wheeler, 2010). Tal propiedad nos ayudaría a entender la noción de quiebres parciales en autopoiesis sugerida por Varela. Tal propiedad debería 1) presentarse de manera gradual en los sistemas vivos, 2) responder en forma diferencial a situaciones diferentes de acuerdo a las consecuencias que éstas implican para el organismo, 3) a veces no funcionar bien, ser capaz de errores, 4) obedecer el axioma de determinismo estructural (es decir, no tratarse de algo mágico), y aun obedeciéndolo 5) permitir al organismo vivo alterar sus operaciones en el presente con respecto a una evaluación de situaciones potenciales.

Para poder formular esta propiedad lo primero que debemos hacer es reconocer las diferencias que pueden ser apreciadas por un observador externo al observar sistemas autopoieticos. Podemos darnos cuenta cuando a un organismo enfermo le quedan horas o días de vida. De manera similar, podemos distinguir organismos que viven situaciones de riesgo de aquellos que pasan por una situación confortable. Por supuesto que ninguna de estas distinciones es perfecta o absoluta (el organismo enfermo puede recuperarse; el riesgo de una situación puede ser ilusorio o su comodidad ocultar un peligro subyacente). Como observadores, podemos modular de forma normativa nuestras acciones cuando realizamos tales distinciones en otros organismos. Pareciera que bajo la condición de viabilidad “todo-o-nada” dada por la doctrina de la conservación de autopoiesis, yaciera un espacio de diferencias estructurales graduales y cualitativas las cuales pueden ser evaluadas de manera normativa en el acto cognitivo de la observación. Por ejemplo, es generalmente preferible evitar situaciones riesgosas que salir a su búsqueda, el estado de salud es preferible al de la enfermedad, etc. Lo que nuestra historia de naturalización requiere es que no seamos nosotros, los observadores externos, sino el organismo mismo el que tenga la capacidad de generar tales normas (a su manera) mientras esté aún vivo y que a partir de ellas pueda regular sus operaciones de manera adecuada dentro del espacio de opciones estructurales que corresponde a la conservación del organización autopoietica. Esta capacidad es la *adaptividad*.

Los sistemas autopoieticos existen lejos del equilibrio termodinámico y deben tolerar las tendencias entrópicas naturales manteniéndose energética y materialmente abiertos. Dicho de otro modo, estos sistemas son robustos en el sentido de que son capaces de soportar un rango de perturbaciones externas y un rango de cambios estructurales internos sin perder su autopoiesis. Los límites de estos rangos están determinados por la estructura

y el estado actual del sistema. Nos referimos a los estados contenidos dentro de estos límites como el *conjunto de viabilidad* del sistema (en general es posible medir o estimar algunos aspectos de este conjunto, por ejemplo, en variables que deben mantenerse acotadas como la temperatura de la sangre en los mamíferos). Si la trayectoria de los estados del sistema se aproxima a los límites de su conjunto de viabilidad y los atraviesa, el sistema muere. La adaptividad es una manera particular de tolerar los desafíos ya sean internos o del entorno, a través de una sensibilidad a las perturbaciones seguida de operaciones cuyo efecto es la compensación de las tendencias negativas que ellas puedan crear. Definimos a la adaptividad como:

la capacidad que tiene un sistema, en ciertas circunstancias, de modular la dinámica de sus estados y su relación con el entorno con el resultado que, si los estados están suficientemente cerca de los límites de viabilidad,

1. las tendencias de los estados se distinguen a través de respuestas sistémicas diferenciales dependiendo de si los estados se aproximan o se alejan de los límites de viabilidad próximos y, como consecuencia de esta distinción,
2. las tendencias de los estados que tienden aproximarse a los límites de viabilidad, se acercan a, o se transforman en, tendencias que no tienden a aproximarse a esos límites de manera de que los estados que sobrevienen en tales tendencias no cruzan estos límites de viabilidad, (ver Di Paolo, 2005, p. 438).

Un sistema autopoietico adaptivo es capaz de operar de manera diferencial en (al menos algunas) situaciones en las que, si se dejaran desarrollar sin cambio alguno, tenderían en general hacia la pérdida de la autopoiesis. Esta propiedad es perfectamente operacional, y sin embargo no esta implicada en la definición de autopoiesis. La adaptividad nos permite recuperar la interpretación homeostática sin los problemas que de otra forma conllevaría. Un *quiebre*, para el sistema mismo, es simplemente la severidad de una tendencia negativa (una tendencia de los estados del sistema a aproximarse al los límites cercanos de la frontera de viabilidad). Tal severidad se distingue y se mide por el sistema mismo a través de las restricciones temporales que se le impone y la cantidad de recursos regulativos que demanda su compensación. Un quiebre que da poco tiempo en relación a los recursos y posibles respuestas que el sistema puede ofrecer para compensar sus

tendencias negativas es más serio que uno que permite un tiempo mayor (nótese cómo la adaptividad nos brinda una manera natural de hablar de temporalidad). Un quiebre es típicamente (aunque no siempre) el resultado de una perturbación externa. Además de responder a quiebres cuando se presentan, la capacidad de adaptividad le permite al organismo la posibilidad de eludirlos, de evitar situaciones riesgosas y buscar aquellas más seguras.

Solo en el sub-conjunto de sistemas autopoieticos que son no sólo robustos sino también adaptivos, podemos encontrar sistemas capaces de distinguir potencialmente entre las distintas consecuencias virtuales que poseen los encuentros con el entorno que, en el presente, son viables en igual medida. Esta operación diferencial es lo que llamamos la búsqueda de sentido o “*sense-making*” (Weber y Varela, 2002, Di Paolo, 2005, Thompson, 2007, etc.). Podemos definir este concepto como la capacidad de regular la dinámica de los estados (internos y relacionales) en función de sus consecuencias virtuales para la conservación de una forma de vida. Si, como se propone, la búsqueda de sentido requiere de “una valencia con una base dual: atracción o repulsión, aproximación o escape”, (Weber y Varela, 2002, p. 117), un sistema que se encuentra en la búsqueda de sentido requiere, además de la norma que surge de su propia auto-construcción, un acceso a cómo se sitúa en este momento con respecto a la barrera todo-o-nada que expresa esa norma. Para poder disponer de tal acceso mientras opera en base a sus estados actuales, el sistema autopoietico deber ser capaz de reconocer en esos estados, y sólo en ellos, las tendencias virtuales que describen su relación como un todo con respecto a la pérdida potencial de su propia viabilidad, y además, deber ser capaz de actuar de forma apropiada en base a esas tendencias. Tales intervenciones bien pueden fallar dado que en el reconocimiento de tendencias virtuales en estados actuales no se garantiza la realización de tales tendencias. Es casi seguro que saltar el precipicio tendrá consecuencias poco saludables para la autopoiesis, pero no es certero (por ejemplo si activamos el paracaídas a tiempo o una fuerte corriente de aire ascendente amortigua la caída). Las intervenciones también pueden fallar aun cuando las tendencias hayan sido identificadas correctamente, simplemente porque estas pueden ya ser tan severas que no permiten una compensación efectiva. Lo que hace a la adaptividad un concepto operacional no es la certeza sobre un futuro que no se puede conocer hasta que sucede, sino la virtualidad ya presente en los estados actuales tal cual son. La búsqueda de sentido no es más que la regulación, más o menos

sofisticada, de estados del sistema con respecto a esta virtualidad. Esta es la base de la regulación normativa de la operación del organismo y que incluye la forma en que el organismo regula como una unidad sus interacciones con el entorno. Semejante regulación normativa con respecto a las consecuencias virtuales de las tendencias actuales es el sello de todo tipo de cognición (aunque, como veremos, no todas las normas surgen directamente a partir de la lógica del metabolismo).

La adaptividad no sólo nos provee de una herramienta conceptual que allana el camino entre la autopoiesis y la cognición, sino que también nos ayuda a tematizar varios fenómenos biológicos de importancia sin entrar en contradicciones con la teoría autopoietica. Para más discusión, ver (Di Paolo, 2005). Para nuestros motivos presentes, un punto crucial es que cuando los mecanismos adaptivos operan más allá de la frontera física del organismo, de modo de que terminan regulando su acople con el entorno, nos desplazamos conceptualmente desde una perspectiva de *acople estructural* (que es esencialmente simétrica e implica que tanto sistema como entorno se influyen mutuamente sin pérdida de viabilidad) hacia una perspectiva de *comportamiento* (concepto asimétrico según el cual el organismo es el origen de una regulación en los parámetros y condiciones del acople estructural). Esta regulación de las interacciones con el entorno dentro del contexto de búsqueda de sentido es lo que nos permite definir a ciertos sistemas autopoieticos y adaptivos como *agentes* (Barandiaran, Di Paolo, Rohde, 2009). El concepto de agencialidad es asimétrico porque implica una *regulación* del acople, el cual ya de por sí es de entrada un bucle de mutuas influencias entre organismo y entorno, no su mera existencia. La modulación implica un segundo bucle de orden superior siguiendo las normas dadas por la adaptividad del agente (Figura 1). Este bucle de segundo orden rara vez se considera de manera explícita (aunque ya aparece como idea central en la teoría del comportamiento adaptivo de Ashby, 1960). La distinción entre acople con el entorno y la regulación normativa de este acople no figura en el funcionalismo (ni en la teoría de autopoiesis de forma explícita). Es precisamente gracias a esta distinción que ahora podemos definir a la *cognición* como *la búsqueda de sentido que un agente realiza en un dominio de interacciones con el entorno*. Se entiende esta noción como la de una característica básica y común a todo tipo de cognición, sin distinciones por el momento entre elementos racionales, afectivos, de memoria o aprendizaje, predictivos, perceptivos, etc. De esta forma, podemos capturar la mayoría de

las intuiciones importantes asociadas a la cognición (quizá incluso todas): se trata de un concepto normativo, asimétrico (existe un centro de actividad cognitiva), relacional, orientado y extendido en el tiempo, capaz a veces de fallar, e implica la existencia de una identidad auto-constituida de cuya continuación surgen las normas a seguir.

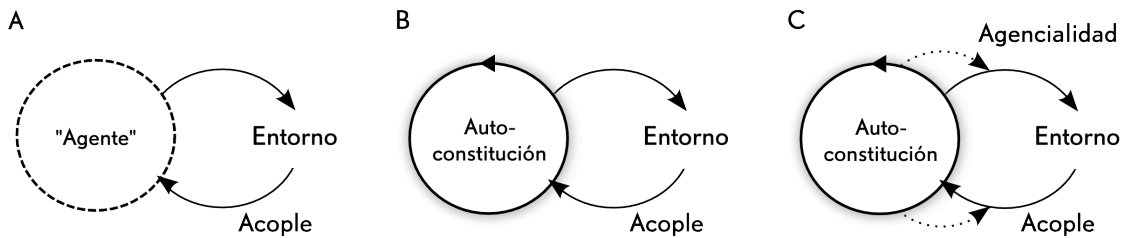


Figura 1 Nociones de agencialidad. A) Un agente definido por convención en interacción con su entorno, situación tradicional en el funcionalismo. B) Un sistema auto-definido a través de un proceso de auto-mantenimiento (como la autopoiesis). C) Definición de un agente: un sistema auto-definido que regula de manera adaptiva su acople con el entorno.

Autonomía, substrato material y precariedad

Como hemos dicho, el enactivismo es un proyecto de naturalización de la mente. En la sección anterior hemos esbozado con algún detalle cómo el enactivismo es capaz de proponer y desarrollar definiciones operacionales y no-circulares de algunos términos básicos que necesita toda teoría de la mente: autonomía, identidad, sentido, significado, agencialidad, comportamiento, cognición. Estos términos, aunque parezca mentira, no tienen definición operacional en teorías funcionalistas como el cognitivismo clásico o la mente extendida. Pero, para afianzar este avance, es necesario plantearse y responder una serie de preguntas. Después de todo es posible que este desarrollo teórico del enactivismo pudiera ser asimilado a una versión funcionalista que complementa al computacionalismo tradicional. Y también cabe preguntarse si la propuesta del enactivismo no es un poco limitada al ligar las raíces de la mente demasiado fuertemente a la lógica del organismo viviente. En esta sección trataremos estas cuestiones. Veremos cómo la introducción de otras nociones clave como la de la identidad no-metabólica, la clausura operacional y la precariedad, ayudan a responder a

estas preguntas de forma simultánea. La otra pregunta fundamental es cuál es la diferencia que aporta el enfoque enactivo a la investigación concreta de la cognición. Volveremos sobre ese tema en forma breve sobre el final del capítulo.

Aunque podemos considerar un gran avance el poder superar a través de la noción de adaptividad los obstáculos presentados por la doctrina de conservación en la teoría de la autopoiesis – confirmando de esta manera la posibilidad de naturalizar la cognición partiendo de la organización de la vida – existen obvios problemas en proponer que la mente no es más que una consecuencia de la autopoiesis adaptiva. Por ejemplo, ¿cómo entender los valores y las normas que guían al comportamiento y que sin embargo están sub-determinados por el metabolismo? ¿No es acaso la posibilidad de una disfunción adaptiva un indicador de que la conexión entre vida y mente es meramente contingente? ¿Cómo se entiende que podamos actuar siguiendo valores que van en contra de una satisfacción inmediata o incluso futura de una necesidad metabólica? ¿Es la mente necesariamente realizable solamente a partir de un substrato biológico? Resolver estas cuestiones demanda la adopción de una perspectiva más amplia, una perspectiva que nos permita tematizar la irreducibilidad del dominio cognitivo al substrato metabólico, algunas de cuyas propiedades le proveen su condición de posibilidad (y de paso eliminar de una vez por todas la sospecha de es necesario interpretar lo mental y lo viviente como dominios co-extensivos).

En el enfoque enactivo, los orígenes de la búsqueda de sentido se encuentran en las capacidades regulativas y normativas del organismo. Pero para ser más precisos y para aclarar la relación vida–mente, debemos decir que estos orígenes están ligados no tanto a la vida en todos sus aspectos, sino a uno en particular: la presencia de un proceso de generación de identidad. Esto es lo que caracteriza a los sistemas llamados autónomos – adaptando la idea de autonomía propuesta por Varela (1979) y sumándole el requerimiento de precariedad, como veremos. La vida del organismo es una forma de autonomía precaria, pero no la única posible. El razonamiento presentado más arriba se aplica en general a los sistemas autónomos precarios y adaptivos, no sólo a la vida orgánica. Por eso debemos aclarar que se entiende precisamente por autonomía.

Un *sistema autónomo*, en cualquier dominio, se define como un sistema compuesto de varios procesos, los cuales de forma activa generan y sostienen

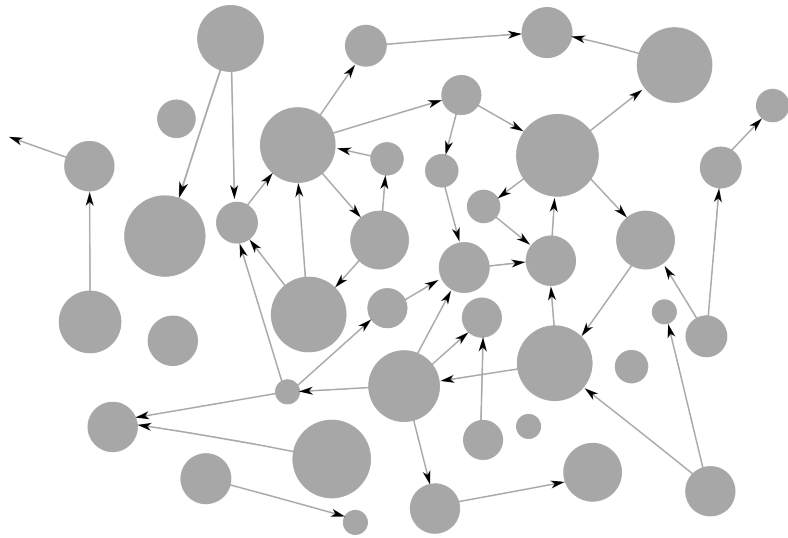
una identidad bajo condiciones de precariedad. En este contexto, “generar y sostener una identidad” significa poseer la propiedad de *clausura operacional*. Esta es la propiedad que especifica que entre las condiciones que afectan y permiten la realización de cada proceso constituyente del sistema siempre encontraremos uno o más procesos que también pertenecen al sistema. Y, además, cada proceso en el sistema es una condición para la existencia de al menos otro proceso constituyente, formando así una red de relaciones de condicionamiento entre todos los procesos que conforman el sistema.

Dicho de otro modo, no existen procesos en el sistema que no estén condicionados por otros procesos en la red cerrada de relaciones, lo que no significa, esta claro, que procesos externos a la red no puedan también influir en los procesos constituyentes, solo que tales procesos externos no forman parte de la red cerrada porque no dependen a su vez de ningún proceso constituyente (ejemplo: la fotosíntesis en plantas depende de la luz solar para existir como proceso; no así en el sentido inverso por eso el sol no es parte de la planta). De manera similar, pueden existir procesos que sean afectados por procesos constituyentes del sistema, pero que en sí mismos no condicionan a ninguno de ellos y por lo tanto no forman parte de la red.

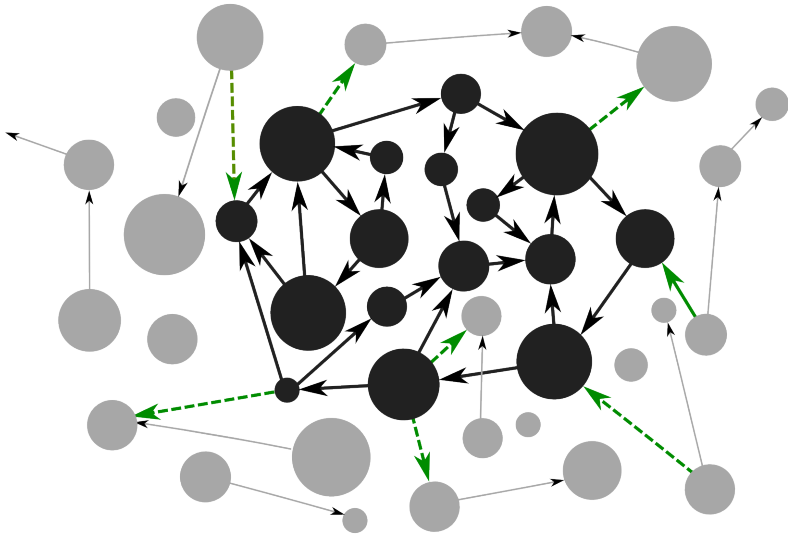
La figura 2 ilustra este concepto. Supongamos que como observadores investigamos una serie de procesos (midiendo temperaturas, movimientos, interacciones, reacciones químicas, etc.). Representamos cada uno de estas observaciones como los círculos en la figura 2.A. La elección de qué procesos estudiar dependerá de qué decidimos observar. La investigación y manipulación de estos procesos lleva a diferentes observaciones sobre las relaciones entre los mismos. Si al intervenir en uno de ellos observamos un efecto sobre otro, dibujamos una flecha. Estas son relaciones de condicionamiento. Más estrictamente, vamos a interpretar las flechas como relaciones de necesidad, significando que sin la presencia o funcionamiento de un proceso, el proceso afectado (indicado con la flecha) no podría existir o funcionar (por ejemplo, sin la presencia de una fuente de calor, el proceso de ebullición del agua a temperatura y presión atmosféricas normales no puede ocurrir). Es posible que a medida que descubrimos, a través de intervenciones, este tipo de relaciones de necesidad entre los círculos encontremos un sub-conjunto de ellos con la siguiente característica: en cada círculo en el sub-conjunto vemos que hay flechas que llegan a él y flechas que salen de él con respectivos orígenes y destinos en otros círculos del sub-conjunto (figura 2.B). Topológicamente, este subconjunto conforma una red

cerrada de condiciones de necesidad. A esto llamamos clausura operacional. A pesar de que como observadores elegimos qué es lo que vamos a observar (o sea cuáles son los círculos que nos interesan), el hecho de que los procesos observados formen o no una red cerrada con las características indicadas estado por las relaciones que exhiben, es decir, depende del observador que elige el nivel de observación. Una vez elegido el nivel, la granularidad espacio-temporal, etc., la presencia o no de clausura operacional no es arbitraria. Si decidimos observar a los procesos de la manera adecuada, vamos a descubrir que existe una clausura operacional (si es el caso) y si los observamos de manera distinta, casi seguro no la veamos. La clausura operacional permite que el sistema cerrado se auto-distinga, dando continuidad a sus propias condiciones de constitución (las relaciones entre procesos esbozadas). De forma interesante, vemos que el concepto de autonomía yace en un intersticio entre la epistemología y la ontología. Depende de nuestra forma de observar el mundo (no vamos a descubrir un ser vivo si decidimos seguir la trayectoria de una molécula o de un planeta), pero en el nivel correcto, la realidad de la clausura operacional se anuncia al observador independientemente de sus intereses dada simplemente la lógica de su organización.

Está claro que la definición de clausura operacional no prescribe la posibilidad de que haya flechas entre procesos que pertenecen a la red y otros que no pertenecen a ella. Sólo que estas flechas son unidireccionales entre tales procesos y la red. Estos procesos que afectan o son afectados por los procesos que constituyen la red conforman el entorno de la red, su ambiente (figura 2.B). Un sistema operacionalmente cerrado no implica de ninguna manera una separación del entorno, o una independencia del mismo. Precisamente la forma de determinar si un proceso es constituyente de la red es verificar operacionalmente si cumple la relación de clausura (en la figura si salen flechas hacia procesos constituyentes y a la vez si recibe flechas de procesos constituyentes). En este sentido estricto decimos que la clausura operacional es la definición de una *identidad*. En principio, un observador científico puede determinar experimentalmente estas relaciones. Y en consecuencia lo mismo podemos decir de la noción de identidad. En su dependencia mutua, la red de procesos se cierra sobre sí misma y define una unidad que se auto-distingue de su entorno (permaneciendo en interacción con el mismo y de hecho definiendo qué procesos lo conforman) y se regenera a sí misma en el espacio donde ocurren estos procesos.



A)



B)

Figura 2. Ilustración del concepto de clausura operacional utilizado para definir a la autonomía de un sistema auto-mantenido. A) Procesos identificados por un observador (círculos) y sus relaciones de necesidad. B) Sub-conjunto autónomo definido por procesos e interacciones que conforman un red cerrada de condicionamiento (círculos y flechas negras); flechas a rayas: relaciones de condicionamiento entre sistema autónomo y su entorno.

Hemos mencionado que las condiciones en que estas relaciones entre procesos aparecen son condiciones *precarias*. Por esto entendemos circunstancias en las que los procesos constituyentes, si fueran aislados, tenderían a detenerse y desaparecer. Los procesos individuales se extinguen en la ausencia de la organización del sistema, aun cuando el resto de las condiciones físicas fueran idénticas. Los procesos constituyentes no están simplemente condicionados (modulados, ajustados, o modificados por otros procesos) sino que también dependen para su continuación de la propia existencia de la red organizacional que ellos mismos sostienen. La red toda, en consecuencia, les permite seguir existiendo y no podrían hacerlo de manera aislada.

Es importante recalcar que el concepto de precariedad juega un papel crucial en el enfoque enactivo. Es un concepto que como mucho puede a veces interpretarse como implícito en algunos pasajes de la literatura autopoietica, pero que nunca se propone de manera explícita como un factor importante, quizá el factor que realmente hace que la idea de clausura operacional cobre su verdadero sentido. La precariedad es un requisito esencial para evitar interpretaciones triviales de la clausura operacional. De este modo se convierte en una herramienta conceptual efectiva para generar explicaciones capaces de capturar con cierta justicia la fenomenología de la vida y de la mente. No es novedoso hablar de auto-organización en sistemas dinámicos. Los patrones de activación en una red neuronal artificial pueden llegar a puntos en que se “auto-mantienen”. O autómatas celulares como el llamado Bittorio (propuesto como ejemplo de clausura operacional en Varela et al., 1991) también convergen en atractores donde la aplicación de las reglas de transformación al un estado resulta en el mismo estado. Tales ejemplos podrían considerarse como operacionalmente cerrados, si interpretáramos de forma suficientemente liberal la definición de clausura operacional.

Sin embargo, si en estos ejemplos tomáramos como procesos constituyentes relevantes a las mismas reglas dinámicas de transformación de estados (que después de todo son las que definen la operación del sistema), estas reglas en sí mismas no dependen ni son constituidas por la organización de patrones auto-mantenidos en los estados. Las reglas de transformación (por ejemplo, en el caso de una red neuronal, una suma ponderada de valores de entrada para determinar el grado de activación de cada nodo) son independientes y

permanecen inalteradas existan o no estados auto-mantenidos en estos sistemas (cierta recursividad es posible, por ejemplo en sistemas capaces de aprendizaje, pero el argumento simplemente de transfiere de las reglas a las meta-reglas de transformación las cuales permanecen fijas). Por lo tanto estos procesos constituyentes (las reglas) no cumplen la condición de precariedad.

Una vez establecido un sistema operacionalmente cerrado (en el caso en que esto suceda) a partir de procesos que no sólo conforman una red cerrada, sino que además, no serían capaces de sostenerse a sí mismos a nivel individual a menos que estuvieran organizados a través de semejante red, podemos entonces decir que estamos en presencia de un sistema precario operacionalmente cerrado. Esto es lo que llamamos un *sistema autónomo*. De forma esencial, en tal sistema, los procesos que lo componen no estarían presentes o no durarían lo suficiente, de no ser por las condiciones que se mantienen gracias al resto de la red. Tal sistema operacionalmente cerrado no sólo se auto-mantiene: se brinda a si mismo las condiciones de su propia existencia. Como consecuencia los procesos que componen al sistema autónomo desvanecen la distinción (fundamental para el funcionalismo) entre operador y aquello sobre lo que el operador actúa. El decaimiento de cada proceso debe prevenirse a través de la actividad de otros procesos cuya operación es a la vez afectada por la manera específica en que su propio decaimiento es evitado. De esta manera, por tener que actuar en contra de las tendencias naturales de su propio substrato material, un sistema autónomo y precario debe “auto-postularse”. No es fortuito entonces que el sistema dependa de recursos materiales y energéticos (como sería el caso de cualquier sistema lejos del equilibrio termodinámico), sino que la organización misma del sistema, su forma de contrarrestar su propio decaimiento, depende de manera específica de la precariedad material interna que debe contrarrestar para seguir existiendo.

Un sistema autónomo y precario, a cualquier nivel en que exista, interviene en su propio substrato y como consecuencia mantiene una forma que se constituye de componentes que paradójicamente, dada su precariedad, proveen las tendencias hacia la disolución misma de la forma. Esta persistencia a través de una doble negación, esta forma de “*suicidio frustrado*” que es la verdadera autonomía y la vida como caso particular, esta en desacuerdo con el funcionalismo, entendido como el intento de dar una explicación de la operación del sistema en términos que son independientes

del substrato. La precariedad no se refiere a ninguna propiedad positiva de la materia, la cual puede ser capturada funcionalmente, sino a la no permanencia, en las escalas temporales pertinentes, de cualquier propiedad y configuración relevante del substrato. Respondemos entonces negativamente a una de nuestras preguntas: *el enfoque enactivo no puede ser capturado de manera completa por el funcionalismo*. Es por lo tanto un enfoque abiertamente no-funcionalista.

Es importante destacar que el papel que juega la precariedad en esta historia y que permite al enactivismo radicarse de forma no-reduccionista en la materialidad, es un papel necesario para que la noción fundamental de autonomía sólo tenga instancias empíricas no triviales, como el caso de la vida misma. No se trata de algo meramente inevitable (universal como la segunda ley de la termodinámica): la precariedad del substrato material es *necesaria* para que la vida pueda tener definición no trivial. La vida no tendría un mejor pasar en la ausencia de precariedad sino que simplemente no sería vida. Sería una permanencia indiferente. Por el contrario, la precariedad nos presenta con una visión de la vida como inherentemente inquieta y activa, en constante necesidad; una visión que se ajusta mejor a nuestra experiencia propia de la vida, nunca en estado completamente seguro, siempre negociando tiempo para sí misma.

Por otra parte, el explicitar los aspectos fundamentales que garantizan el silogismo entre autonomía y cognición, nos permite entender que la mente no es un fenómeno sólo biológico, sino un fenómeno que surge de la autonomía de ciertas organizaciones materiales precarias. Tales organizaciones, en principio podrían existir sin auto-producción molecular (autopoiesis), en redes de relaciones de producción más genéricas. Esto dato es relevante para proyectos de vida e inteligencia artificial y para entender las ideas de transiciones en formas de identidad mental, o más simplemente en formas de vida.

Con raíces en la vida, la mente se mueve a sí misma

La noción de autonomía nos permite establecer un espectro más amplio para la relación entre un substrato de identidad auto-sostenida (que puede ser la vida misma o, como veremos, identidades a otros niveles de autonomía) y la cognición (entendida como hemos dicho en general y sin intervenir todavía

distinciones entre sus distintos aspectos). Es necesario para esto considerar la lógica que vincula a los procesos que resultan en el mantenimiento autónomo de una identidad precaria con la relación normativa, teleológica, teñida de valores que existe entre tal identidad y su entorno (es decir, la búsqueda de sentido). La autopoiesis adaptiva es sólo una instancia (quizá la más fundamental) donde somos testigos de dicha lógica en funcionamiento.

El filósofo Hans Jonas ha explorado precisamente esta lógica en su enfoque existencial sobre la vida (Jonas, 1966). El hecho de que el metabolismo sostiene una forma dinámica de identidad (la cual no coincide con su constitución material instantánea en ningún momento, excepto cuando el sistema está muerto) es el que permite a un organismo ser libre. Esta libertad se expresa en la capacidad del organismo de vincularse con su medio en términos del significado de una situación, contribuyendo de ese modo a la continuación de su autonomía y abriendo incluso la posibilidad de crear nuevos valores y nuevos significados. Sin embargo, esta libertad está condicionada a necesidades materiales estrictas y específicas. Es una “libertad en la necesidad” en las palabras de Jonas. Lejos de ser paradójico, este concepto de libertad evita las objeciones del determinismo porque opera en la relación de *mediación* que existe entre una identidad auto-sostenida y el “objeto” de sus interacciones cognitivas. En ese sentido, el modo de realización de una red autónoma de mantenimiento de identidad (como la autopoiesis) establece el tipo de acceso, por parte del sistema, a las normas que describen sus diferentes formas de viabilidad. Tal acceso puede ser menos o más mediado (la diferencia, por dar un ejemplo, entre reaccionar negativamente al contacto con una superficie caliente versus planificar nuestro movimiento para evitar entrar en contacto con ella, siendo ambos casos ejemplos de regulación adaptiva). La propuesta de Jonas es que en la historia de la vida y la mente han aparecido formas novedosas de vínculos entre sistema autónomo y entorno. Estas formas están caracterizadas por una relación de mediación creciente, permitiendo una mayor libertad a costo de una mayor precariedad.

Un claro ejemplo de tales transiciones en las formas de mediación lo da la vida animal, o más específicamente, su especial forma de auto-movimiento y sus propiedades espaciales (por ejemplo, la reversibilidad). El advenimiento de la movilidad animal establece un nuevo orden de valores y la co-emergencia de la *percepción*, la *acción* y la *emoción*, hasta este punto formas no diferenciadas de la búsqueda del sentido. Al poner una distancia, tanto en el espacio como

en el tiempo, entre las tensiones de la necesidad y la consumación de su satisfacción, la temporalidad inherente a la adaptividad (Di Paolo, 2005) se “espacializa”. Un animal capaz de buscar sentido en eventos distantes puede apreciar en este mismo instante el peligro que acecha a lo lejos y que aún no está presente en sus efectos directos sobre la conservación de su modo de vida. Este es el origen de una relación particular con el mundo, la de la percepción y la acción, la cual está cargada de significados internos múltiples en correspondencia a la multiplicidad de posibilidades en esta relación. Es por lo tanto a la vez el desarrollo de una dimensión emocional (lo que podría ser la vida interior de los ritmos de necesidad y satisfacción se enriquece en posibilidades mediadas por las potencialidades de movimiento, tales como el temor, el deseo, el esfuerzo, la aprehensión, el relajamiento, la curiosidad, etc.)

Este enriquecimiento, tanto en los aspectos externos de la agencialidad como en la dimensión interior de la cognición animal, surge, sin embargo, al costo de nuevas formas de riesgo y de demandas energéticas más severas (por ejemplo, para permitir el movimiento continuo y a menudo veloz a través de distintas condiciones del entorno que por largos períodos no permiten la renovación de energía). Tales transiciones no sólo ocurren con la aparición del auto-movimiento animal, también ocurren ante el surgimiento de distintas formas de movimiento. Por ejemplo, varias especies de insectos como la familia *Corixidae* de la orden Hemiptera, son capaces de respirar bajo el agua gracias a unas espinas hidrófugas abdominales que atrapan burbujas de aire e impiden que colapsen. Las burbujas renuevan su nivel de oxígeno gracias a la diferencia de presión parcial con el agua a medida que este se consume por la respiración del insecto. En principio estos “pulmones externos” podrían funcionar de manera casi indefinida (Thorpe, 1950, Turner, 2000). Esta regulación adaptiva y mediada a través de un elemento extra-corporal (las burbujas de aire) permite el acceso a largos períodos de movimiento bajo el agua y, es de suponerse, acceso a recursos distintos (nutrientes, protección, etc.). Aun así, la ampliación de posibilidades de movimiento conlleva un mayor riesgo que la respiración normal en el aire. La mediación en tales casos está tan íntimamente conectada con las funciones vitales que podríamos decir que el sistema vivo mismo se ha extendido. Tal aseveración puede ser justificada a través del concepto de clausura operacional si se puede identificar a la burbuja (o a algún proceso que la burbuja permite) como perteneciente a la red cerrada de procesos

metabólicos. Lo que esto sugiere es que en ciertos casos, si la nueva relación de clausura se verifica, estas formas de mediación de orden interactivo que se conservan de forma sostenida, ponen en cuestión la identidad misma de tales sistemas extendidos. En otras palabras, si se cumplen los requisitos necesarios, un elemento externo, la burbuja, puede pasar a ser una parte constitutiva del organismo.

Jonas identifica otras transiciones en las formas de mediación a través de la historia de la vida y la mente. No sólo el auto-movimiento animal en sus diversas formas abre la puerta a transiciones en formas de identidad y consecuentes transiciones en el tipo y la complejidad de la búsqueda de sentido, sino que también existen muchas otras transiciones. Un ejemplo son las transiciones que permite la presencia de un sistema visual complejo capaz de percibir una variedad de estímulos en relación mutua en el mismo instante, y por ende una red compleja de relaciones de valores. Otro ejemplo, es la capacidad de entender, crear y manipular imágenes (Jonas asocia esta capacidad específicamente con los seres humanos), lo cual permite la regulación del comportamiento respecto a eventos no presentes, ya sean del pasado o que aún no forman parte de la realidad y hacia los cuales podemos apuntar. También permite la instauración de criterios de verdad y falsedad (adecuación entre imagen y percepción o memoria). Y esta capacidad puede, en una transición adicional, ser recursivamente aplicada al agente cognitivo mismo, convirtiéndolo en su propio proyecto y regulador no sólo de sus comportamientos y mediaciones sino también de su identidad misma (ver Jonas, 1966 para la elaboración de estas relaciones entre formas de mediación y formas de autonomía). Lo que queda claro es que estas transiciones no implican una ganancia intrínseca, ciertamente no a nivel metabólico, pues lo que logran es expandir el dominio de la autonomía a costas de una precariedad mayor. Sin embargo, Jonas apunta que “el estándar de la sobrevivencia no es adecuado para una evaluación de la vida” (Jonas, 1966, p 106). Sigue:

“Es una de las paradojas de la vida el emplear medios que modifican el fin y que se convierten ellos mismos en parte de él. El animal que siente puja por preservarse como una entidad que siente y no sólo como una entidad que metaboliza, es decir, puja por continuar la actividad misma del sentir: el animal que percibe puja por preservarse como una entidad que percibe — y así en general. Sin estas facultades, habría mucho menos que preservar, y es eso que en tal caso

no se preservaría lo que al mismo tiempo permite su propia preservación” (ibid. traducción del autor)

En efecto, tales transiciones en formas de mediación inauguran un dominio que se retro-alimenta: implican una nueva *forma de vida*. No sólo en un sentido metafórico, sino en el sentido estricto de que dan origen a un proceso de mantenimiento de identidad que esta sub-determinado por el metabolismo: una mente que se mueve a sí misma.

¿Cómo es posible todo esto? ¿Podemos entenderlo estas transiciones y subsecuentes transformaciones en términos sistémicos? El problema que plantea la relación entre los aspectos constructivos e interactivos de la organización viviente (es decir el problema de cómo una relación de *mediación* se puede transformar en una relación de *constitución*) se encuentra ya presente en la teoría de la autopoiesis. Pero se hace difícil apreciar este problema (y mucho menos resolverlo) desde los términos propuestos por esta teoría. El obstáculo es la imposibilidad de pasar del dominio operacional/constitutivo al dominio relacional/interactivo dado que se consideran dominios “sin intersección”. El primero corresponde al funcionamiento de la red autopoiética que termina por constituir una unidad, un sistema compuesto. El segundo corresponde a las relaciones en las que tal unidad entra en su acople estructural con el entorno (ver, por ejemplo, Maturana, 2002). A pesar de la contabilidad lógica que esta separación entre “dominios sin intersección” permite, poco ha dicho la teoría de autopoiesis respecto al inconveniente que tal división ha creado: un análogo sistémico del dualismo mente-cuerpo. De hecho, la teoría de autopoiesis no se pronuncia respecto a cómo se coordinan las relaciones interactivas del organismo con las compensaciones internas de su constitución. Simplemente éstas deben estar coordinadas o no existe el sistema autopoiético. De modo que en la interpretación originalmente propuesta (la irreducibilidad entre dominios entendida como no-intersección) la teoría de autopoiesis es estrictamente Cartesiana en la forma en que Jonas y enactivismo (y hasta los mismos Maturana y Varela en otros escritos) intentan evitar[†]. La solución al estilo Malebranche de este dualismo

[†] Debemos aclarar la intención que hay detrás de la distinción entre estos dos dominios: a saber, prevenir un proyecto de reducción de fenómenos interactivos y constructivos entre los mismos. Tal reducción es, por ejemplo, el proyecto del representacionalismo el cual busca dar eficacia funcional y mecánica a relaciones intencionales, como el acto de representar, cometiendo de esta manera una instancia particularmente severa de la falacia mereológica, es decir asignar a partes

— a saber, que un dios se encarga de coordinar las sustancias que no se intersectan causalmente — toma la forma en este caso de una apelación a la evolución biológica; apelación latente en la frase “si no fuera así el sistema se desintegraría”; si la coordinación entre dominios no existiese, no habría ser vivo. Si los hay, es porque los dominios están en armonía. De esta manera, la evolución, como el dios malebrancheano, se encarga de filtrar aquellos organismos poco felices para los cuales los dos dominios no están coordinados.

El principal problema con esta solución no es sólo que el dualismo permanece sin resolución sino que además la relación entre los dominios que resulta es puramente contingente. Simplemente sucede que es la relación que es porque esto ha ayudado al organismo a sobrevivir, y no en virtud de ninguna implicación interna entre el dominio de constitución y el de interacción. Esto no es suficiente para arraigar la cognición en los aspectos sistémicos de la autonomía y por lo tanto un problema para cualquier proyecto naturalizador. Pues, sin negar el rol importante de la evolución, está claro que fundamentar la mente en la vida (o la autonomía en general) requiere establecer los vínculos necesarios entre los fenómenos en ambos dominios. Lo que un organismo es y lo que hace no son procesos mutuamente externos. De no ser así el proyecto enactivo falla. En contraste con la solución autopoietica clásica, la solución que aporta Jonas es proponer que la transición a una nueva y sostenida forma de búsqueda de sentido y

propiedades del todo (Bennett y Hacker, 2003). Dado que uno de los dominios se establece por la presencia de una unidad en su totalidad en relación con su entorno, hay motivos lógicos para distinguir esas relaciones de los procesos constitutivos que conforman al sistema. Reducir fenómenos de un dominio a otro es ignorar que es posible tomar perspectivas alternativas cuando consideramos un sistema: la perspectiva de la unidad total en relación y la perspectiva de una multiplicidad de componentes que conforman la unidad. Confundir estas perspectivas nos llevaría a falacias como buscar la velocidad de un coche dentro de su motor (o describir nuestras memorias y experiencias como patrones de activación neuronal). Este es un punto importante y no debemos desecharlo. El daño, sin embargo, lo hace la noción de “no intersección”. Esta idea connota una separación estricta, cuando en realidad es posible mantener una relación de no-reducibilidad sin necesidad de aislar a los fenómenos entre dominios, los cuales pueden modularse y limitarse mutuamente. De esta forma, es de hecho posible que las explicaciones en el dominio A dependan de fenómenos en el dominio B, pero no de manera exclusiva o determinante; un motor potente condiciona las posibilidades para explicar por qué el coche es veloz, aun si no es posible deducir esto último de forma exclusiva a partir de la potencia del motor (ya que la velocidad es una relación entre coche y entorno). Lo que es importante y requiere cautela es la forma en que tales dependencias pueden tomar dado que cualquier relación entre los dominios será siempre una relación de modulación y no de determinación.

valor (como en el movimiento animal o en la creación de imágenes) regenera las mismas condiciones de organización que han hecho posible tal transición. O bien cambia la forma de generación de identidad que sostiene el nuevo dominio interactivo, o bien establece una nueva identidad autónoma (el animal que siente, el animal que percibe, etc.).

A pesar de las dificultades que hemos expuesto más arriba con la teoría de autopoiesis a secas, la existencia de otras formas de autonomía (en términos de dinámicas operacionalmente cerradas) es de hecho una posibilidad aceptada desde dentro de esta teoría. Se ha destacado la clausura operacional del sistema nervioso (Maturana y Varela, 1980, p. 127) y del sistema inmunológico (Varela, 1991). Varela también ha sugerido que sistemas autónomos también pueden darse en otros dominios, aun si no es en términos de producción material. Tal puede ser el caso de bucles senso-motores, conversaciones e interacciones sociales (Varela, 1979, 1991, 1997, ver también De Jaegher y Di Paolo, 2007). Sin embargo, lo que no se especifica es el tipo de relaciones que tales identidades pueden tener entre sí, cuáles serían, por ejemplo, las relaciones entre mi cuerpo biológico, mi identidad de género, mis habilidades y costumbres, mis roles sociales, etc. Las relaciones transformadoras entre los aspectos constructivos e interactivos de la autonomía que llevan a nuevas formas de identidad no pueden ser encaradas desde la teoría de autopoiesis. Esto simplemente se debe a la barrera lógica que se interpone entre los dos dominios y porque el énfasis en la conservación de la autopoiesis oscurece la posibilidad de un *devenir estructural* de nuevas formas de organización que abarquen tanto aspectos constructivos como interactivos de la vida.

Podemos ahora entender por qué las transiciones en mediación que propone Jonas (y muchas otras que él no hace explícitas) tienen un carácter irrevocable. Se tratan de auténticos nacimientos de nuevas formas de vida. Estas formas de vida pueden relacionarse con el substrato metabólico en una variedad de maneras, demandando la construcción de una genuina topología de procesos de mantenimiento de identidad (imbricados, anidados, en intersección, jerárquicos, compartidos, etc., de hecho la figura 2 permite imaginar varias posibilidades, basta con identificar bucles cerrados). Se abre también la posibilidad de que la dependencia de una forma de vida modifique la autonomía básica del metabolismo a tal punto que la identidad superior interviene de manera esencial en la condición de clausura operacional de la autopoiesis, haciendo que ésta dependa de hábitos y formas

de vida adquirida que se re-inscriben en la fisiología del organismo (Di Paolo, 2009b).

Como se puede apreciar el enactivismo consiste en una re-apropiación crítica y dialéctica de la idea de autopoiesis. Este paso no sólo permite al enactivismo fundamentar su proyecto de sentar las raíces de la cognición en procesos autónomos de mantenimiento de identidad, sino que abre posibilidades interesante para una teoría de la organización de la vida. Desde esta perspectiva, la relación entre los diferentes procesos auto-mantenidos que son sostenidos (pero no determinados) por un substrato autopoietico no es necesariamente una relación de armonía perfecta. Por el contrario, las tendencias regulativas inherentes en los procesos de mantenimiento de identidad a niveles más sofisticados van a entrar en ocasiones en conflicto con las valores metabólicos básicos. Se ha propuesto que los hábitos, en su plasticidad, deben ser considerados como estructuras autónomas que abarcan parcialmente al sistema nervioso, procesos fisiológicos y estructurales del cuerpo, regularidades del comportamiento y procesos en el entorno (Di Paolo, 2003, 2005, 2009b). Y como se sabe puede haber “malos” hábitos. El hecho es que como estructuras auto-mantenidas, los “malos” hábitos no son malos para sí mismos, sino para alguna otra identidad en la que el organismo participa (típicamente, en el caso humano, una combinación entre el yo metabólico y el socio-lingüístico).

Es en este punto en el que enactivismo puede aportar posibilidades de clarificación en los debates sobre mente extendida (Clark y Chalmers, 1993, Thompson y Stapleton, 2009, Wheeler, 2010). El uso de herramientas epistémicas indudablemente participa de procesos que sostienen el búsqueda de sentido. Pero la mera participación no implica constitución (Adams y Aizawa 2008). Aun así, quienes critican a la idea de la mente extendida en este punto han obviado proveer una noción operacional de constitución en la que se pueda basar su crítica. En la idea de clausura operacional el enactivismo encuentra tal noción.

La cognición, al ser inherentemente relacional no tiene ubicación espacial. Es absurdo intentar localizar la relación “la botella esta sobre la mesa” ya sea en la botella o en la mesa, o en algún punto indefinido entre ambas. Del mismo modo es absurdo localizar la apreciación de la ventaja de trepar un gradiente de azúcar en el interior de la bacteria. Por consiguiente, a pesar de que puede costar romper con esquemas metafóricos y lingüísticos, es

igualmente absurdo considerar que la mente está en la cabeza como que está fuera de ella. Sin embargo, sí es posible entender la noción de un sistema cognitivo que se extiende por sobre su substrato biológico en base a la idea de autonomía que hemos venido desarrollando. Las nuevas formas de vida tienden a conservar en forma modificada los niveles a partir de los cuales se originan. La imagen más clara es la de sistemas anidados cuyas identidades participan mutuamente en su conservación, pero que potencialmente pueden por momentos entrar en conflicto. La dependencia y las relaciones entre distintas formas de identidad (por ejemplo, entre el metabolismo y los hábitos que son posibles gracias a la agencialidad) puede resultar en un sistema integrado. Este sistema no es co-extensivo con el nivel básico de autonomía. Tal sistema *vida/mente* conserva su vida pero es diferente al organismo en sí y bien puede incorporar a su constitución relaciones de mediación, es decir interacciones con el entorno que recurren y también se auto-mantienen (Di Paolo, 2009b).

En cada caso particular si esta incorporación se da o no se da puede determinarse, en principio, operacionalmente. La noción de incorporación (asimilar al propio cuerpo) pertenece al espíritu de la idea de mente extendida, pero es la vez más general y más concreta. Más general porque la nueva identidad puede no ser simplemente una “extensión” de la identidad original (por ejemplo el metabolismo) la cual existía antes de la incorporación de una mediación auto-sostenida. Se trata simplemente de un sistema distinto — en algunos casos, como el de los insectos bajo el agua o el ciego con su bastón, la imagen de extensión es adecuada, en otros como el de mamífero que aprende el hábito de trepar árboles y navegar por la selva con pericia, la nueva forma de vida no es una extensión en el sentido dado en la literatura sobre mente extendida, sino que se trata de un sistema nuevo. La descripción enactiva es también más concreta porque nos permite distinguir de manera específica entre interacciones reguladas que pasan a constituir una nueva identidad de aquellas que no lo hacen, resolviendo así el problema de acople-constitución inherente en la noción de mente extendida. Esta distinción es posible en principio a través de la relación de clausura operacional bajo condiciones de precariedad, que debe siempre verificarse para poder hablar de la identidad de un sistema de cualquier orden, de manera no atribuida externamente.

Una transformación de orden cognitivo no necesita ser meramente una extensión espacial del cuerpo del agente (ni siquiera, metafóricamente, un

incremento con respecto a la situación original), basta con que sea un cambio cuyo efecto es el de conservar sus propias condiciones de posibilidad. Entender tales cambios operacionalmente puede ser un problema a nivel práctico. En los seres humanos, los hábitos pueden formar manojos organizados de forma compleja en redes cerradas que aún requieren una mejor elucidación teórica. En un cierto sentido es veraz afirmar que el agente que soy cuando estoy nadando en el mar es diferente al agente que soy cuando estoy tomando fotografías. En este punto, las relaciones e interacciones sociales intervienen en la estabilización de una identidad con distintas capacidades y habilidades y una cierta continuidad fenoménica y narrativa, en lugar de estabilizar la co-existencia de varias identidades regionales. De no ser así, en términos sistémicos, la continuidad de una misma identidad cognitiva a través e diferentes formas de agencialidad sería difícil de establecer, dado que no sería suficiente proveer los vínculos con una misma identidad metabólica subyacente dado que la misma, como hemos visto, puede involucrarse en varias redes cerradas a la vez. Es la acción de otras entidades cognitivas y el trasfondo social de significado que existe en las actividades humanas las que ayudan a sostener una identidad socio-lingüística asociada a un cuerpo con capacidades y habilidades múltiples. El yo cognitivo humano, por la tanto, esta literalmente co-determinado en interacción con el otro (De Jaegher y Di Paolo, 2007).

Enacción en acción

En las secciones anteriores hemos desarrollado el marco teórico central del enactivismo. Esta base permite la articulación conceptual necesaria para examinar varios de los puntos ciegos de los enfoques funcionalistas. Este trabajo no sólo ha generado debates interesantes, como en el caso entre la relación entre el enactivismo y la mente extendida (Di Paolo, 2009a, Thompson y Stapleton, 2009, Wheeler, 2010), sino que ha dado origen una variedad de propuestas positivas – no sólo críticas – con consecuencias teóricas y empíricas concretas. La lista que sigue no es exhaustiva y pretende simplemente mencionar sólo algunas de las referencias recientes a modo de orientación.

- El concepto enactivo de agencialidad (Barandiaran et al 2009) ha inspirado modelos computacionales de quemo-taxis basada en el metabolismo en bacterias (Egbert y Di Paolo 2009, Egbert,

Barandiaran y Di Paolo, 2010, 2012), verificando y unificando varias observaciones experimentales que pueden explicarse de considerarse la relación directa entre auto-constitución y regulación del comportamiento. Los modelos también proponen una nueva hipótesis sobre el papel comportamiento en la evolución de bacterias.

- Las nociones de autonomía y búsqueda de sentido se han aplicado a una nueva teoría de la cognición social, menos interesada en postular capacidades individuales para la inferencia de estados mentales en otros y más focalizada en los procesos de interacción corporizada, introduciendo la noción de búsqueda de sentido en participación (“participatory sense-making”) para describir la manera en que estados mentales, intenciones y acciones son en general co-construidas en participación (De Jaegher y Di Paolo 2007, De Jaegher 2009, Di Paolo, Rohde y Iizuka 2008, Froese y Di Paolo 2009, McGann y De Jaegher 2009; ver también De Jaegher, en este volumen). Recientemente se han desarrollado las consecuencias que este enfoque enactivo sobre la inter-subjetividad tiene para las neurociencias (De Jaegher, Di Paolo y Gallagher 2010, Di Paolo y De Jaegher, 2012, De Jaegher y Di Paolo, 2013), la normatividad socio-cultural (Froese y Di Paolo, 2011, Stewart y Steiner, 2009, De Jaegher, 2013b), la psicología del desarrollo (Wilby, 2011), la comprensión de textos (Caracciolo, 2011) y el uso de gestos durante interacciones (Cuffari, 2012) y la psiquiatría de enfoque fenomenológico (Fuchs y De Jaegher 2009, De Jaegher, 2013a).

- El foco en nociones de interioridad y experiencia subjetiva que permite el análisis sistémico del enactivismo ha generado propuestas empíricas en neurociencias. A través del estudio combinado de correlaciones espacio-temporales en varias bandas de frecuencias de activación neuronal y del uso riguroso de experiencia en primera persona se ha avanzado sobre el estudio de la estructura temporal fina de la percepción visual y el entendimiento de episodios epilépticos (Varela et al., 2001), el desarrollo de perspectivas no-representacionistas sobre las imágenes mentales (Thompson, 2007), el “yo” neuronal entendido como procesos de “auto-especificación” (Christoff et al. 2011) y un enfoque neuro-fenomenológico sobre la dinámica de la conciencia personal del tiempo (Varela, 1999).

- La operacionalización de la noción de autonomía y de la búsqueda de sentido en participación ha inaugurado una revisión de ideas sobre enfoques corporizados sobre las emociones (Colombetti, 2010) cuyas consecuencias éticas apenas han empezado a explorarse (Torrance y Colombetti, 2009).
- Los conceptos de autonomía y mediación ha generado una nueva visión sobre nociones como la de los hábitos (Di Paolo, 2005), la orientación de la acción hacia metas (McGann 2007) y las ideas de habilidad, pericia y modalidades de la percepción (McGann 2010).
- Muchas de las ideas enactivas se han explorado en primera instancia a través de modelos computacionales, notablemente siguiendo la metodología de robótica evolutiva (Harvey et al. 2005, Di Paolo, 1999, 2010, Di Paolo y Iizuka 2008, Rohde 2010) y robótica del desarrollo (Vernon, 2010).

Lo que estos ejemplos demuestran es que si bien el enactivismo es un paradigma en desarrollo, no se ha quedado en una etapa meramente crítica de las ideas tradicionales y sus defectos, sino que ha propuesto un marco conceptual positivo y coherente (aunque sin duda aún no completo) con consecuencias directas sobre una importante variedad de problemas e investigaciones en neurociencia, psicología, inteligencia artificial, robótica, y filosofía de la mente.

Conclusión

Las ideas que forman la base del enactivismo, pintadas a trazo grueso por Jonas y Varela y desarrolladas a través de enfoques sistémicos y fenomenológicos merecen sin duda una mayor exploración crítica y un desarrollo más profundo. Pero no cabe duda de que este enfoque intenta sentar unas nuevas bases claras y precisas para la resolución de problemas que han eludido a los enfoques tradicionales durante las últimas seis décadas. Podemos resumir estas ideas en sus consecuencias más radicales para las ciencias cognitivas:

1. La introducción de un concepto operacional de *identidad*, a partir del cual un agente cognitivo se considera como auto-constituido, manteniendo su organización en condiciones precarias que lo hacen

depender de manera profunda de su substrato material.

2. El desarrollo como consecuencia de esta forma de identidad de la noción del *búsqueda de sentido*, a partir de la cual se considera que un agente autónomo adaptivo se relaciona con su mundo con intereses y normas específicos, y una perspectiva propia basada en el acceso adaptivo a sus condiciones de viabilidad.
3. Definiciones claras y precisas de la *cognición*, del *comportamiento* y de la *agencialidad*
4. La relación dialéctica entre el agente autónomo y el tipo de *mediación* con su mundo a partir de la cual emergen nuevas formas de vida con creciente libertad y precariedad
5. La sugerencia de que tales formas de vida se re-inscriben en los modos de identidad que les dan origen, estableciéndose diversas relaciones “internas”, incluyendo conflictos y la posibilidad a ciertos niveles de complejidad de una estabilización de identidad a través del interacciones sociales sostenidas.

A modo de nota final, debemos recordar una vez más que el enactivismo es un programa en pleno desarrollo. Existen muchos puntos que requieren trabajo y una constante evaluación crítica. Si algo ha dejado en claro este enfoque es que el método más fructífero es el de la crítica dialéctica de conceptos como la autopoiesis, logrando superar sus limitaciones y conservando sus aspectos más fuertes. Tal crítica ha de aplicarse también a todas las nociones enactivas que hemos presentado y este trabajo no es solamente teórico sino que va acompañado de ramas empíricas, fenomenológicas y de conocimiento práctico. Sólo este método puede guiarnos hacia una creciente consciencia de nuestros propios puntos ciegos.

Bibliografía

- Adams F. y Aizawa, K., (2008). *The bounds of cognition*, Blackwell, Oxford.
- Ashby W. R., (1960). *Design for a brain: The origin of adaptive behaviour*, Chapman and Hall, London
- Barandiaran, X., Rohde, M. y Di Paolo, E. A., (2009). Defining agency: individuality, normativity, asymmetry and spatio-temporality in action. *Adaptive Behavior*, 17: 367-386
- Bennett, M. R., and Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*, Wiley-Blackwell, Oxford.
- Boden M. A., (1999). Is metabolism necessary? *British Journal of Philosophy of Science* 50:231– 248
- Bourgine P. y Stewart J., (2004). Autopoiesis and cognition. *Artificial Life*, 10:327–346
- Caracciolo, M., (2012). Narrative, meaning, interpretation: an enactivist approach, *Phenomenology and the Cognitive Sciences*, 11, 367-384.
- Christoff K., Cosmelli D., Legrand D., y Thompson E., (2011). Specifying the self for cognitive neuroscience. *Trends in Cognitive Science*, 15(3). 104-112
- Clark A. y Chalmers D., (1998). The extended mind. *Analysis* 58(1): 7–19.
- Colombetti, G., (2010). Enaction, sense-making and emotion, in Stewart J, Gapenne O, Di Paolo E (eds) *Enaction: Toward a New Paradigm for Cognitive Science*, Cambridge MA: MIT Press, pp 145-164.
- Colombetti, G. y Torrance, S., (2009). Emotion and ethics: an inter-(en)active approach, *Phenomenology and the Cognitive Sciences*, 8(4), 505-526.
- Cuffari, E., (2012). Gestural sense-making: hand gestures as intersubjective linguistic enactments, *Phenomenology and the Cognitive Sciences*, 11, 599 – 622.
- De Jaegher, H., (2009). Social understanding through direct perception? Yes, by interacting. *Consciousness and Cognition* 18 (2): 535–542.
- De Jaegher H (2013a) Embodiment and sense-making in autism. *Front. Integr. Neurosci.* 7:15. doi: 10.3389/fnint.2013.00015
- De Jaegher, H. (2013b) Rigid and fluid interactions with institutions. *Cognitive Systems Research*, in press, <http://dx.doi.org/10.1016/j.cogsys.2013.03.002>
- De Jaegher, H., y Di Paolo, E. A., (2007). Participatory sense-making: An enactive approach to social cognition, *Phenomenology and the Cognitive Sciences*, 6 (4): 485–507.
- De Jaegher, H., y Di Paolo E. A., (2013). Enactivism is not interactionism. *Front. Hum. Neurosci.* 6:345. doi: 10.3389/fnhum.2012.00345.

- De Jaegher, H., Di Paolo, E. A. y Gallagher, S., (2010). Can social interaction constitute social cognition?, *Trends in Cognitive Sciences*, 14(10), 441 – 447.
- Di Paolo E. A., (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4:97–125.
- Di Paolo, E. A., (2009a). Extended life, *Topoi* 28:9-21.
- Di Paolo, E. A. (2009b) Overcoming autopoiesis: a enactive detour on the way from life to society. In Magalhaes, R., and Sanchez, R. (Eds) *Autopoiesis in Organizations and Information Systems*, Elsevier, pp. 43 – 68.
- Di Paolo, E. A., (2010). Robotics inspired in the organism, *Intellectica*, 53-54: 129 – 162.
- Di Paolo, E. A. y De Jaegher, H., (2012). The Interactive Brain Hypothesis, *Frontiers in Human Neuroscience*, 6:163. doi:10.3389/fnhum.2012.00163.
- Di Paolo E. A., Rohde M., y De Jaegher H., (2010). Horizons for the enactive mind: values, social interaction, and play. In: Stewart J, Gapenne O., Di Paolo E. (eds) *Enaction: towards a new paradigm for cognitive science*. MIT Press, Cambridge, MA.
- Di Paolo, E. A., e Iizuka, H., (2008). How (not) to model autonomous behaviour. *BioSystems* 91:409-423.
- Di Paolo, E. A., Rohde, M., e Iizuka, H., (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing, *New Ideas in Psychology*, 26 (2): 278–294.
- Egbert, M. y Di Paolo, E. A., (2009). Integrating behavior and autopoiesis: An exploration in computational chemo-ethology. *Adaptive Behavior*, 17: 387-401.
- Egbert, M., Barandiaran, X. y Di Paolo, E. A., (2010). A minimal model of metabolism-based chemotaxis', *PLoS Computational Biology*, 6(12): e1001004.
- Egbert, M. D., Barandiaran, X. E., y Di Paolo, E. A., (2012). Behavioral metabolution: The adaptive and evolutionary potential of metabolism-based chemotaxis. *Artificial Life*, 18(1), 1-25. doi:10.1162/artl_a_00047.
- Fleischaker G. R., (1988). Autopoiesis: the status of its system logic, *BioSystems* 22(1):37–49
- Froese, T. y Di Paolo, E. A., (2009). Sociality and the life-mind continuity thesis, *Phenomenology and the Cognitive Sciences*, 8(4), 439-463.
- Froese, T. y Di Paolo, E. A., (2011). The enactive approach: Theoretical sketches from cell to society, *Pragmatics and Cognition*, 19, 1-36.

- Fuchs, T. y De Jaegher, H., (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation, *Phenomenology and the Cognitive Sciences*, 8(4), 465-486.
- Harvey, I., Di Paolo, E. A., Tuci, E., Wood, R., y Quinn, M., (2005). Evolutionary robotics: A new scientific tool for studying cognition. *Artificial Life*, 11(1/2), pp. 79 – 98.
- Jonas H., (1966). *The phenomenon of life: towards a philosophical biology*. Northwestern University Press, Evanston.
- McGann, M., (2007). Enactive theorists do it on purpose, *Phenomenology and the Cognitive Sciences*, 6(4), 463-483.
- McGann, M., (2010). Perceptual modalities: Modes of presentation or modes of action?, *Journal of Consciousness Studies*. 17: 72-94.
- McGann, M., y De Jaegher, H., (2009). Self-other contingencies: Enacting social perception. *Phenomenology and the Cognitive Sciences* 8 (4): 417–437.
- Maturana H., (2002). Autopoiesis, structural coupling and cognition: a history of these and other notions in the biology of cognition, *Cybernetics and Human Knowing*, 9:5–34.
- Maturana H. y Varela F. J., (1980). *Autopoiesis and cognition: the realization of the living*. D. Reidel Publishing, Dordrecht.
- Moreno A. y Ruiz-Mirazo K., (1999). Metabolism and the problem of its universalization. *BioSystems* 49:45–61.
- Rohde, M. (2009), *Enaction, embodiment, evolutionary robotics. Simulation models for a post-cognitivist science of mind*. Amsterdam and Paris: Atlantis Press.
- Stewart J., (1995). Cognition = life: implications for higher-level cognition, *Behavioral Processes*, 35(1–3): 311–326.
- Thompson E., (2007). *Mind in life: biology, phenomenology and the sciences of mind*. Harvard University Press, Cambridge.
- Thompson, E. y Stapleton, M., (2009). Making sense of sense-making: Reflections on enactive and extended mind theories, *Topoi*, 28: 23-30.
- Thorpe W. H., (1950). Plastron respiration in aquatic insects. *Biological Review* 25:344–390.
- Turner, J. S., (2000). *The extended organism: The physiology of animal-built structures*, Cambridge, Mass.: Harvard University Press.
- Varela F. J., (1979). *Principles of biological autonomy*. Elsevier, New York.

- Varela F. J., (1991). Organism: a meshwork of selfless selves. In: Tauber AI (ed) *Organism and the origin of the self*. Kluwer Academic Publishers, The Netherlands, pp 79–107
- Varela F. J., (1997). Patterns of life: intertwining identity and cognition. *Brain and Cognition*, 34:72–87.
- Varela, F. J. (1999), 'The specious present: A neurophenomenology of time consciousness', in J. Petitot, F. J. Varela, B. Pachoud and J.-M. Roy (eds.), *Naturalizing Phenomenology*. Stanford, CA: Stanford University Press, pp. 266–314.
- Varela F. J., Thompson E., y Rosch E., (1991). *The embodied mind: cognitive science and human experience*. MIT Press, Cambridge, MA.
- Varela, F. J., Lachaux, J.-P., Rodriguez, E., y Matinerie, J., (2001). The brainweb: phase synchronization and large-scale integration, *Nature Reviews. Neuroscience* 2: 229–230.
- Vernon, D., (2010). Enaction as a conceptual framework for developmental cognitive robotics, *Paladyn Journal of Behavioral Robotics*, 1(2): 89–98.
- Weber A. y Varela F. J., (2002). Life after Kant: natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences* 1:97–125.
- Wheeler, M., (2010). 'Minds, things, and materiality', in Renfrew C. and Malafouris L. (eds.), *The cognitive life of things: Recasting the boundaries of the mind*, Cambridge: McDonald Institute for Archaeological Research Publications.
- Wilby, M., (2012). Embodying false belief tasks, *Phenomenology and the Cognitive Sciences*, 11, 519–540.