



Universidad Nacional Autónoma de México

Programa de Maestría en Filosofía

Facultad de Filosofía y Letras

Instituto de Investigaciones Filosóficas

*Mente extendida y agencia:
la interacción cerebro, cuerpo y mundo*

Tesis

que para optar por el grado de
Maestro en Filosofía

Presenta:

Franklenin Sierra Casiano

Tutor: Dr. Sergio Martínez Muñoz
Instituto de Investigaciones Filosóficas

Ciudad de México, Septiembre 2015.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

*Para Mary,
por recordarme siempre
lo realmente importante de la vida*

Agradecimientos

Este proyecto de investigación inició en el *Instituto de Investigaciones Filosóficas* y fue completado en el *Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas*, ambos pertenecientes a la UNAM. Por ello, me permito agradecer a mi asesor de tesis el Dr. Sergio Martínez Muñoz por toda su invaluable ayuda en estos últimos tres años de mi vida académica, pero sobre todo porque su seminario “*Racionalidad, razonamiento y cognición*” fue mi puerta de entrada al mundo de las Ciencias cognitivas.

Mi más sincero agradecimiento al Dr. Tom Froese por toda su ayuda; sus comentarios, sugerencias y críticas, hicieron de mi tesis una mejor versión. Además, su curso del semestre 2015-2 “*Autonomous agents, social systems, and the new cognitive science*” fue pieza clave en la culminación de mi proyecto de investigación. No puedo dejar de mencionar al Dr. Maximiliano Martínez, al Dr. Miguel Ángel Sebastian y al Dr. Jorge Reyes, por haberse tomado el tiempo de leer y examinar mi tesis. Gracias a sus comentarios, y su ayuda desinteresada, he podido completar este proyecto de investigación.

La tesis que aquí presento no hubiera sido posible sin el apoyo del CONACYT (Consejo Nacional de Ciencia y Tecnología) y el PAEP (Programa de Apoyo a los Estudios de Posgrado). Gracias a su apoyo pude dedicarme de lleno en esta investigación y además me permitió presentar los primeros avances de mi tesis en el *XVII Congreso Internacional de Filosofía* de la AFM, en la ciudad de Morelia, Michoacan. Así como en el *Tercer Congreso Americano de Filosofía Analítica* y el *Tercer Congreso de la Sociedad Brasileña para la Filosofía Analítica*, en Fortaleza Ceará, Brasil.

Finalmente, no puedo dejar de agradecer la ayuda de mis colegas y amigos Jimena Clavel, Jesús Javier Godoy, Henry David Pinto, Julia Gonzalez, Abraham Sapien, Carlos Prieto y Edgar René Ruíz, por haber leído las primeras versiones de los capítulos. Pero sobre todo, por las tantas charlas de café en donde una y otra vez tuvieron que escuchar la exposición de gran parte de los temas que en este trabajo se presentan.

Coyoacán, Ciudad de México.
Agosto 2015.

Índice

Introducción	1
I. Sobre la idea de conceptos unificados	1
II. El problema de la agencia	4
III. La tesis de la mente extendida como marco de partida	5
IV. Objetivos de este proyecto de investigación	6
V. Estructura de esta tesis.....	7
Capítulo 1. Breve recorrido histórico	9
1.1 La relación mente y mundo.....	9
1.2 El mito de Descartes	9
1.3 El cuerpo y el mundo	10
1.4 El solipsismo metodológico.....	12
1.5 Ciencia cognitiva cartesiana	12
1.6 Las 4Es de la mente y la cognición	13
1.7 Un nuevo planteamiento de la agencia	15
Capítulo 2. La tesis de la mente extendida	17
Introducción.....	17
2.1 Cognición extendida: ¿el iPhone es parte de mi mente?.....	18
2.1.1 Externismo activo vs externismo semántico.....	18
2.1.2 Postulados de la tesis de la mente extendida	20
2.2 Argumentos a favor de la mente extendida	20
2.2.1 Jugando Tetris	21
2.2.2 La memoria de Otto	22
2.2.3 El principio de paridad.....	23
2.2.4 Conclusiones de los argumentos.....	24
2.3 Defendiendo los límites de la cognición.....	25
2.3.1 Críticas al principio de paridad.....	26
2.3.2. La marca de lo cognitivo	27
2.3.3 La falacia constitución-acoplamiento	29
2.4 Conclusiones del capítulo.....	30
Capítulo 3. Tres respuestas a la <i>falacia constitución-acoplamiento</i>	32
Introducción.....	32
3.1 Cognición integrada	33
3.1.1 La cognición como un sistema biológico	34
3.1.2 Externismo dinámico	35

3.1.3 El principio de complementaridad	37
3.1.4 La tesis de la manipulación.....	38
3.1.5 Respuesta a la falacia constitución-acoplamiento	39
3.2. Sistemas cognitivos-fenomenológicos extendidos	43
3.2.1 Como trabajar sin representaciones.....	44
3.2.2 Argumento para los sistemas dinámicos	45
3.2.3 Respuesta a la falacia constitución-acoplamiento	46
3.3 Enactivismo: cognición sin representaciones.....	48
3.3.1 El enactivismo radical: cognición sin contenido mental.....	49
3.3.2 Mentes extensivas: respuesta a la falacia constitución-acoplamiento	50
3.4 Conclusiones del capítulo	51
Capítulo 4. El problema de la agencia: ¿cómo diseñar <i>fungus eaters</i>?	53
Introducción.....	53
4.1 El problema metodológico: usos, distinciones y aclaraciones de la agencia.....	56
4.1.1 La agencia en la filosofía.....	56
4.1.2 La agencia en las neurociencias.....	59
4.1.3 La agencia en la inteligencia artificial, la robótica y la vida artificial	60
4.2 Un esquema tripartito de la agencia	63
4.2.1 Acciones intencionales y emociones	63
4.2.2 Agencia humana	65
4.2.3 Agencia animal	66
4.2.4 Agencia artificial	66
4.3 Aplicaciones del marco tripartito	66
4.3.1 Compatibilidad de los diferentes enfoques de la agencia	66
4.3.2 Respuesta al problema normativo de la agencia	67
4.3.3 Construcción de robots	68
4.3.4 Futuras investigaciones	69
4.4 El problema externista de la agencia.....	70
4.4.1 Argumento internista antropomórfico	71
4.4.2 Argumento externista material.....	73
4.4.2.1 Intencionalidad en la acción	75
4.5 Crítica al argumento internista	79
4.6 Implicaciones del argumento externista material	80
Conclusión.....	82
Bibliografía.....	85

Mentes, cerebros y ciencia

“La palabra ‘ciencia’ tiende a sugerir un numeroso grupo de investigadores en bata blanca agitando tubos de ensayo y mirando de cerca los instrumentos. Para muchas mentes esto sugiere una infalibilidad arcana. El cuadro rival que quiero sugerir es éste: a lo que aspiramos en las disciplinas intelectuales es al conocimiento y a la comprensión. Ya lo tengamos en matemáticas, crítica literaria, historia, física o filosofía, hay solamente conocimiento y comprensión. Algunas disciplinas son más sistemáticas que otras, y podríamos querer reservar la palabra ‘ciencia’ para ellas” (1984, p. 12).

John Searle

Introducción

*¿Qué es la vida? ¿Qué es la mente? Es más probable que estas preguntas sean preguntadas por un filósofo más que por un biólogo o científico cognitivo, incluso a pesar del hecho de que estos fenómenos naturales supuestamente constituyen los respectivos campos de investigación de los científicos. Pero si nosotros los científicos no intentamos de alguna manera responder estas preguntas definicionales, ¿entonces cómo podemos decidir qué investigar?*¹ (2010, p. 7).

Tom Froese y John Stewart - *Life After Ashby*

I. Sobre la idea de conceptos unificados

Las ciencias cognitivas nacieron como un esfuerzo unificado de diferentes campos de estudio para dar cuenta de las interrogantes que surgen en torno a la mente, el cerebro y la cognición. Ha sido este esfuerzo conjunto de la antropología, la psicología, la lingüística, la filosofía, la inteligencia artificial y las neurociencias lo que ha permitido que a pesar de ser un campo muy joven (un poco más de cincuenta años) haya dado ya grandes pasos en la tarea de responder a estos retos. Como bien escribiera Gardner en su clásico libro *The Mind's New Science*: “[u]no podría decir que las ciencias cognitivas tienen un largo pasado pero una corta historia”² (1985, p. 9). En este corto tiempo ha logrado implementar tres estrategias metodológicas (el *cognitivismo*, el *conexionismo* y los *sistemas dinámicos*) que tienen diversas maneras de ver la mente y entender la cognición.

En la actualidad estos tres enfoques co-existen y se combinan de diversas maneras (Thompson 2007). A raíz de esto y además porque las ciencias cognitivas son un campo muy joven y multidisciplinario (véase la *Figura 1*), una de las principales dificultades que surgen cuando trabajamos en las ciencias cognitivas es la vaguedad con la que los conceptos son utilizados. Términos como consciencia, racionalidad, conocimiento, inteligencia, experiencia, vida, mente y agencia, son tan solo algunos ejemplos de conceptos que varían drásticamente dentro de los campos de las ciencias cognitivas. Si a estos campos agregamos la arqueología, la robótica, la psiquiatría y

¹ “What is life? What is mind? These questions are more likely to be asked by a philosopher than a biologist or cognitive scientist, even despite the fact that these natural phenomena supposedly constitute the scientists’ respective fields of inquiry. But if we scientists do not try to address these definitional questions in some way, then how do we decide what to investigate?” (Froese y Stewart 2010, p. 7).

² “One might say that cognitive science has a very long past but a relatively short history” (Gardner 1985, p. 9).

otras áreas de la filosofía —que van más allá de la filosofía de la mente— como lo son la fenomenología y la epistemología, de pronto nos vemos inmersos en un mar de literatura que en algunos casos nos confunde y en algunos otros nos hace pensar si realmente estamos hablando de los mismos fenómenos. La racionalidad, por ejemplo, es un término de un uso tan común pero que tiene diferentes acepciones dentro de la psicología, la teoría de juegos, la inteligencia artificial y la biología evolutiva³.

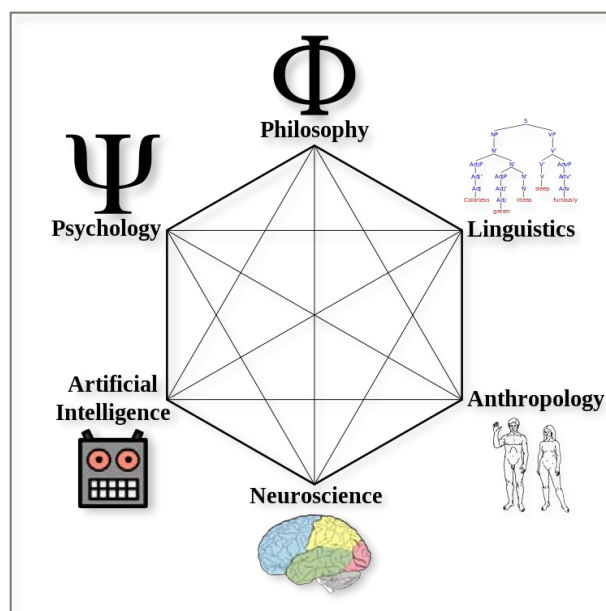


Figura 1. El hexágono cognitivo de Gardner (1985), aquí podemos observar los campos bajo los cuales nacieron las ciencias cognitivas.⁴

Esta preocupación que parece ser un mero enredo conceptual involucra diferentes tesis metafísicas y ontológicas que impactan directamente en los proyectos de investigación. No se trata, por ejemplo, simplemente de tener un concepto unificado de procesos mentales, sino que la manera de entender dichos procesos implica el uso de planteamientos precisos en la investigación llevada a cabo en las neurociencias. Por ejemplo, si partimos de hipótesis internistas acerca de la cognición entonces nuestra investigación empírica estará solo enfocada en lo que sucede dentro de la cabeza de los

³ Mientras que de manera tradicional se aceptaba la idea de que razonamos de acuerdo a los principios lógicos (es decir mediante principios formales y sintácticos), el trabajo de [Gigerenzer \(1998, 2005\)](#) demostró que nuestra toma de decisiones se lleva mediante *heurísticas*. Es decir, mediante decisiones que nos dan mejores resultados pero que en muchos casos contradicen los criterios normativos impuestos por la lógica.

⁴ Imagen tomada de la Enciclopedia Wikipedia.

sujetos, mientras que si nuestro punto de partida es un enfoque externista, entonces estudiaremos no solo lo que sucede dentro de nuestras cabezas sino también la manera en que las herramientas, la tecnología y el medio ambiente influyen en la cognición. Por su puesto que gran parte del problema es que contamos con muchos argumentos a favor pero también con muchos en contra.

Estos enredos conceptuales conllevan un sin fin de consecuencias empíricas que impactan (de manera negativa) en la creación de un marco sólido bajo el cual establecer proyectos de investigación. Uno de los ejemplos más citados para mostrar cómo planteamientos erróneos pueden llevarnos a investigaciones empíricas erróneas es el proyecto CYC⁵ (el cual partió con recursos económicos de más de 50 millones de dólares) iniciado en el año de 1984 por Douglas Lenat y Edward Feigenbaum.

El objetivo de dicho proyecto es hacer de una computadora una base de todos con todo el conocimiento del mundo, la idea de fondo es que una vez que completemos esta base las computadoras podrán *pensar* como lo hacemos los humanos. Así, la hipótesis es que si tenemos todo el conocimiento del mundo (computado en una base de datos) de algún modo la inteligencia emergerá. Por ello, proveer a ésta computadora con conocimiento científico no basta; se tienen que meter también proposiciones de uso cotidiano del tipo “si alguien llora es porque está triste”, si “alguien está triste entonces necesita un pañuelo”. Pero, ¿cómo podría emerger de esta base de datos la inteligencia? (Véase [Dreyfus 1979](#)) o ¿cómo podría emerger la semántica del conocimiento meramente sintáctico? (Véase [Searle 1984](#)). Se pueden hacer un sinfín de severas críticas a este tipo de proyectos, pero el único punto que quiero mostrar es cómo la idea que tengamos de inteligencia (o de racionalidad) impacta de manera contundente en la manera en que desarrollamos investigaciones empíricas.

Uno de los muchos pasos que se tienen que dar para el avance del conocimiento en las ciencias cognitivas es el esclarecimiento de muchas de las nociones que utilizamos, este esclarecimiento tiene que partir de investigaciones que se centren en conceptos centrales y de uso cotidiano que, a pesar de ello, despliegan un sinfín de problemas filosóficos. Ejemplos de ellos son vida, inteligencia, mundo, cuerpo y conocimiento. El tema de tesis estará abocado a la noción de agencia, la cual definiremos en términos generales como la capacidad de actuar ([Schlosser 2015](#)).

⁵ Este proyecto tomó el nombre de CYC por la palabra en inglés para enciclopedia (*encyclopedia*); para mayor información véase [Clark \(1997\)](#) y la entrada “CYC” de Wikipedia.

II. El problema de la agencia

La psicóloga Barrett presenta —en *Beyond the Brain: How Body and Environment Shape Animal and Human Minds* (2001)— el caso de Santino, un gorila que causó sensación en la prensa por llevar cabo tareas que sugerían que tenía nociones acerca del futuro (una característica que usualmente es atribuida solo a los animales humanos). El caso de Santino causó sorpresa porque usualmente nos concebimos a nosotros mismos como los únicos entes con capacidad de actuar libremente en el mundo y también porque somos conscientes de esa capacidad que tenemos para actuar. Santino, por ejemplo, al preocuparse por el futuro parecía llevar a cabo no solo acciones sino tener consciencia de por qué realizaba dicha acciones.

Al igual que Santino, los humanos somos autores de ciertos eventos en el mundo a los cuales hemos llamado acciones y estas han sido encasilladas como procesos o eventos de la agencia (Ross 2007). Dentro de esta capacidad que tenemos para actuar podemos empezar a notar ciertos términos que resultan útiles al tratar de explicar los eventos de la agencia: libertad, acción, intenciones. Pero también, surgen un sin fin de preguntas de diferente indole; por ejemplo, podemos trazar una pregunta fenomenológica: ¿qué significa experimentar una acción cómo mi acción?

Además, tenemos las preguntas ontológicas y normativas: ¿las acciones por sí mismas cuentan como un criterio para escribir agencia o necesitamos tener *acciones intencionales*?, ¿es la agencia una propiedad enteramente humana o podemos compartir este rasgo con los animales? Si esto es así, ¿podemos modelar la agencia y utilizarla en el diseño de robots completamente autónomos? Aunado a esto, Ross (2007) pregunta: ¿la agencia envuelve racionalidad, y si sí, de qué tipo y hasta donde llega su límite?⁶ (p . 1). En suma, todas estas preguntas pueden resumirse en una sola pregunta: ¿qué tipo de agentes somos?

Al sumergirnos dentro del problema de la agencia, nos encontramos en un vasto océano en donde notamos que diferentes proyectos hablan de la agencia pero sin especificar bien de qué están hablando; el error en éstas hipótesis es metodológico: parten de la idea de que la agencia no es término controversial. Sin embargo, basta un análisis rápido para darnos cuenta de que es un término que a pesar de que goza un parecido de familia (que algunas veces es asociado a la intencionalidad, a veces a la consciencia o la autonomía) su uso varía enormemente de un campo a otro.

⁶ “Does agency essentially involve rationality, and if so, of what kind and beyond what threshold?” (Ross 2007, p . 1).

III. La tesis de la mente extendida como marco de partida

Por sí solo el tema de la agencia constituye un tema a desarrollar dentro de muchas ramas de la filosofía, pero debido a que no podemos responder a las diferentes interrogantes que surgen en torno a ella sin apelar a la cognición humana, debemos partir de un marco que nos permita entender los procesos cognitivos y nos provean una idea general de lo que entendemos por fenómenos mentales. Por este motivo, utilizaré la tesis de la mente extendida (Clark y Chalmers 1998), pues dicha propuesta rompe con la idea tradicional de que los estados mentales son un fenómeno totalmente interno que se lleva a cabo dentro de la cabeza de los sujetos. Si hay casos que apoyan la tesis de que la mente se extiende al entorno, entonces deberíamos de poner en consideración muchas de nuestras suposiciones habituales y analizar qué consecuencias tiene el hecho de que muchos de nuestros procesos cognitivos pueden ser producidos por elementos que están más allá de nuestro cerebro.

La razón por la cual tomo este tipo de externismo como punto de partida es la notable influencia que dicha tesis ha hecho en diferentes campos de estudio; pero además, porque dicha propuesta aún no ha sido llevada completamente a la epistemología⁷. Aún falta mucha investigación para re-pensar críticamente muchas nociones epistemológicas con las cuales trabajamos:

La tesis de la cognición extendida ha sido influyente en la filosofía de la mente, las ciencias cognitivas, la lingüística, la informática y la ética, pero sorprendentemente no en la epistemología contemporánea. La disciplina preocupada con uno de los más destacables productos de la cognición —el conocimiento— ha ignorado ampliamente la sugerencia de que su principal objeto de estudio puede ser producido por un proceso cognitivo fuera de la piel humana⁸ (Clark, Pritchard y Vaesen 2012, p. 87).

⁷ Antes de Clark y Chalmers (1998) hubo dos tipos de externismo semántico en la filosofía: Putnam (1975) y Burge (1979). Sin embargo, el externismo de Clark y Chalmers es diferente porque defienden una tesis externista constitutiva: que los elementos externos que desempeña un rol activo en la cognición son parte de la mente. Véase la sección 2.1.1 de esta tesis.

⁸ “The thesis of extended cognition has been influential in the philosophy of mind, cognitive science, linguistics, informatics, and ethics, but, surprisingly, not in contemporary epistemology. The discipline concerned with one of the most remarkable products of human cognition – viz., knowledge – has largely ignored the suggestion that its main object of study might be produced by cognitive processes outside the human skin” (Clark, Pritchard y Vaesen 2012, p. 87).

A pesar de que la tesis de Clark y Chalmers ha tenido una enorme influencia, también ha recibido severas críticas de parte de los internistas; por este motivo, es necesario hacer una apología y mostrar por qué los casos de cognición extendida tienen que ser tomados en serio. Aunque tomaré como punto de partida la tesis tradicional de la mente extendida (Clark y Chalmers 1998) este enfoque será el pretexto para abrir el debate entre el internismo y el externismo, pero una vez que tenga que responder a las críticas internistas tomaré partido por el externismo anti-representacionista de Chemero (2009).

En el transcurso de esta tesis quedará patente por qué la tesis de Clark y Chalmers, a pesar de que es una tesis revolucionaria, sigue manteniendo algunos supuestos tradicionales como es el caso de las representaciones. Por este motivo, la tesis tradicional de la mente extendida me servirá como una escalera para llegar al debate sobre internismo y externismo. Pero una vez que estemos dentro del debate abandonaré esa escalera para optar por una postura más radical que defiende no que algunas veces tenemos casos de cognición extendida sino que la mente está siempre extendida hacia el mundo (Chemero 2009).

Resumiendo, el presente proyecto de investigación parte de un condicional simple: si nos tomamos en serio los casos de cognición extendida, ¿cómo debemos entender la agencia?

IV. Objetivos de este proyecto de investigación

Dado que las ciencias cognitivas se encuentran todavía en una etapa joven sin saber bien no sólo qué fenómenos estamos estudiando, sino también cómo deberíamos de estudiarlos, centrarse en la agencia es apenas uno de los muchos pasos para lograr dar cuenta de diferentes de nuestros conceptos que a raíz de las nuevas concepciones de la mente tienen que ser analizados. La investigación en torno a la agencia no se agota sólo en buscar una definición teórica, sino que dicha definición tiene que ser capaz de llevarse al terreno empírico de diferentes disciplinas. Algunas de las ventajas de la agencia y la razón por la cual constituye un término tan fundamental es su amplio rango de alcance⁹:

⁹ Las implicaciones de la agencia incluso llegan al terreno de la ética, sin embargo debido a las limitaciones de esta tesis me centraré en aspectos ontológicos y epistemológicos de la agencia sin ahondar, por ejemplo, en las implicaciones que tiene la idea de agencia en el sistema judicial.

Aunque un insecto robótico parece ser ya un ejemplo mínimo de agencia, el concepto es lo suficientemente amplio para abarcar humanos e incluso organizaciones colectivas. Desde el punto de vista de la agencia, es posible concebir un programa de investigación que proceda desde abajo, del comportamiento encarnado más sencillo, dándole base a fenómenos de nivel superior en formas gradualmente más complejas de interacciones situadas y de sus mecanismos subyacentes¹⁰ (Barandiaran, Di Paolo y Rohde 2009, p 1).

A raíz de la importancia que despliega la agencia, los objetivos de este proyecto de investigación son dar un reporte de la agencia mirada desde una postura externista de la cognición y proveer un esquema de la agencia que pueda ser útil no solo en la filosofía sino también en las diferentes ramas de las ciencias cognitivas.

Cumplir este propósito implicará sobre todo mostrar por qué la agencia es un problema (esto nos llevará a analizar cómo es utilizada la noción de agencia en diferentes campos de estudio como lo son la psiquiatría, las neurociencias, la arqueología, la robótica y la vida artificial, etc.). Presentar una clasificación de cómo debería de entenderse la agencia para que pueda ser útil en las diferentes áreas de las ciencias cognitivas y finalmente defender que la agencia es un proceso emergente que surge cuando interactúan el cuerpo, el cerebro y el mundo.

Así, lo que defenderé es que si nos tomamos en serio los casos de cognición y mente extendida, entonces la agencia tiene que ser entendida como un proceso extendido que surge solo cuando interactúan el cerebro, el cuerpo y el mundo.

V. Estructura de esta tesis

Mi primer paso en esta investigación será dar un breve recorrido histórico para entender en dónde se sitúan las diferentes concepciones de la mente que emergieron en el siglo XX, esto con el fin de ver de dónde surgen, qué rechazan y cuáles son sus principales tesis (**capítulo 1**). Una vez hecho esto y ya teniendo una idea global de las diferentes corrientes de la cognición que emergieron sobre todo en la última década del siglo pasado, presentaré la tesis de la mente extendida tal y como fue

¹⁰ “While an insect-like robot already seems to be a minimal instance of agency, the concept is open enough to also cover humans or even collective organizations. From the departure point of agency it is possible to envision a research program that proceeds from the bottom up, from the simplest embodied behavior, grounding higher level phenomena on increasingly complex forms of situated interactions and their underlying mechanisms” (Barandiaran, Di Paolo y Rohde 2009, p 1).

presentada por los filósofos Clark y Chalmers en el año de 1998. Aquí analizaré las nuevas ideas que proponen y por qué, como bien señala [Theiner \(2011\)](#), su propuesta constituye una revolución científica; además, también presentaré las objeciones a dicha tesis (**capítulo 2**).

Mirar a la cognición como un proceso extendido hacia el mundo nos brinda un nuevo marco para analizar nuestros conceptos que estaban sobre todo basados en tesis internistas; sin embargo, antes de entrar a analizar las implicaciones que estas ideas tienen sobre la agencia, el siguiente apartado estará enteramente dedicado a dar respuesta a la principal crítica en contra de la idea de que la cognición se extiende al entorno: la *falacia constitución-acoplamiento*. Por ello, en esta parte presentaré tres respuestas ([Menary 2007](#); [Chemero 2009](#); [Hutto y Myin 2013](#)) que podemos dar a dicha falacia. Una vez que hayamos trazado una respuesta a ésta crítica podemos tomarnos en serio la tesis de que hay casos de cognición extendida. Aunque tomo partido por [Chemero \(2009\)](#), presento tres respuestas porque cada una de ellas tiene diferentes supuestos de la cognición; con ello quiero mostrar que para dar respuesta a la crítica de los internistas podemos hacerlo desde posturas representacionistas, anti-representacionistas o enactivas (**capítulo 3**).

Finalmente, una vez que el terreno está preparado podremos adentrarnos al problema de la agencia (**capítulo 4**). Dado que este proyecto de investigación se centra totalmente en la agencia (pero no solo en la filosofía sino en los diferentes campos de las ciencias cognitivas) esto implicará analizar primeramente la manera en que dicho término es usado en la inteligencia artificial, las neurociencias, la robótica y la filosofía. Posteriormente, en la parte final de este capítulo, presentaré un par de argumentos que partiendo de casos de casos de cognición extendida, llegan a conclusiones diferentes: i) que la agencia es una propiedad enteramente humana, razón por lo cual no puede ser ni distribuida ni extendida. ii) Que la agencia es un fenómeno emergente que surge solo cuando interactúan en el cerebro, el cuerpo y el mundo.

Finalmente, daré un recuento de la ruta argumentativa que he seguido en esta tesis y de sus implicaciones para nuestra noción de agencia (**Conclusión**).

Capítulo 1

Breve recorrido histórico

¿Dónde termina la mente y dónde empieza el resto del mundo?¹¹ (1998, p. 7).

Clark y Chalmers - *The Extended Mind*

1.1 La relación mente y mundo

Durante más de veinte siglos el estudio de la mente partió bajo el presupuesto de que podemos estudiar al hombre como algo separado del mundo, acorde con [Dreyfus \(1991\)](#) la idea de que podemos estudiar al universo como algo separado del hombre la podemos rastrear hasta Platón; de este modo, si el universo puede ser estudiado como algo separado, por simple deducción lógica, también podemos estudiar al hombre como separado del mundo. Con esta simple suposición se obvió un paso crucial en el estudio de la relación entre la mente y mundo, de tal modo que aunque se estudió esta relación de diferentes maneras, se mantuvo un presupuesto dualista común: que podemos trazar una línea divisoria entre sujeto y objeto, o dicho en otras palabras: entre el hombre y el mundo.

1.2 El mito de Descartes

En el siglo XVII el filósofo Rene Descartes, en las *Meditaciones metafísicas* (1641), propuso otro tipo de dualismo, que el mundo está dividido en dos tipos de entidades: la *res extensa* y la *res cogitans*. Mientras que la primera refiere a todas las cosas físicas (todas las cosas que tienen extensión), la *res cogitans* refiere a los fenómenos mentales ya que estos, a diferencia de las cosas físicas, no tienen extensión y son inmateriales.

A pesar de las crítica, durante más de tres siglos esta tesis metafísica dualista se mantuvo vigente y de él se empezaron a derivar un par de tesis ontológicas de la mente. Para los herederos de la tradición cartesiana la mente reside única y exclusivamente dentro del cerebro; es decir, los

¹¹ “Where does the mind stop and the rest of the world begin?” ([Clark y Chalmers 1998, p. 7](#)).

fenómenos mentales son vistos desde una concepción internista. Además, otra consecuencia del dualismo cartesiano es que para esta tradición la mente (en tanto que existe dentro del cerebro del sujeto) no es extendida y es inmaterial.

A la par que el dualismo cartesiano, el dualismo sujeto-objeto también se mantuvo vigente. Así, aunque la manera de estudiar la relación hombre-mundo recibió diferentes cambios —como el famoso giro copernicano hecho por Immanuel Kant en la *Crítica de la razón pura* (1781), o la manera en que Ludwig Wittgenstein introdujo, en el *Tractatus Logico-Philosophicus* (1921), al lenguaje como mediador entre los pensamientos y el mundo— la idea central se mantuvo: podemos delimitar perfectamente entre lo que es el hombre y el mundo, por ello, para estudiar al hombre no necesitamos estudiar al mundo.

1.3 El cuerpo y el mundo

En la primera mitad del siglo XX aparecieron nuevas ideas que a la postre modificarían la manera en que se había pensando y entendido la mente, la obra de Martin Heidegger *Ser y tiempo* (1927) supuso un rompimiento con la idea de que podemos separar al mundo del hombre. Heidegger sostenía que la existencia estaba enmarcada e incrustada en el mundo; por ello, cualquier estudio que quisiera hacerse del hombre o de la existencia no puede ser separado del mundo, pues somos seres en el mundo. Con esto, las anteriores maneras de tratar la relación entre el hombre y el mundo quedaban obsoletas.

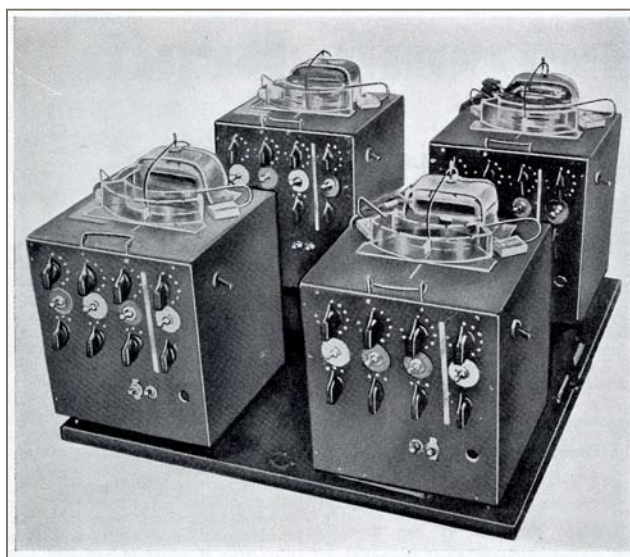
Además, dado que la mente era vista como una sustancia interna de los sujetos, una de la consecuencias directas fue que si no podemos separar al hombre del mundo entonces tampoco podemos estudiar a la mente separada del entorno. Por ello, Maurice Merleau-Ponty en la *Fenomenología de la percepción* (1945) también enfatizó el hecho de que el estudio de la percepción y la conciencia debe partir con el hecho de que el mundo siempre está ahí como una presencia anterior. Mientras que en *La estructura del comportamiento* (1942) hizo hincapié en el estudio del cuerpo para la cognición, un tema obviado y dejado de lado por la tradición filosófica analítica del siglo pasado.

Otro de los campos en donde nuevas ideas acerca de la vida, el comportamiento y de agentes inteligentes empezaron a surgir fue la cibernética.¹² Algunos de los trabajos pioneros en ésta área lo

¹² Véase *“Cybernetics and the Mangle: Ashby, Beer and Pask”* (2002) de Andrew Pickering.

constituyen *An Imitation of Life* (1950), de William Grey Walter, en donde expone por primera vez su trabajo al crear tortugas robóticas, las cuales han sido consideradas los primeros robots móviles y a pesar de su arquitectura minimalista fueron capaces de exhibir comportamiento cognitivo al interactuar entre ellas. Estas famosas tortugas de Walter han sido consideradas no solo los primeros robots móviles sino también los primeros robots creados e inspirados por un punto de vista biológico (Holland 2002). El homeostato de William Rosh Ashby, expuesto en *Design For a Brain* (1952), es un modelo del cerebro que no usa representaciones y que sin embargo es adaptativo al medio ambiente. Este es otro de los primeros ejemplos de cómo construir aparatos que pueden adaptarse al medio sin utilizar representaciones, véase la *Figura 2*.

A pesar de estos trabajos pioneros, que ponían mayor énfasis en la interacción de los agentes artificiales con el mundo más que en el uso de representaciones, y el señalamiento hecho por Heidegger y Merleau-Ponty (que la existencia del hombre está anclada al mundo) la influencia del dualismo cartesiano y del dualismo sujeto-objeto se mantuvo todavía gran parte del siglo pasado. Por ello, la cognición se siguió estudiando como un fenómeno aislado que sucede dentro de los sujetos, razón por la cual Jerry Fodor encasilló a estos proyectos bajo el título de *solipsismo metodológico*.



*Figura 2. Fotografía del homeostato de Ashby.*¹³

¹³ Imagen tomada de Ashby (1952/1960, p. 101).

1.4 El solipsismo metodológico

La naciente ciencia cognitiva ignoró totalmente estos trabajos y también las críticas tempranas hechas por Dreyfus (al proyecto de la inteligencia artificial) en *What Computers Can't Do (1979)*. Así, el proyecto de la ciencia cognitiva partió de la idea de que para entender y comprender la inteligencia, ésta tenía que ser modelada de acuerdo a la imagen que nos dan las computadoras. La combinación del internismo, junto con la influencia del entonces comienzo de la computación creó una nueva corriente en el estudio de la mente, a saber, el cognitivismo; dicha corriente parte de dos supuestos principales: la representación mental y el lenguaje de la computación para explicar la cognición, por ello sostenía que:

[...] la inteligencia (incluida la inteligencia humana) se parece tanto a un ordenador en sus características esenciales, que la cognición se puede definir como la computación de representaciones simbólicas. [Además] el cognitivista afirma que el único modo de dar cuenta de la inteligencia y la intencionalidad consiste en formular la hipótesis de que la cognición consiste en actuar sobre la base de representaciones que adquieren realidad física con la forma de un código simbólico en el cerebro o una maquina (Varela 1988, p. 37-39).

El cognitivismo planteó que la mente era como una computadora y que la manera en que funcionaba era mediante algoritmos; por lo tanto emular la inteligencia consistía entonces en llevar a cabo la resolución de problemas o completar ciertas tareas como, por ejemplo, jugar ajedrez. El proyecto del cognitivismo surgió en los años sesentas y se mantuvo como corriente principal durante más de veinte años en las ciencias cognitivas.

1.5 Ciencia cognitiva cartesiana

Como ya hemos visto, desde Descartes la mente había sido entendida como un proceso que ocurre dentro de nuestras cabezas, durante casi todo el siglo XX el solipsismo metodológico se mantuvo en pie y aunque el cognitivismo empezó a dejarse de lado y empezó a emerger el *conexionismo* (el cual defiende que la mente es como una red neuronal dinámica) las principales hipótesis seguían siendo las mismas: el internismo y el representacionalismo.

La influencia de estas ideas no sólo influyó en la filosofía sino también también en otros campos; el *internalismo psicológico*, por ejemplo, defiende la idea de que los estados mentales son

independientes de la relación entre el sujeto y su medio ambiente (Theiner 2011). Si bien el cognitivismo, el conexionismo y las nuevas ciencias de la mente tenían como objetivo superar el dualismo cartesiano, solo lo lograron a medias, tanto las ciencias cognitivas y la filosofía se habían ufano de haber abandonado todos los estigmas del dualismo cartesiano:

Contrario a lo que Descartes creyó, la mayor parte de la investigación dominante en psicología, ciencias cognitivas, y neurociencia cognitiva está predicada sobre suposiciones materialistas de que la mente es una clase (aunque una muy especial) de cosa física, abierta en principio al mismo tipo de investigación objetiva como cualquier otro fenómeno natural tal como las nebulosas espiral, la fotosíntesis y la tuberculosis¹⁴ (Theiner 2011, p. 1).

La investigación empírica de la cognición, en tanto que es materialista, va en contra de una de las ideas centrales del cartesianismo; sin embargo, ninguno de los proyectos del solipsismo metodológico se daba cuenta de que, sin querer, mantenían uno de los postulados esenciales del dualismo cartesiano, a saber, que los fenómenos mentales son algo interno del sujeto. Por ésta razón Rowlands (2010) ha denominado a esta manera de investigar la mente y la cognición como *ciencia cognitiva cartesiana*.

1.6 Las 4Es de la mente y la cognición

No fue sino hasta la última década del siglo XX que el paradigma de la ciencia cognitiva cartesiana empezó a ser seriamente criticado, desde diferentes campos del conocimiento empezaron a surgir diferentes tesis que rechazaron la suposición internista de la cognición. Uno de los primeros enfoques en cambiar este punto de vista fue la obra *The Embodied Mind* (1991), escrita por el biólogo Francisco Varela, el filósofo Evan Thompson y la psicóloga Eleanor Rosch, en donde defienden que el cuerpo es parte central de la cognición y que además estos están anclados al mundo. Así, a la manera en que Heidegger había señalado (que estamos puestos en un mundo pre-dado) ésta obra supuso también, en palabras de los propios autores, la continuación del proyecto empezado por Merleau-Ponty:

¹⁴ “Contrary to what Descartes believed, the bulk of mainstream research in psychology, cognitive science, and cognitive neuroscience is predicated on the materialist assumption that the mind is -albeit a very special- a kind of physical thing, open in principle to the same kind of objective investigation as any other natural phenomena such as spiral nebulae, photosynthesis, and tuberculosis” (Theiner 2011, p. 1).

Nos gusta considerar nuestro viaje en este libro como una moderna continuación del programa de investigación fundado una generación atrás por el filósofo francés Maurice Merleau-Ponty. [...] Nosotros sostenemos junto con Merleau-Ponty que la cultura científica occidental requiere que veamos a nuestros cuerpos como ambas cosas como estructuras físicas y como vividas, estructuras vivenciales —en suma, como ambos “externo” e “interno”, biológico y fenomenológico¹⁵ (Varela et al., 1991, p. xv).

La obra de Varela fue un rompimiento con lo que entonces eran las corrientes principales en las ciencias cognitivas, además supuso el punto de partida para estudiar la mente desde un punto de vista no solo psicológico o computacional sino también rescatando los aspectos biológicos y fenomenológicos. Esta corriente llamada *cognición corporizada dinámica* [dynamical embodiment cognition] (la tercera en las ciencias cognitivas) vino a sustituir a cognitivismo y al conexionismo; partiendo de este nuevo enfoque surgió una corriente que ha sido llamada enactivismo¹⁶.

Además de la obra de Varela y compañía, en el campo de la inteligencia artificial hubo un movimiento paralelo que rechazó otro de los presupuestos básicos del cognitivismo: el representacionalismo. Dicho movimiento fue llevado a cabo por Rodney Brooks, quien en su artículo *Elephants don't play chess* (1990) hizo un cambio de paradigma en la tarea de emular la inteligencia y la construcción de robots ya que dejó de lado las representaciones pues “[...] el mundo es su propio mejor modelo¹⁷” (1990, p. 3).

Brooks sostenía (en contra de los proyectos tradicionales de las ciencias cognitivas) que más que construir robots con algoritmos y representaciones lo que necesitamos es robots que interactúen con el mundo, por ello en *Intelligence without representation* (1991) escribe: “[...] deseo construir agentes móviles completamente autónomos que co-existan en el mundo con los humanos y que sean visto por los humanos como seres inteligentes en su propio derecho¹⁸” (p. 142). De este manera se inició con el diseño y construcción de robots que aprenden a base de prueba y error, y que encajan

¹⁵ “We like to consider our journey in this book as a modern continuation of a program of research founded over a generation ago by the French philosopher, Maurice Merleau-Ponty. [...] We hold with Merleau-Ponty that Western scientific culture requires that we see our bodies both as physical structures and as lived, experiential structures—in short, as both “outer” and “inner”, biological and phenomenological” (Rosch, Thompson et al. 1991 p. XV).

¹⁶ Véase la sección 3.3 del siguiente capítulo para una explicación más a fondo del enactivismo.

¹⁷ “[...] the world is its own best model” (Brooks 1990, p. 3).

¹⁸ “I wish to build completely autonomous mobile agents that co-exist in the world with humans, and are seen by those humans as intelligent beings in their own right” (Brooks 1991, p. 142).

de mejor manera con la idea que nos da la biología acerca de cómo aprendemos a interactuar con el mundo; los elefantes, por ejemplo, no saben jugar ajedrez pero el nivel de inteligencia que desempeñan implica un alto nivel cognitivo.

Por otro lado, en el año de 1995 Hutchins publicó *Cognition in the Wild*, en el cual muestra que la cognición es situada y que para la resolución de ciertas tareas cognitivas es importante la interacción entre varios elementos en donde se involucran: agentes, artefactos y recursos del medio ambiente. Mientras que en la filosofía, en el año de 1997 Clark publicó *Being There: Putting Brain, Body and World Together Again* en donde defiende que la cognición es corporizada y activa, aquí Clark analiza como los agentes interactúan en el mundo; además, en este trabajo podemos ya notar los esbozos y lo que podrían ser los prolegómenos a la tesis de la mente extendida.

En el año de 1998 la revista *Analysis* publicó un artículo de los filósofos Clark y Chalmers, *The Extended Mind*, en donde defendieron que la cognición y algunos estados mentales se extienden hacia el mundo y que por lo tanto algunos elementos externos al sujeto son también parte constitutiva de la mente. Así, con este nuevo tipo de externismo se hizo un rechazo total al internismo; además, esta tesis ontológica también rompió con las posturas tradicionales materialistas, pues éstas aunque rechazaban la idea de que la mente fuera algo inmaterial, aceptaban el carácter internista de la mente. Es por ello que Theiner (2011) afirma que la revolución que hace la tesis de la mente extendida es similar a la hecha por Darwin y Newton. Gallagher¹⁹ bautizó a estas cuatro principales corrientes (corporizada, embebida, enactiva y extendida)²⁰ como las 4Es (por su enunciación en inglés) de la mente y la cognición.

1.7 Un nuevo planteamiento de la agencia

Nuestra manera de vernos y percibirnos como agentes en el mundo cambia al momento de entender la cognición como un proceso extendido e integrado en el entorno; si la cognición es el producto de la interacción del cerebro, el cuerpo y el entorno, ¿qué clase de organismos somos? Clark (2008) expresa de manera perfecta este problema:

¹⁹ De acuerdo con Mark Rowlands, este nombre fue acuñado por Shaun Gallagher durante el desayuno de un taller sobre cognición situada organizado por Alessandra Tanesini y Richard Gray; posteriormente, en 2007, Gallagher organizó un congreso en la Universidad de Florida Central que llevó por título: “4e: The Mind Embodied, Embedded, Enacted, Extended”. Para mayor detalle véase la introducción del libro *The New Science of the Mind* (2010).

²⁰ En inglés las palabras correspondientes son *embodied*, *embedded*, *enacted* y *extended*, de ahí que se les llame las 4E.

¿Deberíamos de pensarnos a nosotros mismos solo como agentes cognitivos que co-optan y explotan las estructuras circundantes (por ejemplo, pluma y papel) en tanto que expandimos nuestras capacidades de resolución de problemas, o hay un sentido real en el cual el agente cognitivo (como opuestos a los organismos biológicos) es por lo tanto revelado como una entidad extendida que incorpora cerebro, cuerpo y algunos aspecto del medio ambiente local?²¹ (p. 14).

Al momento de entender la cognición como extendida surgen bastantes preguntas respecto a lo que somos y al asimilarnos como agentes: ¿la agencia puede ser distribuida o sigue siendo un fenómeno individual? ¿Debemos tener todavía una mirada antropomórfica de la agencia o podemos ampliarla a los demás componentes de un sistema cognitivo? ¿para explicar la agencia necesitamos todavía representaciones? Estas y otras preguntas son las que han inspirado este trabajo de investigación.

²¹ “Should we just think of ourselves as cognitive agents who co-opt and exploit surrounding structures (e.g., pen and paper) so as to expand out problem-solving capacities, or is there a real sense in which the cognitive agent (as opposed to the bare biological organism) is thus revealed as an extended entity incorporating brain, body and some aspects of the local environment?” (Clark 2008, p. 14).

Capítulo 2

La tesis de la mente extendida

“La mente había sido entendida como un espacio ontológico que el cartesianismo había imaginado como independiente y privilegiado, un lugar donde hay modelos internos [...] aislados del cuerpo, del entorno y de la situación” (2005, p. 10).

Jesús Vega Encabo - *Mentes híbridas*

Introducción

En el año de 1998 Clark y Chalmers dejaron de lado las respuestas tradicionales a la pregunta “¿dónde termina la mente y dónde empieza el resto del mundo?”²² las cuales, por un lado, o aceptan que lo que está fuera del cuerpo está fuera de la mente (en otras palabras, la piel y el cráneo son las fronteras de la cognición) o sostienen un externismo semántico y llevan este externismo —del significado— a toda la mente. La respuesta de Clark y Chalmers cobra importancia porque afirma que el entorno juega un papel activo en los procesos cognitivos, la relevancia de esta tesis es que rompe con la idea de que los fenómenos mentales (y en general la cognición) se llevan a cabo sólo dentro de nuestras cabezas. Rowlands en *The New Science of the Mind* señala que gracias la tesis de la mente extendida ahora la pregunta por la mente se formula de manera diferente:

¿Dónde empieza la mente y dónde termina? Esta es una pregunta inusual. La pregunta común [...] es: ¿qué es la mente? Y la respuesta usual es: el cerebro. Si esto es correcto, entonces tu mente empieza y termina donde tu cerebro empieza y termina; para tu mente está simplemente tu cerebro y nada más. Pero ¿dónde empieza tu cerebro y dónde termina?²³ (2010, p. 8).

²² “Where does the mind stop and the rest of the world begin?” (Clark y Chalmers 1998, p. 10).

²³ “Where does your mind begin, and where does it end? This is an unusual question. The usual question [...] is: what is the mind? And the usual answer is: the brain. If this is right, then your mind begins and ends where your brain begins and ends—for your mind is simply your brain and nothing else. But where does your brain begin and end? (Rowlands 2010, p. 8).

Clark y Chalmers formulan una pregunta inusual y su respuesta es que la mente no está confinada al cerebro sino que se expande hacia el mundo; de este modo, que la mente se extiende al mundo significa que los fenómenos mentales ya no pueden ni deben ser explicados (ni tampoco entendidos) sin tomar en cuenta el entorno.

En este capítulo presentaré la tesis de la mente extendida tal y como fue planteada por Clark y Chalmers (**sección 2.1**); mostraré los argumentos en favor de dicha tesis (**sección 2.2**) y expondré sus principales críticas y objeciones (**sección 2.3**). Posteriormente, haré una conclusión de este primer capítulo para, en el capítulo 3, responder a la crítica principal de la tesis de la mente extendida (**sección 2.4**).

2.1 Cognición extendida: ¿el iPhone es parte de mi mente?

La tesis de la mente extendida debe ser entendida literalmente: mi iPhone²⁴ forma parte de mi mente. Dado que en la consecución de procesos cognitivos entra en juego el entorno, las herramientas que me ayudan a llevar a cabo tareas cognitivas tienen que ser aceptadas como partes del proceso cognitivo. Para Clark y Chalmers el iPhone es parte constitutiva de mi mente porque, en la toma de decisiones, cumple el rol funcional y causal de algunos estados mentales.²⁵ Así, la idea es que la mente no es algo solamente interno sino que se extiende hacia el mundo y, por lo tanto, algunos elementos extra-craneales son parte constitutiva de la mente.

Con esta nueva tesis se rompe con una de las principales ideas cartesianas, a saber, el internalismo. Este nuevo punto de vista acerca de los fenómenos mentales es conocido como la tesis de la mente extendida, y para diferenciarlo del externismo de [Putnam \(1975\)](#), ha sido llamado externismo activo.

2.1.1 Externismo activo vs externismo semántico

En contra del psicologismo (la doctrina que afirma que los conceptos son algo mental) [Putnam \(1975\)](#) expone su externismo semántico, el cual sostiene que el contenido mental de los sujetos está determinado, en parte, por aspectos del medio ambiente que son externos a él. La tesis de de la

²⁴ En una conferencia de TED titulada “The Extended Mind” David Chalmers afirma que el iPhone es tan parte de su mente como su memoria biológica y que depende enteramente de él para muchas funciones en su vida cotidiana.

²⁵ Como veremos en el apartado 2.3, uno de los argumentos para sostener esta tesis apela a la relación causal y funcional que sostenemos con los artefactos; por ejemplo, la relación de Otto y su *notebook*.

mente extendida ha sido llamada externismo activo para poder distinguirla de la tesis de [Putnam \(1975\)](#) y [Burge \(1979\)](#), mientras que ellos son externistas acerca del contenido mental de los sujetos, C&C²⁶ son externistas acerca de los estados mentales; en otras palabras, Clark y Chalmers son externistas acerca de los portadores de los estados mentales y sostienen que el medio ambiente puede desempeñar un rol en la constitución y en la conducción de los procesos mentales.

Además, en el caso del externismo semántico los rasgos externos responsables de las creencias (como en el celebre argumento de la Tierra gemela de Putnam) son pasivos; es decir, mis creencias sobre el agua se mantienen gracias a mis anteriores experiencias, mientras que en la tesis de C&C el entorno se vuelve activo en el sentido de que puede ser parte del manejo del proceso cognitivo:

En los casos que describimos nosotros, en cambio, los factores externos relevantes son *activos*, tienen un papel crucial en el aquí y ahora. Dado que existe una complementación con el organismo humano, tienen un impacto directo en el organismo y en su conducta. [...] Al reflexionar sobre este tipo de complementariedad nos vemos abocados a un *externalismo activo*, opuesto al externalismo pasivo de Putnam y Burge²⁷ ([Clark y Chalmers 1998](#), p. 9)²⁸.

C&C sostienen que cuando tomo decisiones, o realizo determinadas tareas, lo hago mediado o con ayuda de factores externos que tienen un rol activo aquí y ahora; además, es en estos casos que el entorno juega un rol crucial en los procesos cognitivos. Pese a esto, es importante señalar que [Menary \(2010b\)](#) afirma que se pueden hacer dos lecturas del externismo activo, una más fuerte que la otra: la primera es una lectura que sostiene que algunos recursos del medio ambiente influyen activamente en los procesos cognitivos llevados a cabo en el cerebro, mientras que la segunda lectura sostiene que algunos procesos cognitivos están constituidos por recursos del ambiente.

En lo que están pensando C&C es en la segunda versión, es decir, defienden que algunos procesos cognitivos están constituidos por características del medio ambiente; razón por la cual no

²⁶ De aquí en adelante algunas veces utilizaré “C&C” para referirme a Clark y Chalmers.

²⁷ “In the cases we describe, by contrast, the relevant external features are active, playing a crucial role in the here-and-now. Because they are coupled with the human organism, they have a direct impact on the organism and on its behavior. [...] Concentrating on this sort of coupling leads us to an active externalism, as opposed to the passive externalism of Putnam and Burge” ([C&C 1998](#), p. 9).

²⁸ Para el artículo de [Clark y Chalmers \(1998\)](#) utilizaré la traducción de Eva Aladro.

podemos reducir el externalismo activo a una mera tesis causal, pues más que causal esta tesis es constitutiva: las herramientas que forman parte en los procesos cognitivos no sólo son importantes por su rol causal sino también porque forman parte de la mente.

2.1.2 Postulados de la tesis de la mente extendida

La tesis de la mente extendida —tal y como la presentaron [Clark y Chalmers \(1998\)](#)— sostiene que el entorno juega un papel activo en los procesos cognitivos. Este eslogan se descompone en diferentes postulados que podemos resumir en una tesis ontológica: la cognición y la mente se extienden al mundo, de este modo parte de lo que llamamos mente está constituido por elementos del entorno que están fuera de la cabeza del sujeto.

Sin embargo para que esta tesis se cumpla, los elementos que están fuera de la cabeza tienen que cumplir con la tesis funcionalista: llevar a cabo el rol funcional de algunos de los mecanismos internos que conforman la cognición; así, una vez que cumplen con este criterio pueden ser consideradas parte de la mente.

Algunas otras posturas externistas están de acuerdo en que los recursos del entorno son centrales para entender la cognición pero, pese a esto, no todos se comprometen con la idea de que por eso, la mente esté conformada por artefactos o elementos del entorno. Veamos cuáles son los argumentos para defender el externismo activo.

2.2 Argumentos a favor de la mente extendida

El argumento que dan C&C (1998) está basado enteramente en el funcionalismo y en el *principio de paridad*²⁹, además, sus premisas están sostenidas en dos ejemplos que son el juego del Tetris y el caso de Otto. Veamos las premisas del argumento:

i) Un estado cognitivo (o mental) E se define por su rol funcional F.

ii) Si tenemos casos en donde un elemento externo cumple el rol funcional F,

entonces algunos elementos externos son parte de un estado estado cognitivo E y por lo tanto parte constitutiva de la mente.

²⁹ Para una explicación del *principio de paridad* veas la sección 2.2.3

iii) Tenemos casos (como el caso de Otto) en donde un elemento externo cumple el rol funcional F.

∴ por ii y iii podemos concluir que en algunos casos ciertos elementos externos son parte de un estado cognitivo y por lo tanto son parte constitutiva de la mente.

Los casos que se plantean para defender la premisa iii) son el juego del Tetris y el caso de Otto, ambos están pensados para defender la tesis ontológica de la mente y la cognición; pero, mientras que el primer caso busca mostrar que algunos procesos cognitivos se extienden al entorno, el segundo defiende que no todos los estados mentales están determinados internamente y que, por lo tanto, no solo la cognición se extiende al entorno sino también la mente.

2.2.1 Jugando Tetris

El primer ejemplo para defender la premisa tres nos hace pensar en tres escenarios en los que un sujeto se enfrenta al mismo problema: insertar figuras geométricas en un espacio; veamos los tres escenarios:

1. En el primer caso Aleks³⁰ está sentado frente a una computadora y se le muestran figuras bidimensionales; la tarea de Aleks consiste en responder si dichas figuras encajan con una conexión que aparece dibujada. (Para completar la tarea, Aleks tiene que rotar mentalmente las imágenes y ver si encajan en las conexiones).
2. En el segundo caso se le exige a Aleks la misma tarea que en el escenario uno, pero esta vez, puede escoger entre rotar físicamente (apretando un botón) las imágenes de la pantalla o hacerlo mentalmente como en el caso uno. Además, citando el trabajo de [Kirsh y Maglio \(1994\)](#), C&C sostienen que hay una mayor velocidad en la rotación física que en la rotación mental del escenario 1.
3. En el tercer escenario se nos invita a pensar en un futuro cibernético en donde Aleks tiene un implante neuronal y este implante puede llevar a cabo la rotación de las figuras tan rápido como

³⁰ Clark y Chalmers no utilizan ningún nombre en particular en este argumento, yo utilizo el nombre Aleks en honor a Alekséi Pázhitnov ya que fue él quien inventó el tetris en 1984.

la opción de aplastar físicamente el botón del caso dos; de esta manera, al estar jugando Tetris, Aleks tiene que escoger entre los recursos internos que posee: utilizar el implante neuronal o la rotación mental. C&C afirman que cada uno de estos recursos demandan diferente atención y otras actividades cerebrales.

Clark y Chalmers defienden que el caso 3 es similar al caso 1 y que, además, el escenario 2 es igual al caso 3 (salvo con la diferencia de que en el escenario 3 el implante está dentro del sujeto). Pero, si consideramos que en el escenario 3 la rotación es cognitiva, ¿por qué rechazaríamos que en el escenario 2 la rotación sea cognitiva?

Dado que en los tres casos se está haciendo uso del mismo rol funcional (girar las piezas) sería meramente arbitrario decir que la rotación en el escenario 2 no es cognitiva; además, apelar a la piel o el cráneo como límite cognitivo (y decir que los casos 2 y 3 no son similares) sería una petición de principio porque justo la legitimidad de ese límite es lo que se está discutiendo.

2.2.2 La memoria de Otto

Inga y Otto están en Manhattan y desean ir a una exposición al MoMA. Cuando Inga escucha que hay una nueva exposición decide ir, por lo tanto piensa un momento y entonces recuerda que este museo está ubicado en la calle 53; acto seguido camina a la calle en cuestión y llega a su destino. El caso de Otto es diferente porque él padece de alzheimer, “y como otros enfermos de este mal, confía en la información del entorno para ayudarse a estructurar su vida” (C&C 1998, p 7). Otto lleva su libreta de notas a todos lados, así que cuando decide ir al MoMA, primero consulta en su libreta en qué calle se encuentra el museo, después de que su notebook le informa que éste se encuentra en la calle 53, así que camina y llega para ver la exposición.

En términos funcionales da igual cómo ambos llegaron al museo, pues los dos cumplieron su propósito de llegar al MoMA; pero, y aquí es dónde C&C ponen todo el peso de argumentación, para ellos hay algo crucial en este proceso de recordar: mientras que Inga utiliza su memoria biológica, Otto apela a una memoria artificial. Inga lo que tiene que hacer es recordar y buscar en su mente la respuesta a la pregunta ¿dónde queda el MoMA?, mientras que Otto necesita de su libreta de notas para recordar. Así, para Otto, su libreta desempeña el rol que antes desempeñaba su

memoria biológica; el comportamiento de Otto está causalmente determinado por un artefacto que es externo a él.

Para C&C decir que la libreta no es parte del estado mental recordar —porque es externa a Otto— mientras que considerar la memoria biológica de Inga como parte constitutiva del proceso de recordar (por el simple hecho de ser algo interno) es algo completamente arbitrario.

2.2.3 El principio de paridad

Lo que llamamos el principio de paridad es la tesis que C&C defienden para dar soporte a sus argumentos, fue enunciado de la siguiente manera:

Si, cuando abordamos una tarea, una parte del mundo funciona como un proceso que, *si se hiciera dentro de nuestra cabeza*, no dudaríamos que es parte del proceso cognitivo, entonces esa parte del mundo (defendemos nosotros) es parte del proceso cognitivo³¹ (C&C 1998, p. 29).

Para [Menary \(2010b\)](#) el principio de paridad desempeña dos roles en la argumentación: 1) nos invita a reconsiderar nuestros prejuicios cartesianos: si un mecanismo interno es parte de la cognición ¿por qué uno externo no? 2) Introduce las credenciales funcionalistas de la tesis de la mente extendida: mientras un proceso tenga una función cognitiva no importa dónde se encuentre, da igual si se encuentra dentro o fuera del sujeto, debe ser tomado como parte del proceso cognitivo. Vemos lo que [Sutton \(2010\)](#) nos dice al respecto:

El principio de paridad destaca el isomorfismo funcional de los procesos internos y externos de los estados. Podría ser chauvinista e injusto tratar la libreta de Otto diferentemente del cerebro de Inga solo porque uno es externo: y, como defienden Clark y Chalmers, dado que juegan el mismo rol funcional en la conducción de comportamiento en ambos casos, y cuando no están en uso ambos

³¹ “If, as we confront some task, a part of the world functions as a process which were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process” (C&C 1998, p. 29).

están todavía preparados y disponibles para jugar esos mismos roles, no hay una buena razón en absoluto para tratarlos diferentemente³² (2010, p. 195).

De este modo, el caso de Otto y el *principio de paridad* sirven para desechar las ideas chovinistas que se tienen acerca de la mente. Finalmente, siguiendo a [Menary \(2010\)](#), es importante señalar que el principio de paridad se preocupa más por la *función* que por la *localización*; es decir, lo que importa cuando queremos determinar si un proceso es cognitivo o no, no es su localización sino su *rol funcional*. Además, no importa si los procesos internos tienen diferentes características a los elementos externos, de hecho pueden ser del todo diferentes, lo que importa es el rol funcional que puedan desempeñar.

Así, otra característica de la tesis de la mente extendida es que, en el fondo, tiene un carácter funcionalista: mientras un artefacto externo logre emular una de las funciones de los procesos internos, éstos tienen que ser considerados como parte de la mente.

2.2.4 Conclusiones de los argumentos

Clark y Chalmers defienden que, muchas veces, el organismo humano está vinculado a una entidad externa (en una interacción en dos direcciones) y que esto crea un sistema ensamblado que debe ser considerado cognitivo. Ahora bien, para defender que este sistema ensamblado es cognitivo fundamentan la premisa iii) en el caso del Tetris y esto los lleva a la siguiente conclusión:

a). Este tipo de proceso ensamblado (girar las piezas del Tetris) es igualmente un proceso cognitivo, no importa si ocurre completamente dentro de la cabeza o no.

Los procesos cognitivos bien pueden extenderse al entorno, pero tal vez el entorno no juegue un papel importante o determinante en los estados mentales; el segundo caso, el de Otto, justo lo que busca es mostrar que el entorno sí juega un papel activo e importante en la consecución de los

³² “The parity principle stresses the functional isomorphism of inner and outer process and states. It would be chauvinistic and unfair to treat Otto’s notebook and Inga’s brain differently just because one is external: and, claim Clark and Chalmers, since they play the same functional role in driving behavior in both cases, and when not in use are both still poised and available to play those same roles, there is no good reason at all to treat them differently” ([Sutton 2010, p. 195](#)).

estados mentales y por lo tanto extienden no sólo la cognición sino también la mente. Uniendo la premisa tres con el principio de paridad y el caso de Otto podemos concluir que:

b). Todos los componentes del sistema ensamblado juegan un papel causal activo; así, la conducta estará gobernada conjuntamente por todos los componentes del mismo modo que se hace en la cognición usual. “Si eliminamos el componente externo del sistema de competencia conductual, es como si elimináramos una parte del cerebro mismo” (C&C 1998, p. 3).

Esta segunda conclusión podemos perfectamente aplicarla al caso de Otto: si al sistema ensamblado de Otto y su notebook le eliminamos una parte (por ejemplo la notebook) la conducta de Otto cambiará radicalmente y por lo tanto no podrá llegar al MoMA. C&C postulan el caso de Otto porque bien podemos aceptar que la cognición se extiende al entorno, pero rechazar que el entorno juegue un papel activo y que, por lo tanto, sea determinante en los estados mentales; pues dicha conclusión sería perfectamente compatible con la idea de que, a pesar de que la cognición sea extendida, los verdaderos estados mentales (creencias, deseos, emociones, etc.) sean determinados por procesos cerebrales internos. Así, mientras que el primer caso se enfoca en defender que la cognición se extiende al entorno, el caso de Otto tiene un objetivo mucho más fuerte y complicado: defender que —no sólo la cognición sino también— la mente se extiende al entorno.

La argumentación de C&C sigue tres pasos: 1) defienden que la cognición se extiende al entorno, 2) argumentan que el entorno juega un papel causal relevante en algunos estados mentales y que, por lo tanto, 3) la mente también se extiende al entorno. De esta manera, se llega a la conclusión de que la creencia de Otto no está solo dentro de su cabeza.

2.3 Defendiendo los límites de la cognición

Hasta aquí, hemos presentado la tesis central de la propuesta de Clark y Chalmers, a su vez hemos presentado el argumento funcionalista que subyace en la tesis de la mente extendida; también hemos analizado como la tercera premisa de este argumento está fundamentada en los casos del Tetris y de Otto. En esta sección presentaré algunas de las críticas más contundentes que se han hecho a ésta tesis.

Un rasgo distintivo de las diferentes críticas que se han hecho es que aceptan la premisa i) y ii) pero su estrategia consiste en rechazar enteramente la premisa iii). El rechazo de esta premisa

implica 1) criticar el principio de paridad y 2) tratar de caracterizar lo cognitivo para, de este modo, descartar que el caso de Otto sea considerado cognitivo.

2.3.1 Críticas al principio de paridad

Como hemos visto el principio de paridad cuestiona de manera crítica los límites entre el cerebro, el cuerpo y el mundo; “[l]a idea esencial es que, en términos cognitivos, lo relevante no es la localización espacial del proceso” (Martínez 2013, p. 38). De este modo, dicho principio va en contra de la aceptada suposición de que lo que está dentro del sujeto es cognitivo y lo que está fuera no lo es (Sutton 2010). Sin embargo, uno de los primeros problemas con el principio de paridad es que al centrarse únicamente en el rol funcional deja de lado diferencias significativas entre los recursos externos e internos; por ejemplo, la memoria biológica de Inga y la notebook de Otto (aunque jueguen el mismo rol funcional) poseen diferentes características que no se pueden obviar al momento de hacer investigación interdisciplinaria:

Si el cerebro de Inga y la libreta de Otto están jugando un rol similar en impulsar lo que cada uno de ellos este impulsando, entonces no hay un punto especial en estudiar las propiedades particulares o de los cerebros o de nuestras libretas y otros medios externos. [...] La paridad no deja un espacio obvio para investigar diferencias individuales en relación con EM [la tesis de la mente extendida], porque nos pide centrarnos en características genéricas de los estados y procesos cognitivos [...]³³ (Sutton 2010, p. 198-199).

Además de esta falla, Rupert, en *Cognitive Systems and the Extended Mind* (2009),³⁴ señala otros problemas con el principio de paridad: el poco alcance que el principio de paridad logra. Dado que no arroja nada de luz sobre la cognición (es decir, teóricamente no revela nada acerca de las propiedades del procesamiento cognitivo), al momento de decidir si un proceso es parte de la mente o no, no tenemos una manera clara de responder si sí o si no.

En lugar de responder directamente a esta crítica, en el capítulo tres presentaré una nueva versión de la tesis de la mente extendida que se deshace del principio de paridad y propone otro

³³ “If Inga’s brain and Otto’s notebook are playing relevantly similar roles in driving what they each do, then there’s no special point in studying the peculiar properties either of brains or of notebooks and other external medias. [...] Parity leaves no obvious space for investigating individual differences in relation to EM [the extended mind hypothesis], because it asks us to focus in generic features of cognitive states and processes [...]” (Sutton 2010, p. 198-199).

³⁴ Para una análisis detallado de éstas críticas véase el capítulo dos: *Principles of Demarcation* de este libro.

principio que no se basa en la equivalente funcional entre los recursos internos y externos³⁵; esta nueva versión es de [Menary \(2007\)](#) y es llamada integración cognitiva. Pero antes de pasar al siguiente capítulo veamos una de las críticas más citadas para descartar la tesis de la mente extendida.

2.3.2 La marca de lo cognitivo

Otra de las estrategias para responder a los argumentos de Clark y Chalmers es proponiendo un principio que nos permita distinguir entre lo que es cognitivo y lo que no lo es. Como escribe Rupert: ¿qué pensaríamos que distingue una genuina contribución cognitiva de una que no; de una que es, como uno podría decir, meramente causal?³⁶ ([2009, p. 16](#)). Dicho en otras palabras lo que necesitamos es buscar la marca de lo cognitivo, esta marca no es otra cosa que un criterio normativo que nos permita saber cuándo un proceso es cognitivo y cuando no.

Ésta crítica esgrimida también por [Adams y Aizawa \(2001, 2010\)](#) consiste en demarcar lo que constituye un proceso cognitivo, la finalidad de esta crítica es descartar el caso de Otto sugiriendo que no podemos considerar cognitivo a su proceso porque no cumple con lo que hemos llamado cognitivo. Sin embargo, antes de proseguir, es importante aclarar que este argumento no busca desarticular el principio de paridad, o al menos no de manera directa; la marca de lo cognitivo no está negando de entrada que pueda haber procesos cognitivos externos. La estrategia consiste en postular algunas condiciones para aplicar el término cognitivo, después de que tenemos estas condiciones vamos a los casos y entonces notaremos que ninguno de estos casos de cognición es externista como sugiere la tesis de la mente extendida.

Otro punto que debe ser tomado en cuenta es que la marca de lo cognitivo debe de ser diferenciada de la marca de lo mental, A&A³⁷ afirman que dichas marcas son diferentes. El problema es que toman esta distinción como una suposición de partida sin explicar por qué tendríamos que trazar dicha división, nosotros bien podríamos preguntar si puede haber cognición sin mente. Y si esto es el caso, entonces ¿qué sería tener una mente? Sin embargo, por *mor* de la argumentación aceptemos esta distinción y enunciemos que la marca de lo cognitivo es aquella que

³⁵ Véase el capítulo 3, sección 3.1.3

³⁶ “What, thought, distinguishes a genuinely cognitive contribution from one that is not—one that is, as one might say, merely causal?” ([Rupert 2009, p. 16](#)).

³⁷ De aquí en adelante algunas veces utilizaré “A&A” para referirme a Adams y Aizawa.

nos va a permitir saber la naturaleza de los procesos cognitivos como aprender, recordar, percibir, pensar, etc.

Dado que no podemos simplemente decir que lo cognitivo es aquello que se da entre el sujeto y las herramientas porque entonces “[...] necesitamos una teoría de lo que hace a algo un agente cognitivo”³⁸ (A&A 2010, p. 68), Adams y Aizawa prosiguen con su estrategia argumentativa. Así, postulan dos condiciones que deben de satisfacer todas las cosas del mundo si es que queremos aplicarles el término cognitivo: i) los estados cognitivos deben envolver intrínsecamente contenido no derivado y ii) el procesamiento cognitivo es de una cierta clase que no puede ser imitada por una computadora.

La idea de i) es que mientras que los números o símbolos adquieren su significado mediante prácticas y convenciones sociales, lo que llamamos estados cognitivos no debe derivar su significado de estas prácticas; por otro lado, ii) está basada en la idea de que la manera en que las computadoras llevan a cabo procesos cognitivos es diferente a como los humanos hacen esos procesos cognitivos. Ésta segunda condición tiene que ver con la naturaleza del procesamiento cognitivo y afirma que la cognición es algo más que meramente pasar la prueba de Turing; así, uniendo a i) y ii) tenemos que la “[...] cognición involucra tipos particulares de procesos involucrando representaciones no derivadas”³⁹ (A&A 2001, p. 52-53).

Adams y Aizawa están defendiendo que la cognición es de cierta clase natural que no puede ser igualada por otros vehículos (no biológicos) que procesan información, como por ejemplo las computadoras. Una vez teniendo esta marca de lo cognitivo, A&A concluyen que, por ejemplo, el argumento del Tetris⁴⁰ no es válido porque los escenarios 1 y 2 difieren en su uso de representaciones de contenido no derivado; mientras que para el caso de Inga y Otto afirman que sus creencias no son de la misma clase natural, o mejor aún, el caso de Otto no cae dentro de la clase natural creencia.⁴¹

En resumen, lo que hace esta crítica es plantear requisitos normativos a la cognición: postula condiciones para delimitar lo que podemos llamar procesos cognitivos y de esta manera descartar

³⁸ “[o]ne still needs a theory of what makes something a cognitive agent” (A&A 2010, p. 68).

³⁹ “[...]cognition involves particular kinds of processes involving non-derived representations” (A&A 2001, p. 52-53).

⁴⁰ Véase la sección 1.3.1 de este capítulo.

⁴¹ Para una respuesta a estos argumentos véase: *Memento's Revenge: The Extended Mind, Extended* (2010a) de Clark.

muchos procesos que —al no cumplir con las condiciones estipuladas— no pueden ser considerados como procesos cognitivos. Al igual que con las demás críticas, se han ensayado diferentes respuestas a éstas; algunas salidas, como la de [Rowlands \(2010\)](#)⁴², proponen otro tipo de criterio cognitivo de tal manera que los procesos híbridos (a los cuales nos lleva la tesis de la mente extendida) puedan contar como cognitivos; además, algunas críticas señalan que se tiene que saber exactamente qué es eso que se llama “contenido no derivado”.⁴³ En el capítulo tres (sección 3.3) veamos cómo desde el enactivismo radical esta crítica queda descartada.

2.3.3 La falacia constitución-acoplamiento: ¿algunas mentes tienen más partes que otras?

La principal crítica en contra de la tesis de la mente extendida se pregunta ¿cómo es posible que de una simple relación causal (o funcional) se siga que un objeto externo es parte de la mente? Fodor en *Where is my mind?* (2009) muestra, a través de una reducción al absurdo, que si el principio de paridad es aceptado entonces debemos aceptar conclusiones bizarras del tipo “Otto tiene más mente que Inga”. El punto que desea señalar Fodor es la crítica que Adams y Aizawa han hecho desde el año 2001: ¿cómo puede mi conexión causal con mi iPhone ser parte de mi mente?

Además, las críticas de Fodor preguntan cómo es posible que las mentes tengan partes y exactamente cuantas partes diríamos que tienen nuestras mentes. [Adams y Aizawa \(2001, 2010\)](#) han hecho esta misma crítica y la han llamado la *falacia constitución-acoplamiento* (*coupling-constitution fallacy*); esta crítica puede resumirse en un diálogo de dos renglones que A&A presentan en su celebre artículo *Defending the Bounds of Cognition* (2010):

Pregunta: ¿Por qué el lápiz piensa que $2 + 2 = 4$?

Respuesta de Clark: Porque el lápiz está ensamblado al matemático.

Para A&A estos dos renglones nos muestra lo que está mal con la tesis de la mente extendida, es incorrecto hacer cognitivo a cualquier objeto por el simple hecho de ensamblarlo a un agente cognitivo; veamos por qué este paso es falaz:

⁴² Véase el capítulo 8 “The Mark of the Cognitive” del libro *The New Science of the Mind*.

⁴³ Los mismos [Adams y Aizawa \(20001, p. 49\)](#) señalan que la tesis de que la cognición involucra contenido no derivado tiene que ser contrastada con, por ejemplo, el *representacionalismo*, el cual nos dice que la cognición envuelve el uso de representaciones con contenido no derivado.

El patrón de la falacia es llamar la atención para casos, real o imaginados, en los cuales algunos objetos o proceso están ensamblados de alguna manera para algún agente cognitivo. De esto, uno se desliza a la conclusión de que el objeto o proceso constituye parte del aparato cognitivo del agente o del proceso cognitivo⁴⁴ (Adams & Aizawa 2010, p. 68).

Del hecho de que un objeto esté ensamblado —en algún sentido— a un agente, no sigue que ese objeto forme parte del agente cognitivo; en otras palabras, del hecho de que algunas herramientas externas jueguen un rol importante en la cognición, no se sigue la tesis ontológica que afirma que esas herramientas son parte constitutiva de la mente. Además, Adams y Aizawa sostienen que esta crítica muestra que aceptar el caso de Otto es una pendiente resbaladiza: no sabremos donde poner los límites de aquello que constituye parte de la mente y aquello que no.⁴⁵

Tanto la crítica de Fodor como la Adams y Aizawa han hecho que Clark (2010) tenga que agregar algunas restricciones, por que de lo contrario, casi cualquier cosa podría entonces ser parte de la mente; de manera preliminar dicha estrategia pueda funcionar pero una mejor salida sería rechazar el principio de paridad y postular otro principio que no se centre exclusivamente en el rol funcional. El capítulo tres estará dedicado enteramente a responder o mostrar algunas salidas a esta crítica.

2.4 Conclusiones del capítulo

Con la tesis de Clark y Chalmers, la cognición se contempla como un proceso continuo con el entorno; llamaremos a esta propuesta la tesis tradicional de la mente extendida para poder diferenciarla de otros tipos de cognición, que aunque externistas, hacen diferencias significativas en la manera de entender los procesos cognitivos. La propuesta filosófica de la tesis tradicional de la mente extendida es una tesis ontológica que sostiene que la mente se extiende hacia el mundo, en el corazón de dicha propuesta subyace la tesis funcionalista que es expresada principalmente por el

⁴⁴ “The fallacious pattern is to draw attention to cases, real or imagined, in which some object or process is coupled in some fashion to some cognitive agent. From this, one slides to the conclusion that the object or process constitutes part of the agent’s cognitive apparatus or cognitive processing” (Adams & Aizawa 2010, p. 68).

⁴⁵ Para ver la replica de Andy Clark, a estas objeciones, véase *Coupling, Constitution, and the Cognitive Kind: A reply to Adams and Aizawa* (2010b).

principio de paridad: si un artefacto logra llevar a cabo el mismo rol funcional que un proceso interno, este artefacto debe ser considerado como parte de la mente.

En contra de esta tesis, algunas de las críticas han optado por rechazar dicho principio; mientras que otras señalan que la tesis de la mente extendida comete la *falacia constitución-acoplamiento*: ¿por qué mi relación causal con un artefacto (como el iPhone, por ejemplo) hace que dicho artefacto se vuelva parte de mi mente? Además, esta crítica implica otra consecuencia: si aceptamos casos de cognición extendida ¿cómo podemos trazar un límite entre lo que es parte de la mente y lo que no? El siguiente capítulo estará dedicado enteramente a dar respuestas a la falacia constitución-acoplamiento, las respuestas se esbozaran desde las posturas de [Menary \(2007\)](#), [Chemero \(2009\)](#) y [Hutto & Myin \(2013\)](#).

Capítulo 3

Tres respuestas a la *falacia constitución-acoplamiento*

*“Una de las ideas filosóficas más peligrosas es, curiosamente, la de que pensamos con la cabeza o en la cabeza. La idea del pensar como un proceso en la cabeza, en un espacio absolutamente cerrado, le da el carácter de algo oculto”*⁴⁶.
(1967, p. 605-606).

Ludwig Wittgenstein - *Zettel*

Introducción

En una escena de la película *Matrix*, Morfeo le dice a Neo que desde los inicios de la humanidad el hombre siempre ha necesitado de máquinas para sobrevivir. Si recordamos bien, algo similar han dicho [Clark y Chalmers \(1998\)](#): no es la avanzada tecnología lo que ha originado el debate de la mente extendida sino la dependencia que los seres humanos tenemos hacia el uso de las herramientas. Ejemplos de esta necesidad la constituyen inventos milenarios como la rueda, los arcos, el alfabeto, la imprenta, la calculadora, la brújula, modelos en la ciencia e incluso —muchas veces— recursos de nuestro alrededor como es el caso las estrellas⁴⁷.

Viendo el uso que hacemos de estas tecnologías, [Menary \(2007, 2010\)](#) ha propuesto una nueva versión acerca de la tesis de la mente extendida; esta versión (que recibe el nombre de cognición integrada [cognitive integration]) sostiene que la mente —y con ella la cognición— es híbrida. La cognición involucra procesos neuronales y corporales que a su vez están integrados con elementos que se encuentran en el medio ambiente, la cognición integrada es un intento por entender la naturaleza de la integración de estos elementos ([Menary 2010, p. 229](#)). Como ya hemos visto en el capítulo 2, la *falacia constitución-acoplamiento* es la crítica principal que se ha hecho a la tesis de la mente extendida. Razón por lo cual, el objetivo de este capítulo es hacer una presentación de tres

⁴⁶ “One of the most dangerous ideas for a philosopher is, oddly enough, that we think with or in our heads. The idea of thinking as a process in the head, in a completely enclosed space, give him occult.” ([Wittgenstein 1967, p. 605-606](#)).

⁴⁷ [Hutchins \(1995\)](#) nos señala el papel que jugaron las estrellas en la navegación.

respuestas que podemos trazar para responder a dicha falacia. La primera de ellas es la cognición integrada propuesta por Menary y desarrollada sobre todo en *Cognitive Integration - Mind and Cognition Unbounded* (2007). La estrategia de Menary es mostrar que los procesos cognitivos son híbridos y con ello responder a las críticas internistas hecha por Adams y Aizawa (**sección 3.1**).

La segunda respuesta que presentaré (**sección 3.2**) será el externismo cognitivo dinámico de Chemero (2009) presentada en *Radical Embodied Cognitive Science*, dicha propuesta nos permite mantener casos de cognición extendida pero, a diferencia de la tesis tradicional de la mente extendida, se deshace de las representaciones mentales. Además, su enfoque propone mirar a la cognición desde el punto de vista de los sistemas dinámicos.

Finalmente, la última propuesta que presentaré será el enactivismo radical de Hutto y Myin (2013), expuesto en *Radical Enactivism - Basics Minds Without Mental Content*, que más que responder a la *falacia constitución-acoplamiento* propone disolver dicho problema. Lo novedoso en esta propuesta es que no solo se deshace de las representaciones sino también del *contenido mental*. De acuerdo con Hutto y Myin, solo al hacer esto podemos dar una respuesta clara y definitiva a las críticas internistas (**sección 3.3**).

Finalmente, para terminar este capítulo daré un breve resumen de las propuestas aquí mencionadas todo ello con miras de entrar al tema de la agencia en el capítulo 4 (**sección 3.4**). Es importante recalcar que presento tres respuestas diferentes, a la falacia constitución-acoplamiento, para poder mostrarle al internista que se le puede responder desde posturas representacionistas, anti-representacionistas y enactivas. Sin embargo, debido a los alcances de esta tesis algunas otras respuestas no serán presentadas; tal es el caso de el *enactivismo autopoietico* de Di Paolo (2009), que hace compatible al enactivismo con la tesis de la mente extendida. Tampoco presentaré otras propuestas híbridas de la mente como la de Rowlands (1999) y Wilson (2004).

3.1 Cognición integrada

Cotidianamente nos volvemos agentes cognitivos más poderosos gracias a la ayuda de herramientas tecnológicas más potentes como es el caso del internet y de las computadoras personales, es por ello que pensar en nosotros como agentes híbridos ya no es cosa de ciencia ficción. Los teléfonos inteligentes y nuestro uso de google, por ejemplo, se vuelven cada día una parte central en la resolución de nuestras tareas cognitivas cotidianas; por esta razón, Menary (2007, 2010, 2012)

defiende, al igual que Clark y Chalmers, que los límites de la cognición no pueden estar confinados a los límites del cerebro: si la cognición está limitada al cerebro, entonces ¿por qué no podemos completar todas las tareas cognitivas dentro de nuestras cabezas? (Menary 2010a).

La cognición integrada tiene también un enfoque externista de la cognición, pero incluye nuevas propuestas ya que “[...] [t]oma a los sistemas cognitivos siendo un todo integrado que tienen partes interactuando, pero estas partes pueden ser neuronales, corporales y componentes del entorno”⁴⁸ (Menary 2012, p. 150). Menary parte mostrando que su propuesta tiene una base biológica y evolutiva, por ello su estrategia consiste en presentar algunos casos de la biología en donde queda patente que, en la naturaleza, la relación entre organismo-medio ambiente no respeta los límites tradiciones entre lo interno y externo. De este modo, al igual que en la biología, se pretende estudiar la cognición como un sistema en donde interactúan el agente y el medio ambiente.

3.1.1 La cognición como un sistema biológico

En la naturaleza podemos observar que distintos organismos logran integrarse con elementos de su medio ambiente, un buen ejemplo de estos casos biológicos lo constituye la araña y su telaraña; aunque la telaraña se encuentra fuera del sistema nervioso central del organismo, ésta influye de manera determinante en el comportamiento global del sistema (en este caso la araña). Menary se apoya en el trabajo de Dawkins (1982) acerca de los fenotipos extendidos para mostrar que la cognición no debe respetar como límite la piel de los agentes humanos, la idea de fondo en toda la argumentación es que, del mismo modo en que los organismos se extienden dentro de su medio ambiente, la cognición también es extendida.

Algunos ejemplos que Menary cita de Dawkins son: i) los pájaros, pues estos construyen sus nidos de materiales que encuentran en el medio ambiente; ii) los caracoles, ya que estos construyen sus conchas del calcio que obtiene de su comida y iii) los cangrejos ermitaños (*dardanus calidus*), ya que estos heredan las conchas de caracoles muertos. El punto que Menary quiere traer a colación con estos ejemplos es que, cuando estudiamos a las aves no sólo las estudiamos a ellas sino también a sus nidos. Estudiamos al ave y a su nido como un sistema biológico, por ello “[l]a distinción entre

⁴⁸ “[it] takes cognitive systems to be integrated wholes that have interacting parts, but these parts can include neural, bodily and environmental components” (Menary 2012, p. 150).

lo que está espacialmente ‘adentro’ y lo que está espacialmente ‘afuera’ del pájaro, en cuanto tal, no tiene significado para el estudio del sistema biológico aviar” (Millikan (1993, p. 159)⁴⁹.

La propuesta que subyace en la tesis de Menary es que la cognición debe ser estudiada de igual modo que un sistema biológico: cuando estudiamos el comportamiento de los pájaros estudiamos también los nidos, cuando estudiamos a una araña estudiamos también la telaraña que va dejando. No estudiamos a los organismos como agentes solitarios sino siempre acompañados de su medio ambiente, con la cognición pasa lo mismo: para estudiar los procesos cognitivos necesitamos estudiar no solo los procesos corporales sino también las herramientas externas, ya que solo cuando ponemos juntos a estos elementos es que constituyen un sistema cognitivo.

Pero además, Menary defiende que la relación entre medio ambiente y organismo no es una en donde solo el medio ambiente modifica al organismo sino una en donde también éste modifica al medio ambiente. Esta relación dual es una en donde ambas partes interactúan y cada una modifica a la otra, no se trata solo de que un organismo reciba pasivamente los cambios que el medio ambiente dispone, sino que también de que el comportamiento de los organismos impacta directamente en el medio ambiente. Esta idea es influencia de la propuesta de los nichos cognitivos⁵⁰ sostenida por Sterelny (2004), la tesis es que los agentes no permanecen pasivos ante los cambios del medio ambiente sino que también los agentes modifican su entorno, razón por la cual la influencia es simétrica.

El externismo de la mente sostiene que *algunos* elementos del medio ambiente determinan el contenido de los estados mentales. La tesis de la cognición integrada es externista, pero es un tipo de externismo en donde los elementos internos influyen en los elementos externos y viceversa; esta relación en donde ambas partes se influyen mutuamente motivará a que en la cognición integrada la cognición no solo sea estudiada como un sistema, sino además como un sistema dinámico.

3.1.2 Externismo dinámico

⁴⁹ Citado por Richard Menary (2007, p. 108): “[t]he distinction between what is spatially ‘inside’ and what is spatially ‘outside’ the bird, as such, has no significance for the study of the avian biological system (pág. 159)”.

⁵⁰ La idea de *construcción de nichos* fue introducida en la biología evolutiva por Lewontin (1982, 1983). Laland y O’Brien definen a la construcción de nichos como “[...] the process whereby organisms, through their activities and choices, modify their own and each other’s niches. By transforming natural-selection pressures, niche construction generates feedback in evolution at various different levels” (2012, p. 1).

Cuando estudiamos a la cognición como un sistema cognitivo, tenemos diferentes tipos de relaciones entre los componentes de dicho sistema: funcionales, causales y estructurales. Otra característica esencial de dicho sistema es que su integración tiene que ser dinámica, ya que estudiarla con este punto de vista nos permite darle la misma importancia a los componentes sin importar si son elementos internos o externos. Pero, ¿qué significa que la cognición tiene que ser estudiada como un sistema dinámico?

Decir que la cognición es descrita de una mejor manera usando la teoría de los sistemas dinámicos es decir que los científicos cognitivos deben de tratar de entender la cognición como un comportamiento inteligente y modelar el comportamiento inteligente usando un tipo particular de matemáticas, usualmente conjuntos de ecuaciones diferenciales⁵¹ (Chemero 2009, p. 25).

Cuando proponemos que un sistema cognitivo puede incluir partes del medio ambiente, necesitamos una herramienta que nos permita explicar los límites de esta extensión, y esta herramienta es la teoría de los sistemas dinámicos (Chemero 2009). Además, desde una postura externista no podemos estudiar separadamente unos componentes de otros, sino que se tienen que estudiar de manera conjunta ya que solo cuando logran integrarse es cuando se completan las tareas cognitivas. Por ello, Beer (2003) nos dice que desde una perspectiva dinámica:

El sistema nervioso, el cuerpo y el entorno de un agente son cada uno conceptualizados como sistemas dinámicos que están en constante interacción. Desde este punto de vista, el comportamiento de un agente surge de la interacción entre estos sub-sistemas y no puede ser atribuido propiamente a cualquier componente en aislamiento de los otros⁵² (p. 211).

Cuando estudiamos a la cognición como un proceso dinámico lo que importa es la interacción de los diferentes componentes del sistema cognitivo; por este motivo, otra de las ventajas de este punto de vista es que los diferentes componentes de dicho sistema pueden estar localizados fuera del sistema nervioso del agente ya que será la coordinación entre ambos lo que nos dará el comportamiento total

⁵¹ “To say that cognition is best described using dynamical systems theory is to say that cognitive scientists ought to try to understand cognition as intelligent behavior and to model intelligent behavior using a particular sort of mathematics, most often sets of differential equations” (Chemero 2009, p. 25).

⁵² “The nervous system, body and environment of an agent are each conceptualized as dynamical systems that are in constant interaction. From this viewpoint, an agent’s behavior arises from this interaction between these subsystems and cannot properly be attributed to any one component in isolation from the others” (Beer 2003, p. 211).

del sistema. La interacción entre el cerebro, el cuerpo y el medio ambiente conforman un sistema dinámico, que a su vez forman parte un sistema más grande: el sistema organismo-medio ambiente. El dinamismo del sistema nos ayuda a entender esta integración que Menary llama *bio-causal*.

3.1.3 El principio de complementariedad

Una de las diferencias entre la cognición integrada y la tesis de la mente extendida tradicional es su nula relación con el *principio de paridad*⁵³. Sutton (2010) sostiene que mientras que Clark y Chalmers (1998) postulan el *principio de paridad*, Menary (2007) y Kiverstein & Farina (2011) sostienen el *principio de complementariedad*, el cual nos dice que:

En sistemas cognitivos extendidos, los estados externos y procesos no necesitan imitar o replicar los formatos, dinamismo o funciones de los estados y procesos internos. Más bien, componentes diferentes del sistema total pueden jugar (permanente o temporal) roles absolutamente diferentes y tienen diferentes propiedades mientras colaboran en contribuciones colectivas y complementarias al acto y pensamiento flexible⁵⁴ (p. 194).

Al rechazar el *principio de paridad*, la cognición integrada no se compromete con la tesis de que los elementos externos e internos (y sus procesos) tengan que ser funcionalmente equivalentes, todo lo contrario: lo importante es que los vehículos externos desempeñan diferentes roles y será su coordinación —y acoplamiento— con los vehículos internos lo que constituya un sistema cognitivo.

Nótese que rechazar el *principio de paridad* es parte de la estrategia para responder a las críticas presentadas en el capítulo 2 (sección 2.2.3); la estrategia no consiste en defender dicho principio sino más bien abandonarlo y postular uno nuevo. Al rechazar dicho principio, la cognición integrada se deshace de la tesis funcionalista que subyace en la propuesta de C&C y evita muchas de sus críticas. Además, Menary (2007) señala que el *principio de paridad* adolece de dos factores importantes: 1) el impacto que los recursos externos tienen en nuestras capacidades cognitivas y 2) la normatividad de la manipulación de estos recursos externos en nuestras prácticas cognitivas; es

⁵³ En el capítulo 2 hemos visto de qué manera Clark y Chalmers (1998) presentan el *principio de paridad* y algunas de las críticas que se le han hecho a dicho principio.

⁵⁴ “In extended cognitive systems, external states and processes need not mimic or replicate the formats, dynamics or functions of inner states and processes. Rather, different components of the overall (enduring or temporary) system can play quite different roles and have different properties while coupling in collective and complementary contributions to flexible thinking and acting” (Sutton, 2010. p. 194).

por ello que, además de postular el *principio de complementaridad*, la nueva versión de mente extendida postula la tesis de la manipulación. Veamos en qué consiste dicha tesis.

3.1.4 La tesis de la manipulación

La tesis de la manipulación nos dice que un artefacto es parte de la cognición no porque cumpla un rol causal definido (puede cumplir varios o muchos, o cambiar su rol a lo largo del tiempo) sino porque está integrado dentro del carácter de los agentes. El tener la habilidad de manipular representaciones externas (para completar tareas cognitivas) es lo que Menary llama prácticas cognitivas, la importancia de la tesis de la manipulación de vehículos externos es un pre-requisito para que surjan las habilidades y también para completar tareas cognitivas avanzadas:

Los integracionistas toman el tipo de procesos acoplado que C&C identifican como evidencia de que ahí hay una relación importante entre un individuo y su entorno y entonces coloca estas interacciones en un contexto más grande. La gente aprende cómo manipular representaciones externas con plumas y papel y aprende a leer y escribir, hacer listas, planes, notas, diagramas, etc., y aprende a hacer estas cosas para propósitos particulares. Usualmente la habilidad de manipular herramientas externas y representaciones permite resolver problemas de una manera que no podríamos sin ellas⁵⁵ (Menary 2007, p. 60).

Ya hemos visto que, para Clark y Chalmers, un sistema causal ensamblado [*causal coupling*] se da cuando se unen un agente con un artefacto. Menary llama a este tipo de integración *extensión artefactual* [*artefact extension*], pero él propone un tipo diferente de integración: la *cognición enculturada* [*enculturated cognition*]. En este tipo de integración los artefactos en sí mismos no extienden nuestras habilidades cognitivas sino que más bien es nuestra adquisición de *prácticas cognitivas* —para la manipulación de los artefactos— lo que lo que nos convierte en agentes más

⁵⁵ “The integrationist takes the kind of coupled processes that C&C identify as evidence of there being an important relationship between an individual and her environment and then places these interactions in a wider context. People learn how to manipulate external representations with pens and papers and learn to read and write, make lists, plans, notes, diagrams and so on and they learn to do these things for particular purposes. Often the ability to manipulate external tools and representations allows us to solve problems in a way that we would not without them” (Menary 2007, p. 60).

poderosos. La idea de la tesis de la manipulación es tomada del trabajo de Rowlands (1999), sin embargo Menary (2007) distingue entre diferentes tipos de manipulación de elementos externos⁵⁶.

Al tener en cuenta estos diferentes tipos de manipulación nos damos cuenta que la cognición no tiene que ser estudiada no solo como un sistema dinámico, sino que además debemos tomar en cuenta el aprendizaje de habilidades, pues el uso de herramientas y elementos externos al sujeto implica necesariamente la noción de aprendizaje y habilidades. Poniendo estas nuevas ideas juntas, poco a poco vamos teniendo nuevas estrategias para responder a las críticas hechas a la tesis de la mente extendida.

3.1.5 Respuesta a la falacia constitución-acoplamiento

La tesis de la manipulación es el hilo central de la propuesta filosófica de la cognición integrada, ya que postula que se debe estudiar a la mente como algo híbrido; además, la postulación de prácticas cognitivas depende ampliamente de la noción de manipulación, pues sin ellas es imposible que estas prácticas cognitivas culturales se den en los agentes y además puedan ser heredadas. Otro punto importante es que un reporte causal (como el de la tesis tradicional de la mente extendida) no explica por qué la manipulación de artefactos externos es parte de la cognición, sino que solo muestra que se dan estos casos de cognición extendida. Sin embargo, se necesita dar un reporte de cómo es que al resolver un problema matemático, por ejemplo, interactúan el cerebro, el cuerpo y el medio ambiente.

Mientras que el reporte causal nos muestra que hay casos de cognición extendida, la cognición integrada nos explica por qué y cómo es que se dan estos casos. Para la cognición integrada las prácticas cognitivas para la manipulación de información guardada en artefactos (por ejemplo agendas, recuerde el caso de Otto) están integradas en nuestro *carácter cognitivo*. Otra diferencias entre la tesis de la mente extendida tradicional y la cognición integrada es que, mientras que para la primera los artefactos son parte de nuestro sistema cognitivo porque estamos conectados

⁵⁶ Menary distingue entre cuatro principales tipos de manipulación: **i) acoplamiento biológico**, la idea en este tipo de manipulación es que corporalmente manipulamos e influimos en el medio ambiente para nuestro beneficio; un ejemplo de este tipo de manipulación es el caso de los fenotipos extendidos. **ii) Acciones epistémicas**, este tipo de manipulación se da cuando en lugar de utilizar una representación interna la representamos en el mundo, tal y como hacemos cuando jugamos al *tetris*. **iii) Acciones auto-correctivas**, en este tipo de manipulación encontramos al lenguaje y soportes externos para estructurar acciones prácticas. Finalmente, **iv) las prácticas cognitivas** son el tipo de manipulación que se da cuando utilizamos representaciones externas para completar algunas tareas cognitivas, un ejemplo de este tipo se da cuando resolvemos una multiplicación con ayuda de una pluma y un papel.

causalmente con ellos, para la cognición integrada los artefactos no conforman parte de la mente sino más bien lo que es parte de nosotros son las habilidades para manipular a dichos artefactos. Y es justo esta diferencia la que le permite a Menary responder a la pregunta ¿cómo puede mi conexión causal con mi iPhone ser parte de mi mente? Si recordamos bien, para los internistas el límite natural de la cognición es el cerebro:

Ellos defiende que si quieres estudiar la cognición y la mente, entonces tienes que estudiar los fenómenos cognitivos y mentales implementados en el cerebro, y solo esos fenómenos implementados en el cerebro pueden contar como miembros de una clase cognitiva o mental⁵⁷ (2010, p. 229).

Sin embargo, los procesos cognitivos están constituidos en parte por la manipulación de estructuras físicas y corporales que se encuentran en el entorno de los agentes; de este modo la relación entre vehículos y agente no puede ser meramente una relación causal: “una manipulación de un vehículo externo va mas allá de una simple relación causal porque la manipulación necesita ser descrita en términos de normas y usualmente de contenido”⁵⁸ (Menary 2007, p. 50). Pero al aceptar que esta integración va más allá de una simple relación causal, una de las primeras preguntas que surgen es dónde trazar el límite entre lo que constituye parte de la mente y lo que no. O en otras palabras, cuando completamos una tarea, ¿qué herramientas son parte de la mente y cuáles no?

La manera de responder a esta pregunta es que lo que extiende la cognición no son las herramientas sino nuestras habilidades cognitivas; de acuerdo con Menary lo que forma parte de la mente (o del carácter cognitivo del agente) no son los artefactos sino nuestras habilidades de manipular dichos artefactos, por ello en el caso de Otto su libreta de notas no es parte de su mente sino más bien de su carácter cognitivo dadas sus habilidades para manipularla. El iPhone no es parte de mi mente, por supuesto que gracias a sus aplicaciones puedo completar tareas complejas, pero si lo pierdo no pierdo parte de mi mente (recuerde la crítica de Jerry Fodor) ya que puedo utilizar el iPhone de alguien más. Los artefactos tienen diferentes habilidades y nosotros las manipulamos para

⁵⁷ “They argue that if you want to study cognition and the mind, then you need to study the cognitive and mental phenomena implemented in the brain, and only those phenomena implemented in the brain can count as members of a cognitive or mental kind” (2010, p. 229).

⁵⁸ “[a] manipulation of an external vehicles goes beyond a simply causal relation because the manipulation needs to be described in terms of norms, and very often, content” (Menary 2007, p. 50).

nuestros fines pero esas son propiedades de los artefactos no de los agentes, lo que reside en nosotros es nuestra habilidad para manejarlos y justo esto es lo que nos convierte en agentes cognitivos más poderosos.

Así, lo que extiende nuestra cognición es más bien nuestra capacidad para utilizar artefactos, eso es lo que constituye nuestro carácter cognitivo, razón por lo cual la tesis de la manipulación desempeña el punto de partida de todo el proyecto que pretende entender a la cognición como un proceso integrado. El caso de Otto que vimos en el capítulo 2 nos muestra cómo algunos estados mentales pueden ser extendidos, si a esta manera de extender la cognición le agregamos el punto de vista de la cognición integrada entonces tenemos una manera de explicar esos estados cognitivos que ahora nosotros llamaremos híbridos. Una característica de estos estados es que para que resulten tiene que darse la integración de los elementos internos y externos:

Para los integracionistas cognitivos la imagen es como esta: mi manipulación de la agenda y mis procesos cerebrales juntos constituyen un proceso de recordar. En casos como este, los procesos de recordar no pueden ser descritos exclusivamente en términos de memoria biológica o solamente en términos de la manipulación de representaciones externas, porque es un proceso híbrido⁵⁹ (Menary 2007, p. 63).

Siguiendo este enfoque, al estudiar la cognición debemos tener en cuenta varias cosas: la manera en que los elementos externos e internos se integran, su naturaleza, la manipulación de los elementos externos y el uso de habilidades cognitivas. De este modo, no sólo tenemos una relación causal entre X (en donde X es la manipulación de la libreta de Otto) y Y (en donde Y constituye los procesos neuronales), sino que esta integración de X y Y nos da el estado Z, que en este caso Z constituye el proceso de recordar.

En la **Figura 3** podemos ver cómo el color verde representa los elementos externos (estos son los artefactos, como la libreta de Otto), el color rojo son los elementos internos (los cuales constituyen los procesos neuronales); mientras que el color amarillo (que sólo se produce mediante la integración del rojo y el verde) es lo que constituye el proceso de recordar. Así, mientras que la

⁵⁹ “For the cognitive integrationist the picture is like this: my manipulation of the notebook and my brain processes together constitute a process of remembering. In cases like these, the process of remembering cannot be described exclusively in terms of biological memory or solely in terms of manipulation of external representations, because it is a hybrid process” (Menary 2007, p. 63).

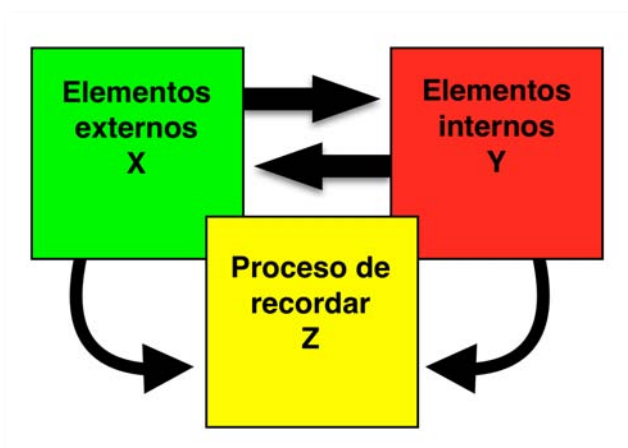
tesis de la mente extendida muestra casos de estados mentales extendidos, la cognición integrada explica cómo es que se dan estos estados y en qué sentido es que son híbridos. Además, la cognición integrada se centra en explicar cómo los vehículos externos modifican y transforman nuestras capacidades cognitivas; de este modo, no va en contra de otros tipos de externismo sino que simplemente mira los mismos fenómenos desde otras perspectivas.

Es importante recalcar que uno de los objetivos de Menary es explicar la relación entre los componentes externos e internos, y es en esta relación en donde podemos encontrar procesos cognitivos híbridos. Los sistemas cognitivos integrados se dan cuando se logra la coordinación de los elementos internos con los elementos externos, solo la coordinación de ambas manipulaciones (la interna y externa) es lo que permite a los agentes completar sus tareas cognitivas.

A pesar de los esfuerzos de Menary por dar un mejor reporte del externismo, su propuesta no está exenta de problemas. Una de las primeras críticas es que las habilidades cognitivas y nuestra capacidad para manipular artefactos, a pesar de que se originen en nuestra interacción con los artefactos, una vez que están en nuestro carácter cognitivo pueden ser nuevamente algo interno del sujeto. Para rechazar estas críticas, Menary tendría que rechazar la distinción sujeto-objeto que hemos visto en el capítulo 1 y así (a la manera en que [Heidegger \(1927\)](#) y [Merleau-Ponty \(1945\)](#) señalaron que somos seres mundo y que el mundo está siempre ahí) defender que el hombre está siempre en el mundo y que en este estar en el mundo su mente está siempre extendida.

Finalmente, otro de los problemas es que a pesar de ser externista el enfoque de Menary sigue siendo representacionalista, por este motivo presentaré otras dos respuestas para mostrar que responder a la falacia constitución-acoplamiento puede hacerse desde una postura anti-representacionalista y también desde el enactivismo.

Figura 3. Aquí queda representado de qué manera se da el proceso de recordar (color amarillo). El color amarillo solo surge cuando se dan cuando interactúan el color verde y el color rojo.



3.2 Sistemas cognitivos-fenomenológicos extendidos

Mientras que podemos distinguir entre posturas internistas y externistas de la cognición, también podemos distinguir entre diferentes versiones de externismo. Podemos ser externistas con un reporte enactivista de la cognición o ser externistas pero manteniendo algunas ideas tradicionales. Chemero, en *Radical Embodied Cognitive Science* (2009), defiende casos de cognición extendida pero rechaza algunas de las ideas de Clark: la idea de que la cognición sigue siendo representacionalista y que ésta es dada en términos de contenido informacional. Además de estar influenciado por la fenomenología y la robótica, su postura está también fuertemente influenciada por la psicología ecológica de Gibson (1979).

Chemero define su anti-representacionalismo como el estudio científico de la percepción, la cognición y la acción, a los cuales entiende como fenómenos encarnados en el mundo. Además, toda su propuesta se apoya en la teoría de los sistemas dinámicos, razón por la que nos habla de sistemas cognitivos-fenomenológicos extendidos:

Los sistemas cognitivos extendidos están compuestos heterogéneamente de cerebro, cuerpo y nicho, acoplados no-linealmente unos a los otros. [...] Lo que llamamos “sistemas cognitivos-fenomenológicos extendidos” son entidades relacionales y dinámicas, con interacciones entre partes heterogéneas en múltiples escalas espaciales y temporales⁶⁰ (Silberstein y Chemero 2012, p. 35).

Dado que los componentes de los sistemas cognitivos extendidos están acoplados de manera no lineal, desde esta perspectiva dinámica el cuerpo interactúa con el sistema nervioso en dos direcciones. Pero al mismo tiempo el cuerpo está también interactuando también con el entorno en dos direcciones, véase la **Figura 4**.

Otra de las diferencias entre la tesis tradicional de la mente extendida y el externismo de Chemero es que aquí se parte de la idea de que la cognición es siempre extendida y se da por la simbiosis entre cuerpo, cerebro y mundo. Mientras que en la tesis tradicional se parte de la idea de que la cognición es algo interno y algunas veces llega a ser extendida, en el caso de Chemero se parte del lado contrario: la cognición está siempre extendida hacia el mundo.

⁶⁰ Extended cognitive systems are heterogeneous composed of brain, body and niche, non-linearly coupled to one another. [...] What we call “extended phenomenological-cognitive systems” are relational and dynamical entities, with interactions among heterogeneous parts at multiple spatial and temporal scales (Silberstein y Chemero 2012, p. 35).

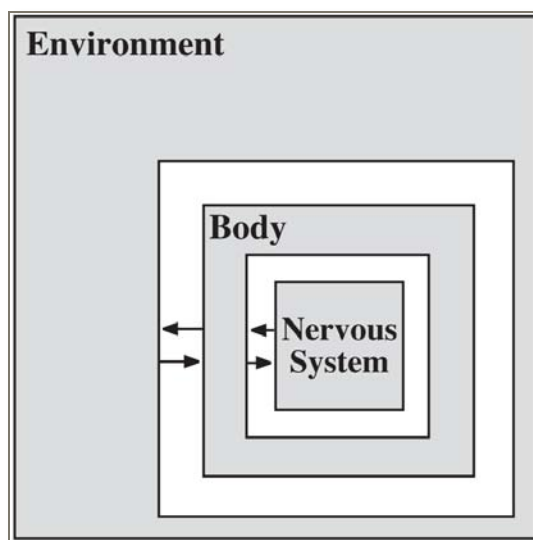


Figura 4. Esta imagen presentada en Beer (2003, p. 211) nos muestra cómo se mira la cognición desde los sistemas dinámicos: el sistema nervioso interactúa con el cuerpo y este a su vez interactúa con el entorno.

3.2.1 Cómo trabajar sin representaciones

Chemero apela al trabajo de Beer (2003) para mostrar cómo su propuesta es compatible con trabajos llevados a cabo en el terreno de la inteligencia artificial. Beer, utilizando sistemas dinámicos, muestra cómo se lleva a cabo el comportamiento de un robot simulado que es controlado por una red neuronal artificial.

El agente propuesto por Beer no utiliza representaciones y sin embargo tiene cierto tipo de percepción. Además, otras características de este modelo es que puede predecir el comportamiento del agente porque: 1) el modelo muestra cómo el sistema —como un todo— evoluciona sobre el tiempo, 2) muestra cómo la relación entre el agente y los objetos cambia sobre el tiempo y 3) muestra cómo el sistema nervioso del agente reacciona a los objetos.

Con este modelo se sabe con certeza el comportamiento del agente pero además permite predecir cómo se comportará; aunado a esto, una enseñanza final de este modelo es que el *conocimiento* del agente no reside en su sistema nervioso, pues para reconocer objetos el agente necesita moverse e interactuar, razón por la cual podemos ver por qué es importante tener en cuenta todos los componentes de un sistema cognitivo:

El cuerpo físico de un agente, la estructura de su entorno y su contexto social pueden jugar un rol tan importante en la obtención de su comportamiento tanto como su cerebro. De hecho, en un sentido muy

real, la cognición ya no puede ser vista como limitada a la cabeza de un agente, sino que puede ser distribuida a través de un grupo de agentes y artefactos⁶¹ (Beer 2003, p. 209).

Mirando este ejemplo concreto podemos notar por qué, al estudiar un sistema cognitivo, no podemos descomponerlo en partes, pues para explicar cómo el agente adquiere conocimiento tenemos que apelar a la interacción entre el cerebro del agente, su cuerpo y el medio ambiente. Por sí solo este caso ya constituye una respuesta a la *falacia constitución-acoplamiento*, pues muestra a través de una noción epistemológica tan básica, a saber el conocimiento, que esta no puede ser reducida a un fenómeno que sucede en la cabeza de un agente. El conocimiento emerge solo cuando el agente empieza a interactuar con su mundo, luego entonces queda patente por qué la cognición tiene que ser extendida. En el siguiente apartado veremos otro ejemplo para mostrar que la cognición es un proceso no lineal, y con ello dejar de lado las visiones tradicionales que miran a la cognición como un proceso lineal estático.

3.2.2 Argumento para los sistemas dinámicos

Desde la robótica evolutiva también se han dado argumentos para mostrar el punto de Chemero: que la *falacia constitución-acoplamiento* solo tiene sentido cuando miramos a la cognición desde una perspectiva tradicional; además, dicha falacia surge porque hay un entendimiento erróneo del papel que los diferentes componentes de un sistema juegan al momento de interactuar no linealmente en la teoría de los sistemas dinámicos.

Froese, Gershenson y Rosenblueth (2013) en *The Dynamically Extended Mind: A Minimal Modeling Case Study* presentan un modelo de dos agentes acoplados acústicamente y muestran que algunas propiedades emergen solo cuando dichos agentes están en un sistema dinámico no lineal. El planteamiento de este objetivo parte con el reto de construir el modelo de un agente encarnado [*embodied agent*] con un sistema nervioso artificial; este sistema nervioso artificial ANS (por sus siglas en inglés) tiene propiedades matemáticas que es imposible que tenga cuando se encuentra aislado, por lo tanto, se tiene que mostrar que cuando el agente esta situado dentro de un *loop*

⁶¹ “An agent’s physical body, the structure of its environment, and its social context can play as important a role in the generation of its behavior as its brain. Indeed, in a very real sense, cognition can no longer be seen as limited to an agent’s head, but can be distributed across a group of agents and artifacts” (Beer 2003, p. 209).

sensor-motor el ANS se transforma en un sistema acoplado diferente, la estrategia utilizada aquí es que para medir esta transformación se utiliza el caos.

Una vez reconstruido el argumento de Froese, Gershenson y Rosenblueth, queda de la siguiente manera:

i) El teorema de *Poincaré-Bendixon* muestra que para que un sistema dinámico presente actividad caótica, dicho sistema debe tener al menos tres dimensiones.

ii) Por lo tanto, si un ANS con menos de 3 dimensiones está acoplado con otro sistema no caótico y su actividad neuronal interna de pronto presenta comportamiento caótico, entonces no hemos cometido la *falacia constitución-acoplamiento*.

iii) El modelo de dos agentes acústicos embebidos y acoplados es un ejemplo de un ANS descrito en la premisa ii) y muestra comportamiento caótico.

∴ El caos solo puede ser explicado como una propiedad emergente que aparece solo cuando interactúan el ANS con otro sistema no caótico.

Cuando estudiamos a los sistemas separadamente no registramos actividad caótica, pues el caos emerge solo cuando los dos sistemas están acoplados. Este ejemplo muestra cómo la cognición extendida es, en principio viable, y pone en entredicho las críticas filosóficas hechas por [Adams y Aizawa \(2001, 2010\)](#). Si el caos solo emerge cuando los dos sistemas están en interacción, ¿no podría pasar lo mismo con la cognición? Lo interesante del argumento de [Froese et al., \(2013\)](#) es que presenta, desde la teoría de los sistemas dinámicos, un argumento en favor de una tesis filosófica.

3.2.3 Respuesta a la falacia constitución-acoplamiento

Una vez que hemos mostrado que la cognición es un sistema dinámico no lineal, se deja entrevisto que la falacia solo adquiere sentido si miramos a la cognición como un sistema lineal en donde podemos separar a sus componentes y estudiar todas las propiedades que tiene. Este punto de vista es altamente cuestionable ya que, como hemos visto, algunas propiedades emergen solamente cuando los elementos de un sistema están en interacción:

Cuando un agente y su entorno están acoplados no-linealmente, ellos juntos, constituyen un sistema no desmontable, y cuando esto es el caso, la falacia constitución-acoplamiento no es una falacia. En otras

palabras, la falacia constitución-acoplamiento es solo una falacia cuando el acople es lineal⁶² (Chemero 2009, p 31-32).

La falacia solo tiene sentido cuando miramos a la cognición como un sistema computacional que procesa información y que su conexión con el mundo es representacional, pero cuando pensamos la cognición como un sistema dinámico la crítica deja de tener sentido porque no hay manera de separar al agente del mundo.

No podemos explicar los fenómenos cognitivos descomponiendo en partes a dicho sistema porque la cognición solo emerge cuando el agente y el mundo forman un sistema juntos: “[...] el agente y su entorno están acoplados no-linealmente, forman un sistema unificado y no desmontable, lo cual quiere decir que ellos forman un sistema cuyo comportamiento no puede ser modelado, incluso aproximado, como un conjunto de partes que van separados”⁶³ (Chemero 2009, p 32). La única manera en que la falacia tenga sentido es que todavía pensemos a la cognición como algo estable que puede ser modelado en un sistema lineal, pero no podemos apelar a este punto de vista porque es lo que estamos poniendo en duda. Como bien señala Froese:

[...] [los] críticos han defendido que la hipótesis de la mente extendida está basada en una falacia constitución-acoplamiento, es decir que cuando un proceso cognitivo en la cabeza está acoplado con una parte del entorno, entonces esa parte no es necesariamente ni típicamente constitutiva del proceso cognitivo, Sin embargo, esta crítica filosófica no es apoyada por la teoría de los sistemas dinámicos no lineales. [...] Un enfoque con sistemas dinámicos tiene, por lo tanto, el potencial para formalmente salvar la hipótesis de la mente extendida de sus críticas⁶⁴ (Froese et al., 2013, p. 7).

⁶² “When the agent and the environment are nonlinearly coupled, they, together, constitute a non decomposable system, and when that is the case, the coupling-constitution fallacy is not a fallacy. In other words, the coupling-constitution fallacy is only a fallacy when the coupling is linear” (Chemero 2009, p 31-32).

⁶³ “[...] the agent and environment are nonlinearly coupled, they form a unified, nondecomposable system, which is to say that they form a system whose behavior cannot be modeled, even approximately, as a set go separate parts” (Chemero 2009, p 32).

⁶⁴ “[...] critics have argued that the extended mind hypothesis is based on a coupling-constitution fallacy, i.e. that when a cognitive process in the head is coupled with a part of the environment, that part is neither necessarily nor typically constitutive of the cognitive process. However, this philosophical critique is not supported by nonlinear dynamical systems theory. [...] A dynamical system approach thus has the potential to formally save the extended mind hypothesis from its critics” (Froese et al 2013, p. 7).

La *falacia constitución-acoplamiento* parte de una visión ya obsoleta y anticuada de la cognición. Trabajos como el de Beer (1995, 2003), Van Orden, Holden y Turvey (2003) y Froese et al., (2013), muestran que entender la cognición extendida como un sistema dinámico no solo queda en un nivel filosófico sino que dicho planteamiento puede llevarse (y de hecho se está llevando) a cabo en la inteligencia artificial y la robótica.

Finalmente, a pesar de estos ejemplos concretos alguien podría querer preguntar todavía dónde trazamos los límites entre lo que es parte de la mente y lo que no, tal pregunta surge nuevamente de la idea de que podemos tener casos de cognición aislados del mundo, o que hay tal cosa como un cerebro desnudo; sin embargo, dado que el enfoque dinámico “[...] enfatiza la dimensión temporal de la cognición, y la manera en que el comportamiento del agente surge de la interacción continua entre su cerebro, su cuerpo y su entorno”⁶⁵ (Beer 2003 p. 209-210), podemos responder que para trazar los límites bastaría hacer un análisis temporal de la cognición y ver cuándo y en qué momentos algo es parte de la cognición y cuándo no.

Tal análisis temporal implicaría demasiadas complicaciones prácticas pero en principio no es imposible de llevar a cabo. Es por simple hábito que en la vida cotidiana separamos al mundo de nosotros, pero en sentido estricto la cognición está siempre extendida pues nuestra capacidad de actuar en el mundo se da solo gracias a la interacción de nuestro cerebro con el cuerpo, que a su vez está interactuando con el mundo. De este modo, podemos ver cómo desde punto de vista llegamos de manera paralela, a algunas de las ideas de Heidegger (1927): que somos seres anclados en el mundo.

3.3 Enactivismo: cognición sin representaciones

El enactivismo aparece por primera vez en el año de 1991 en la obra de Varela, Thompson y Rosch: *The Embodied Mind*. En este libro ellos proponen que la cognición no puede estar dissociada del cuerpo y que ésta emerge de la interacción entre los organismos y su medio ambiente. De este modo el enactivismo parte de la idea de que la actividad encarnada de los seres vivos nos brinda el mejor modelo para entender la mente. De acuerdo con Thompson (2007), el término *enactivismo* fue pensado por Varela y estuvo inspirado en los versos del poeta español Antonio Machado: *caminante*

⁶⁵ “[...] emphasize the temporal dimension of cognition, and the way in which an agent’s behavior arises from the ongoing interaction between its brain, its body and its environment” (Beer 2003 p. 209-210)

no hay camino, se hace camino al andar. Por ello es que el enactivismo sostiene que no hay cognición si no hay acción:

Nosotros proponemos *enactivo* como un nombre para enfatizar la creciente convicción de que la cognición no es la representación de un mundo pre-dado por una mente pre-dada sino más bien el decreto de un mundo y una mente sobre la base de una historia de la variedad de acciones que un ser en el mundo lleva a cabo⁶⁶ (Varela et al, 1991, p. 9).

Para el enactivismo la cognición surge por la percepción y la acción a través de los patrones sensor-motrices; además, como bien señala Menary (2006): desde el enactivismo para que un agente sea cognitivo tiene que tener un cuerpo, y será la interacción entre el cuerpo de los organismos y el mundo lo que de origen a la cognición. De acuerdo con Thompson (2006, p. 13), el enactivismo presentado en 1991 engloba diferentes ideas que podemos resumir en varios puntos: 1) los seres vivos son agentes autónomos que activamente se mantienen a sí mismos, 2) el sistema nervioso es un sistema dinámico autónomo. 3) La cognición es llevar a cabo habilidades en acciones situadas y encarnadas en el mundo, 4) el mundo de un ser cognitivo no está pre-representado internamente en el cerebro y 5) la experiencia debe de ser investigada desde un punto de vista fenomenológico.

Uno de los puntos compartidos por estos diferentes tipos de enactivismo es el punto 4: el rechazo de representaciones internas para explicar la cognición; en los últimos años el enactivismo ha sido defendido por Noë (2004), Thompson (2007) y Hutto & Myin (2013). Dado que el enactivismo radical propone una manera de responder a la *falacia constitución-acoplamiento*, me centraré en explicar exclusivamente este tipo de enactivismo y la respuesta que propone a la crítica hecha a la tesis de la mente extendida.

3.3.1 El enactivismo radical: cognición sin contenido mental

Que esta corriente de la cognición rechace el uso de representaciones es asumido por las diferentes tipos de enactivismo que podemos encontrar, pero una de estas posturas llevadas a sus últimas consecuencias es defendida por Hutto y Myin (2013). Ellos no solo rechazan el uso de representaciones sino que dan un paso más allá: se deshacen del contenido mental. La propuesta de

⁶⁶ “We propose as a name *enactive* to emphasize the growing conviction that cognition is not the representation of a pre-given world by a pre-given mind but is rather the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs” (Varela et al, 1991, p. 9).

Hutto y Myin es explicar la cognición básica y afirman que en este tipo de fenómenos mentales no hay contenido; es decir, cuando interactuamos con el mundo lo único que hay es acción y percepción. El enactivismo radical niega que para que haya interacción inteligente entre el mundo y la mente sea necesario el contenido mental.

La idea de mentes sin contenido parece descabellada ya que nuestra idea de mente, por definición, implica la idea de contenido. Pese a esto, para Hutto y Myin la cognición básica es sin contenido y las mentes básicas están ya involucradas en el mundo. Al rechazar el contenido mental la *falacia constitución-acoplamiento* no tiene cabida porque la manera en que se entiende la cognición no comparte ningún presupuesto internista: ni el uso de representaciones ni tampoco el contenido mental.

3.3.2 Mentes extensivas: respuestas a la *falacia constitución-acoplamiento*

Además de proponer desecharnos de las representaciones y del contenido mental, Hutto y Myin presentan una nueva manera de entender la tesis de la mente extendida: tenemos que entender las mentes no como *extendidas* sino como *extensivas*. El enactivismo radical parte señalando que en el debate en torno a la tesis de la mente extendida tanto las posturas en contra como las que están a favor comparten ciertas ideas: que la cognición (sea esta interna o extendida) tiene que ser entendida como el procesamiento de representaciones y de contenido informacional.

Como hemos visto, la crítica principal a la tesis de la mente extendida es la *falacia constitución-acoplamiento*; pero esta solo es una falacia si seguimos manteniendo un marco en donde la cognición sigue siendo con representaciones. Cuando entendemos la mente no como extendida (pues este punto de vista señala que solo a veces la cognición se extiende al mundo) sino como extensiva, entonces no tenemos que preocuparnos por dar una respuesta a la falacia constitución-acoplamiento pero “[...] sin la cognición radical enactivista no hay manera para los defensores de la hipótesis de la mente extendida motivar su posición decisivamente y silenciar las objeciones internistas de una vez por todas⁶⁷ (Hutto y Myin 2013, p. 151).

Como podemos notar, este tipo de enactivismo no es radical solo respecto al enactivismo sino también con respecto a la tesis de la mente extendida. Partiendo de la postura de Hutto y Myin nos vemos abocados a radicalizar la idea de mente extendida: las mentes tienen que ser extensivas, y

⁶⁷ “[...] without REC [radical enactivism cognition] there is no way for defenders of EMH [extended mind hypothesis] to motivate their position decisively and to silence internalist objections once and for all” (Hutto y Myin 2013, p. 151).

al hacer esto no hay necesidad de apelar ni al *principio de paridad* o al *principio de complementaridad*. Para Hutto y Myin en la tesis de la mente extendida todavía se le da demasiado terreno al internista: se parte de la idea de que la cognición está casi siempre limitada y solo en algunos casos la mente se extiende al mundo; lo que necesitamos hacer es partir del lado contrario: que nuestras mentes son extensivas hacia el mundo y que los procesos cognitivos están constituidos por andamiajes [*scaffoldings*].

Además, si rechazamos el contenido mental, entonces no tiene caso hablar de la distinción entre vehículos y contenidos⁶⁸; la distinción entre lo que es parte de la mente y lo que es necesario causalmente para la cognición deja de tener sentido.⁶⁹ Al entender la mente como extensiva se rechaza la suposición, por default, internista de la mente: que los procesos cognitivos son algo que sucede dentro de nuestras cabezas y que sólo en raras ocasiones pueden ser extendidos. Desde el punto de vista del enactivismo radical las críticas hechas a la tesis de [Clark y Chalmers \(1998\)](#) tampoco pueden ser respondidas desde la propuesta de [Menary \(2007\)](#), la única salida es ser radical. Y al igual que [Chemero \(2009\)](#) el enactivismo propone que para entender la cognición: “[...] es necesario apreciar como los seres vivos interactúan dinámicamente con sus entornos: finalmente, no hay posibilidad de entender las mentes sin referencia a la interacción entre los organismos y sus entornos”⁷⁰ ([Hutto y Myin 2013, p. 4](#)).

3.4 Conclusiones del capítulo

La cognición integrada ([Menary 2007](#)) toma el enfoque externista de la mente extendida tradicional pero modificando varias de sus tesis centrales: se deshace del *principio de paridad* y, contrario a lo que pudiera parecer a primera vista, esto en lugar de restarle poder explicativo le permite expandir sus horizontes en el tratamiento de los estados mentales y los procesos cognitivos. Una vez que se deshace de la tesis funcionalista, Menary puede dar cuenta de las críticas hechas a la tesis de Clark y Chalmers, su propuesta parte estipulando la *tesis de la manipulación* como el eje central de la

⁶⁸ De acuerdo con Hutto y Myin si no podemos hablar de la marca de lo cognitivo, o si entendemos la mente como envueltas por el mundo entonces una de las herramientas metodológicas que podemos utilizar son las *capacidades*.

⁶⁹ Recuerde la crítica de A&A del capítulo pasado acerca de la marca de lo mental, o del criterio para delimitar lo que es cognitivo y lo que no lo es.

⁷⁰ “[...] it is necessary to appreciate how living beings dynamically interact with their environments: ultimately, there is no prospect of understanding minds without reference to interactions between organisms and their environments” ([Hutto & Myin 2013, p. 4](#)).

cognición integrada; además, sostiene que la cognición debe ser estudiada como un sistema dinámico.

Algunos de los problemas con esta propuesta es que sigue siendo representacionista y, además, no explica cómo podemos naturalizar los criterios normativos de la tesis de la manipulación. Otra crítica importante es que las habilidades cognitivas y nuestra capacidad para manipular artefactos, a pesar de que se originen en nuestra interacción con los artefactos, una vez que están en nuestro carácter cognitivo pueden ser nuevamente algo interno del sujeto.

La segunda respuesta que hemos visto es la postura de [Chemero \(2009\)](#); la cual, desde un punto de vista dinámico, defiende casos de cognición extendida y muestra que la *falacia constitución-acoplamiento* solo tiene sentido cuando miramos a la cognición como un proceso lineal. Un punto a tener en consideración es que con la propuesta de Chemero dejamos de hablar de mente extendida y hablamos más bien de cognición extendida, además su enfoque es anti-representacionista.

La tercera respuesta que hemos visto es una que permite disolver la falacia, desde el punto de vista de [Hutto y Myin \(2013\)](#) necesitamos entender a las mentes no como extendidas sino como extensivas. Además, al deshacernos de las representaciones y del contenido mental no hay tal cosa como la falacia, lo que hay es un entendimiento de que nuestras mentes están todo el tiempo enraizadas en el mundo. Una vez que hemos respondido a la *falacia constitución-acoplamiento* veamos qué implicaciones hay para nuestra noción de agencia si la miramos desde un punto de vista que propone que la cognición es extendida y sobre todo dinámica.

Capítulo 4

El problema de la agencia: ¿cómo diseñar *fungus eaters*?

“Si tuvieras que construir un agente inteligente, ¿dónde empezarías? ¿qué te impresionaría como algo especial que separa el mundo irreflexivo de las rocas, las cascadas y volcanes del reino de la inteligencia sensible? ¿qué es lo que permite a algunas partes del orden natural sobrevivir percibiendo y actuando mientras el resto permanece al margen, libre de pensamientos e inerte?”⁷¹ (1997, p. xii).

Andy Clark - *Being There*

Introducción

En el año de 1962 Masanao Toda, en *The design of a fungus-eater: A model of human behavior in an unsophisticated environment*, presentó el reto de construir robots totalmente autónomos: los *fungus eaters*. Su planteamiento del problema, aunque sencillo, involucra un sinnúmero de complicaciones no solo empíricas sino también teóricas y metodológicas: si existiera un planeta con únicamente dos recursos (uranio y hongos) y nosotros quisiéramos enviar a los *fungus eaters* a recolectar uranio (véase la **Figura 5**), ¿cuáles serían las propiedades esenciales de estos robots para que pudieran sobrevivir (comiendo hongos) y lograran recolectar el uranio? Sin duda, algunas propiedades que deberían de tener es la autonomía (para decidir cómo y cuándo interactuar con el medio ambiente), la autosuficiencia (porque adquieren energía comiendo hongos) y sobre todo, la adaptabilidad.⁷²

Independientemente de que tal proyecto sea viable o no, podemos plantear una pregunta hipotética: si lográramos diseñar un *fungus eater* y cumpliera con todas las cláusulas prescritas, ¿tendríamos que decir que son agentes en sentido estricto? Al tratar de responder a esta pregunta nos veremos orillados a antes dar respuestas a otras que tampoco resultan obvias: ¿son estas propiedades (la autonomía, la autosuficiencia y la capacidad de adaptarse al medio) solo humanas?, ¿es necesaria

⁷¹ “If you had to build an intelligent agent, where would you begin? What strikes you as the special something that separates the unthinking world of rocks, waterfalls, and volcanos from the realms of responsive intelligence? What is it that allows some parts of the natural order to survive by perceiving and acting while the rest stay on the sidelines, thought-free and inert?” (Clark 1997, p. xii).

⁷² Véase Ashby (1952/1960) para una explicación a fondo de la adaptabilidad y de cómo se puede ser implementada en agentes artificiales.

la inteligencia para adscribir agencia?, ¿bastarían estas propiedades para que éstos robots logran completar la tarea de recolectar hongos? O aún mejor: ¿podemos trasladar estas propiedades a los agentes artificiales?, ¿podemos distinguir entre agentes y agentes inteligentes?, ¿compartimos estas propiedades con otros animales?, ¿la agencia emerge de estas propiedades o falta algo más?

En gran medida nuestras respuestas dependerán de lo que entendemos por agencia, para Schlosser (2015), por ejemplo, un agente es un ser con la capacidad de actuar, por lo tanto la agencia no sería otra cosa que manifestar esta capacidad. De manera general dicha definición tiene sentido, pero cuando comparamos la acción de un robot con la de un humano nuestro reporte de ambas acciones será distinto y uno será más detallado que el otro. Piense, por ejemplo, en un robot recolectando basura y un pintor dibujando un paisaje. Aunque en ambos casos se hace patente la capacidad de ejecutar acciones, parece que hay diferencias significativas entre una acción y otra. Así, aunque en primera instancia el tema de la agencia parece no ser problemático, al momento que queremos adscribir agencia surgen algunas complicaciones.



Figura 5. Un “fungus eater” en un planeta distante, ilustración de Isabelle Follath.⁷³

El tema de este capítulo estará destinado a tratar de responder a algunas de las preguntas señaladas anteriormente, pero dada la complejidad del tema de la agencia lo descompondré en un par de problemas: i) *el problema metodológico de la agencia* y ii) *el problema ontológico de la agencia*. Mientras que i) tiene que ver con la creencia de que la agencia no es término controversial, ii) es el

⁷³ Esta imagen aparece en Pfeifer (1996), aquí la reproduzco aquí con el permiso de Isabelle Follath.

reto de responder qué es la agencia. A estos dos problemas podemos añadir un tercero que aparece cuando a ii) le añadimos las nuevas tesis externistas de la cognición, pues si la cognición es un fenómeno que se da cuando interactúan el cerebro, el cuerpo y el mundo ¿cómo deberíamos de entender la agencia? Llamaré a este tercer problema iii) *el problema externista de la agencia*.

Mientras que [Giere \(2012\)](#) defiende que, a pesar de que la cognición puede ser extendida o distribuida, los estados cognitivos solo pueden ser asignados a los componentes humanos de un sistema cognitivo. [Knappett y Malafouris \(2008\)](#) defienden que debemos abandonar el punto de vista que sugiere que la agencia es solo un atributo de los seres humanos, de tal modo que debemos entenderla como un proceso material que emerge cuando nuestros cerebros interactúan con el aspecto material del mundo. A este problema de decidir cómo debemos entender la agencia es a lo que he llamado iii) *el problema externista de la agencia*. Así, para llegar a la resolución de este problema seguiré el plan de trabajo descrito a continuación:

En la primera parte de este capítulo presentaré un reporte de cómo es tratada la agencia en los diferentes campos de estudio de las ciencias cognitivas. Con esto daré una respuesta al problema i), mientras que al mismo tiempo se hará visible por qué es importante responder al problema ii). Además, esta sección nos ayudará a entender por qué resulta tan problemático hablar de agencia, por qué el problema de la agencia es importante y sobre todo nos ayudará a entender de qué hablamos cuando hablamos de agencia (**sección 4.1**). Teniendo ya algunas ideas de la agencia presentaré un esquema tripartito que nos permita clasificar las diferentes acepciones de la agencia presentadas en la sección pasada. El marco tripartito será una respuesta al problema ii) (**sección 4.2**). Posteriormente, presentaré algunas aplicaciones del marco tripartito y con ello quedará patente cómo dicho marco nos da varias ventajas metodológicas (**sección 4.3**).

Una vez que tenemos una noción de la agencia entraré de lleno a iii) el problema externista de la agencia. Aquí expondré un par de argumentos, que partiendo de un punto de vista externista de la cognición, llegan a conclusiones diferentes de cómo debe ser entendida la agencia: 1) que la agencia es solo una propiedad humana y no puede ser ni distribuida ni extendida ([Giere 2006, 2012](#)). 2) Que la agencia es un proceso emergente que se da entre la interacción del cuerpo, el medio ambiente y el cerebro, por lo tanto no puede ser una propiedad interna humana sino un fenómeno emergente ([Malafouris 2008, 2013](#)) (**sección 4.4**). En el siguiente apartado llevaré a cabo una crítica

al argumento de Giere (**sección 4.5**), para posteriormente exponer algunas de las implicaciones del argumento de Malafouris a nuestra idea de agencia (**sección 4.6**).

4.1 El problema metodológico: usos, distinciones y aclaraciones de la agencia

El i) problema metodológico de la agencia consiste en creer que dicho término no es un concepto problemático y que, a diferencia de los conceptos de mente, racionalidad e intencionalidad, es incontrovertible (Barandiaran et al., 2009). Esta sección estará destinada a presentar la relevancia de la agencia y con ello dar una respuesta a i), pero además servirá para mostrar algunas de las complicaciones en torno al problema ontológico. Así, con el fin de dar una respuesta a i) y presentar ii) veamos cómo se entiende la agencia en los diferentes campos de las ciencias cognitivas y la filosofía.

4.1.1 La agencia en la filosofía

Gallagher y Zahavi (2008) señalan que ser el causante de algo no basta para atribuir agencia; un huracán, ejemplifican Gallagher y Zahavi, puede causar que nos quedemos sin luz, pero no por ello le atribuiríamos agencia. Por ello, dentro de la filosofía, una manera común de pensar la agencia consiste en asociarla a las intenciones y la acción, de tal modo que cuando hablamos de agencia hablamos casi siempre de *acciones intencionales*.⁷⁴ Pero tampoco las puras intenciones bastan, es necesario tener un cierto grado de consciencia, este no tiene que ser de un nivel alto sino de tener una pre-reflexiva consciencia de nuestras acciones (Gallagher y Zahavi 2008, p. 158). Así, cuando atribuimos agencia estamos hablando de algo más que causar acciones, la adscripción de agencia no se reduce a un recuento causal de las acciones, sino que involucra un sinnúmero de características:

Si le atribuimos agencia a una persona por romper una ventana, no solo describimos a la persona como causalmente responsable por el rompimiento de la ventana, también la describimos como teniendo ciertas concepciones de sí mismo como el agente del rompimiento de la ventana, como teniendo ciertas actitudes prácticas con respecto al rompimiento de la ventana y como teniendo

⁷⁴ El término “intencionalidad” resulta confuso porque dentro de la filosofía tiene dos diferentes significados, i) la capacidad que tiene la mente para estar dirigida a algo o para representar, decimos que los estados tienen *intencionalidad* cuando están dirigidos a algo (véase Searle 1983; Dennett 1987 y Brentano 1995). Mientras que ii) cuando hablamos de *acciones intencionales* nos estamos refiriendo a la capacidad que tenemos para tener intenciones y además para llevarlas a cabo. Además, tenemos que distinguir entre el término *intensional*, el cual es un concepto semántico que es contrapuesto a *extensional*, en inglés podemos distinguir entre “*intentionality*” e “*intensionality*”. Para una explicación a fondo véase Crane (2001) y Jacob (2014).

ciertas experiencias prácticas en la ejecución de la acción del rompimiento de la ventana (sentido de agencia, sensación de control)⁷⁵ (Grünbaum 2010, p. 337).

Por ello, para que yo sea un agente debo de *ser consciente* de que soy quien está llevando a cabo dicha acción, pero ¿basta el hecho de tener acciones intencionales y la capacidad de sentirnos agentes para hablar de agencia? Grünbaum nos señala otro aspecto importante de la agencia: la capacidad de tener movimiento corporal o control motor. Los mecanismos que subyacen a nuestra capacidad para tener movimiento corporal o control motor son una de las muchas cosas que obviamos al hacer estudios sobre la cognición. Por ejemplo, cerrar los ojos y saber que tengo dos manos es un hecho tan básico en nuestra vida cotidiana que lo damos por sentado, no obstante, cuando perdemos este sentido de *pertenencia corporal* surgen trastornos como los presentados por Sacks (1985). En la pérdida del sentido de pertenencia corporal el desorden neuropsicológico que surge es conocido como *somatoparaphrenia*⁷⁶ y consiste en no reconocer una parte de nuestro cuerpo como nuestra; por ejemplo, el caso de somatoparaphrenia que presenta Sacks se trata de un paciente que no sentía que una de sus piernas fueran de él.

El conocimiento de pertenencia corporal está anclado directamente al problema de la agencia y en una primera revisión parecen ser el mismo fenómeno; sin embargo, son dos cosas distintas. De hecho, para tener auto-conciencia dos de los requisitos son saber que sé es un agente y además saber que el cuerpo que tenemos es de nosotros (Synofzik, Vosgerau y Newen 2008). Así, una cosa es saber que soy el sujeto del movimiento y otra distinta es saber que yo soy la causa del movimiento, estos dos fenómenos se han confundido tanto teórica como experimentalmente.

Gallagher (2000, 2007), al embarcarse en un estudio fenomenológico de la acción y de la experiencia, traza una distinción entre *el sentido de pertenencia* y *el sentido de agencia*. Mientras que el primer sentido es la pre-reflexiva experiencia de que *yo soy el sujeto del movimiento*, el segundo es la pre-reflexiva experiencia de sentir que *yo soy la causa o el autor del movimiento*⁷⁷. Algo a notar en el sentido de agencia es nuevamente su conexión con la consciencia, por ello este

⁷⁵ “If we attribute agency to a person for the breaking of a window, we not only describe the person as causally responsible for the breaking of the window, we also describe her as having certain conceptions of herself as the agent of the breaking of the window, as having certain practical attitudes with regards to the breaking of the window, and as having certain practical experiences in the execution of the action of breaking the window (sense of agency, feeling of control)” (Grünbaum 2010, p. 337).

⁷⁶ Para otros desordenes neuropsicológicos relacionados véase Balconi y Bartolotti (2010).

⁷⁷ Para un análisis detallado del sentido de agencia véase Gallagher (2000) y Balconi (2010a).

sentido ha sido definido como “[...] el sentido de que yo soy quien está causando o generando una acción [...]”⁷⁸ (Balconi 2010c, p. v). Así por ejemplo, a la pregunta ¿cómo sé que soy yo el que está llenando la taza con café? La respuesta es gracias a mi sentido de agencia.

Gallagher y Zahavi (2008) proveen evidencia tomada de la psiquiatría para mostrar cómo en algunos casos de esquizofrenia la distinción entre estos dos sentidos se hace presente:

En síntomas esquizofrénicos de delirio de control o inserción de pensamientos, el sentido de pertenencia es retenido de alguna forma, pero el sentido de agencia está ausente. El esquizofrénico que sufre de estos delirios afirmará que su cuerpo se está moviendo pero que alguien más está causando el movimiento; o que hay pensamientos en su mente, pero que alguien más los está poniendo ahí⁷⁹ (p. 160).

Como podemos notar, en los casos mostrados por Gallagher y Zahavi (a diferencia del caso de Sacks en donde el paciente pierde el sentido de pertenencia) lo que se pierde es el sentido de la agencia; por ello, para estos filósofos lo que constituye la acción son estos dos sentidos (a los cuales consideran experiencias de primer orden no-conceptuales) que están integrados [embedded] y sobre todo que están ligados a la estructura temporal de la consciencia.

Una vez que distinguimos entre estos dos sentidos, Gallagher (2007) afirma que podemos distinguir entre tres tipos del sentido de agencia⁸⁰. Nuevamente la utilidad de dicha división es que nos permite explicar otros desordenes mentales, pues cuando alguno de los elementos que constituyen el sentido de la agencia falla, surgen problemas como la esquizofrenia o el trastorno obsesivo-compulsivo⁸¹. Finalmente, dado que el sentido de agencia solo puede surgir cuando se da la comunicación entre las neuronas sensoriales (neuronas aferentes), las neuronas eectoras

⁷⁸ “The sense that I am the one who is causing or generating an action” (Balconi 2010c, p. v).

⁷⁹ “In schizophrenic symptoms of delusions of control or thought insertion, the sense of ownership is retained in some form, but the sense of agency is missing. The schizophrenic who suffers from these delusions will claim that his body is moving but that someone else is causing the movement; or that there are thoughts in his mind, but that someone else is putting them there” (Gallagher y Zahavi 2008, p. 160).

⁸⁰ Para Gallagher (2007) los tres tipos del sentido de agencia son: 1) el sentido de agencia de primer orden ligado a la intencionalidad, 2) el sentido de agencia de primer orden ligado al movimiento corporal y 3) finalmente el sentido de agencia de segundo orden, este tipo de agencia está ligada a la atribución reflexiva.

⁸¹ Para más casos de esquizofrenia y el sentido de agencia véase Synofzik & Voss (2010) y Balconi & Bortolotti (2010).

(neuronas eferentes)⁸² y la retroalimentación corporal dada por la intencionalidad; [Gallagher \(2007\)](#) concluye que la agencia no puede ser entendida sin la *intencionalidad*, pues es esta la que hace que nuestra atención este dirigida hacia el mundo y por ello no puede ser solo un fenómeno que sucede dentro de nuestras cabezas.

Teniendo esta distinción en mente, veamos cómo se ha tratado el sentido de agencia en los últimos años dentro del campo de las neurociencias. Además, notemos que no podemos dissociar el problema de la agencia del problema de la acción y cómo esto nos lleva a otros temas como lo son el problema de la percepción, la auto-consciencia, la temporalidad y la intencionalidad.

4.1.2 La agencia en las neurociencias

Un buen ejemplo de cómo proceder de una manera completamente naturalista en el estudio de la cognición lo constituye las neurociencias, en este campo la investigación consiste en buscar cómo las neuronas están directamente relacionadas con procesos cognitivos precisos. Por ejemplo, las funciones del habla están asociadas al hemisferio izquierdo del cerebro, de tal modo que cualquier daño en el lóbulo frontal repercute en la capacidad que tenemos para hablar. De este modo, si queremos explicar un proceso cognitivo lo asociamos ineludiblemente a ciertas neuronas, el caso de la agencia no es la excepción.

Los intentos en las neurociencias para explicar la acción se han enfocando en el sentido de agencia ya que este sentido no solo tiene implicaciones directas para la acción sino también para el tema de la consciencia y la auto-consciencia. Algunas investigaciones como las de [Chaminade & Decety \(2002\)](#) y [Farrer & Frith \(2001, 2002\)](#) han llevado diferentes experimentos con el fin de encontrar el correlato neuronal del sentido de la agencia y mostrar que esta se origina en los procesos neuronales que son los responsables de los aspectos motores de la acción. El común denominador en estos trabajos es que utilizan la distinción (entre el sentido de agencia y el sentido de pertenencia) hecha por [Gallagher \(2000\)](#). Sin embargo, el problema es que dejan abiertas un sinnúmero de cuestiones sobre cómo pensar la idea de agencia; por ejemplo, no queda claro si debemos entender la agencia como asociada al control motor y corporal o más bien asociarla a la intencionalidad y a las acciones intencionales ([Gallagher 2007, p. 349](#)).

⁸² Las neuronas *aférentes* son las neuronas receptoras y se comunican mediante señales electromagnéticas con las neuronas *eferentes*, pues son éstas últimas quienes transmiten los impulsos eléctricos más allá del sistema nervioso, por ejemplo, a los músculos.

Otros trabajos recientes como el de [Synofzik et al., \(2008\)](#) y [Balconi \(2010b\)](#) arguyen, al igual que [Gallagher \(2007\)](#), que deben distinguirse tres niveles en el sentido de agencia. Ellos proponen que el sentido de agencia puede descomponerse en el nivel de la sensación [the feeling of agency], del juicio [the judgment of agency] y de la interacción social [the sense of a moral responsibility].⁸³ De este modo, si lo que queremos es hacer un reporte de la acción, no solo tenemos que distinguir entre los sentidos de pertenencia y de agencia, sino que dentro de ellos se tienen que distinguir otros niveles ya que una cosa es sentir (no conceptualmente) que somos el agente, otra cosa diferente es saber que somos los agentes (y al mismo tiempo distinguir entre mis acciones y las acciones de alguien más) y finalmente, otra cosa es tener acciones intencionales.

El propósito de [Synofzik et al., \(2008\)](#) no es solo proveer una explicación de la experiencia de la agencia y del sentido de pertenencia sino también ofrecer una taxonomía que pueda ser comprobable para las ciencias cognitivas y las neurociencias. Por ello, al proponer los diferentes niveles dentro de la agencia y la pertenencia, postulan también cuáles pueden ser los procesos neuronales que subyacen dentro de estos dos sentidos.

Con un poco de esfuerzo podemos trazar un paralelismo entre los tres niveles del sentido de agencia y los tres diferentes tipos de agencia postulados por [Gallagher](#). De entrada, tal comparación es viable y podemos encontrar algunos puntos en común como lo son el hecho de que, en ambos casos, el sentido de agencia está ligado desórdenes mentales. Pero uno de los problemas con el marco propuesto por [Synofzik](#) y el de [Gallagher](#) es que dejan abiertas diferentes interpretaciones de lo que puede ser, por ejemplo, una experiencia no conceptual⁸⁴. En el siguiente apartado veremos cómo es definida la agencia, sobre todo en la inteligencia artificial y la robótica.

4.1.3 La agencia en la inteligencia artificial, la robótica y la vida artificial

Cuando nos trasladamos a los campos de la inteligencia artificial, la robótica y la vida artificial, encontramos que la idea de agencia varía un poco. Dado que el diseño y construcción de agentes autónomos inteligentes es uno de los principales objetivos en estos campos, antes de iniciar el

⁸³ Para un exposición detallada de estos tres niveles en el sentido de agencia véase [Synofzik et al., \(2008\)](#) y [Balconi \(2010b\)](#).

⁸⁴ No existe consenso acerca de si el contenido de la experiencia está ya permeado por la racionalidad, es decir no hay seguridad de que todas nuestras acciones (incluidas las acciones no reflexivas) involucren ya el uso de nuestras capacidades conceptuales. Véase el debate entre [Dreyfus \(2013\)](#) y [McDowell \(2013\)](#) para examinar los argumentos en contra y favor, para una introducción a este debate véase [Scheur \(2013\)](#) y [Hernandez \(2015\)](#).

diseño de robots se parte de una idea de lo que tiene que ser un agente. Por ello, dentro de estas áreas, uno de los términos a los cuales es asociada la agencia (o lo que constituye un agente) es la autonomía.

Por ejemplo, [Steels \(1995\)](#) afirma que un agente debe ser capaz no solo de sentir su entorno, sino también de poder actuar. Además, esta capacidad tiene que darse mediante mecanismos automáticos del agente; es decir, para tener robots completamente autónomos estos no tienen que necesitar la intervención de otro agente. La idea de autonomía prescrita en estas cláusulas normativas implica no solo poder actuar en un entorno sino que se debe de actuar en pro del agente y con miras a lograr los objetivos. Por otro lado, [Barandiaran et al., \(2009\)](#) lleva el punto de vista enactivo a la definición de la agencia; por ello define a un agente como “una organización autónoma capaz de regular adaptativamente su acoplamiento con el ambiente de acuerdo a las normas establecidas por sus propias condiciones de viabilidad”⁸⁵ (p. 8)⁸⁶.

Esta definición enactivista permite que seamos incluyentes con muchos sistemas, por ello postulan tres condiciones para que podamos llamar agente a cualquier sistema que a) logre definir su propia individualidad, b) que esté haciendo algo por sí mismo en su medio ambiente y finalmente c) debe hacerlo mediante normas que regulen las actividades. Definir la propia individualidad significa que un sistema debe ser capaz de diferenciarse de su medio ambiente, esta cláusula nos ayuda a saber que estamos hablando de sistemas y no de cualquier proceso.⁸⁷ Por ello, para ser considerado un agente todo sistema tiene que ser persistente; en otras palabras, un sistema es un agente cuando es estable y duradero, por ello la noción de cohesión juega un papel importante: “un sistema debe de tener un grado de cohesión para llegar a ser estable y ser un punto de individualidad y agencia”⁸⁸ ([Barandiaran 2008, p. 93](#)).

Otro de los campos que une la idea de autonomía con de la agencia es la vida artificial, pero a diferencia de la robótica (en donde la autonomía es pensada como la capacidad que tiene un

⁸⁵ “an autonomous organization capable of adaptively regulating its coupling with the environment according to the norms established by its own viability conditions” (p. 8).

⁸⁶ Además, postulan tres condiciones para determinar cuando un sistema es un agente: la individualidad, la asimetría y la normatividad. Teniendo juntos estos tres requisitos podemos llamar agente a cualquier sistema que logre definir su propia individualidad, que esté haciendo algo por sí mismo en su medio ambiente y finalmente debe hacerlo mediante normas que regulen las actividades.

⁸⁷ Véase el capítulo 3 de [Barandiaran \(2008\)](#).

⁸⁸ “A system must have a degree of cohesion in order to become stable and a locus of individuality and agency” ([Barandiaran 2008, p. 93](#)).

sistema para poder interactuar con el mundo sin ayuda humana) aquí la autonomía es pensada como el proceso de auto-producción (Aguilar et al., 2014). El problema con definir la agencia en términos de autonomía es que la noción de agencia, lejos de resolverse, se vuelve circular pues tampoco tenemos una idea clara y esclarecedora de la autonomía. Froese, Virgo e Izquierdo (2007) muestran cómo en el campo de la vida artificial no hay un consenso acerca de la idea de autonomía.⁸⁹

En algunos otros casos, podemos notar cómo la idea de agencia está atada ineludiblemente a la cognición. En Harvey, Di Paolo, Wood y Quinn (2005) podemos ver cómo el concepto de agencia juega un rol central dentro del marco teórico. Harvey et al. (2005) proponen utilizar la robótica evolutiva⁹⁰ para estudiar modelos de cognición mínima y para ello definen a la cognición como:

[...] la capacidad de un agente de interactuar con su ambiente en tanto que conserva algunas viabilidades restrictivas. No es una propiedad interna del agente, sino una propiedad relacional que envuelve al agente, su medio ambiente y la conservación de algunas restricciones⁹¹ (p. 81).

Aquí podemos notar cómo la relación entre cognición y agencia se hace patente pues la agencia resulta indispensable para dar una explicación de la cognición. Además, una vez que Harvey y compañía inician con su investigación el término agencia se hace nuevamente presente en la primera de sus hipótesis: “un agente (humano, animal, robot, etc.) y el mundo en el que vive están hecho de materiales físicos que obedecen las leyes de la física, la química, etc.”⁹² (Harvey et al., 2005, p.82). Hasta aquí podemos empezar a notar un patrón en las diferentes áreas: para definir la cognición apelamos a la agencia, para definir la agencia apelamos a la cognición o a la autonomía.

Con lo que se ha presentado en este apartado basta para responder al problema metodológico de la agencia ya que, a pesar de las diferentes definiciones (e ideas) de la agencia planteadas en esta sección, queda un punto indiscutible: la necesidad de la agencia en el estudio de la cognición y el rol

⁸⁹ Para otro enfoque naturalista de la autonomía aplicada a los agentes véase Barandiaran (2008).

⁹⁰ De acuerdo con Harvey et al., (2005) el término robótica evolutiva [evolutionary robotics] se utiliza para llevar (de manera artificial) el punto de vista darwinista de la evolución al diseño de robots y agentes artificiales (p. 79).

⁹¹ “[...] the capability of an agent of interacting with its environment so as to maintain some viability constraint. It is not an internal property of the agent, but a relational property that involves the agent, its environment, and the maintenance of some constraint” (Harvey et al., 2005, p. 81).

⁹² “An agent (human, animal, robot, etc.) and the world it lives in are made of physical materials obeying the laws of physics, chemistry and so on” (Harvey et al., 2005, p.82).

central que desempeña en los diferentes campos de investigación. Así, una vez que que i) ha sido resuelto, veamos cuál podría ser una manera de resolver ii) el problema ontológico de la agencia.

4.2 Un esquema tripartito de la agencia

Hasta ahora hemos analizado cuál es el papel que la agencia desempeña en diferentes disciplinas, además de ello hemos notado que tan central resulta este concepto. Pero con las diferentes concepciones que se tienen de ella surge lo que he llamado el problema ontológico de la agencia, dicho problema puede ser planteado en un par de preguntas: ¿qué es es la agencia?, ¿es una propiedad solo humana? Si respondemos estas preguntas podemos dar respuesta a otras preguntas de tipo normativo como lo son: ¿a qué entidades aplicamos agencia: a los humanos o también a los animales?, ¿podemos con nuestra noción de agencia incluir a los robots?, ¿tiene sentido hablar de agencia o solo de agentes?

Ya hemos notado que la agencia se relaciona de manera ineludible con el tema de la acción y las intenciones. Sin embargo, al tratar de emularla (como en el caso de la robótica evolutiva y la inteligencia artificial) necesitamos un concepto de agencia que nos permita incluir a los robots y a otros sistemas artificiales. Como hemos visto, [Schlosser \(2015\)](#) define a los agentes como seres con la capacidad de actuar, pero teniendo esta definición ¿cómo podríamos distinguir entre los diferentes tipo de agentes?

En esta sección propondré una división tripartita que nos permita encasillar las diferentes acepciones de la agencia. Una de las ventajas de esta división es que nos permitirá entender el por qué esos diferentes enfoques, aunque tienen diferentes definiciones, hablan del mismo fenómeno. Veamos que características de los agentes nos permitirían trazar esta división tripartita.

4.2.1 Acciones intencionales y emociones

[Di Paolo \(2010\)](#) afirma que los robots aun carecen de una propiedad que los aleja de ser agentes cognitivos animales, la agencia intencional. Es la carencia de intenciones lo que traza la línea entre los animales y los robots:

Un robot fallando en su ejecución no muestra signos de preocupación, la falla o el éxito no afecta su estructura de ninguna manera peligrosa, ni cualquier forma de preocupación acompaña sus acciones

simplemente por que el objetivo deseado no es deseado por el robot sino por el diseñador (Di Paolo 2010, p. 132)⁹³.

A pesar de que los robots pueden ser agentes, un agente natural (humano o animal) tiene la ventaja no solo de que tiene propósitos sino que al mismo tiempo los genera (Di Paolo 2010, p. 135). Aunque tuviéramos robots con cierto grado de inteligencia y de adaptabilidad estaría faltando lo que Di Paolo llama el significado del mundo.

A diferencia de los agentes artificiales el mundo tiene para nosotros un significado, razón por lo cual las acciones que llevamos a cabo tienen una carga emotiva. Cuando un animal o un humano falla en el intento por conseguir comida, hay una preocupación. Esta preocupación se traduce en el caso de los humanos en un miedo a la muerte, en el caso de los animales la necesidad de conseguir comida viene dada por un instinto de supervivencia. Por ello, con las intenciones y con las emociones podemos trazar una distinción entre agentes naturales (animales y humanos) y artificiales.

Una vez que hemos distinguido entre agentes naturales y artificiales necesitamos alguna otra característica que nos permite distinguir entre los animales humanos y otros animales. Esa característica será el *sentido de agencia*, pues es este sentido el que nos permite saber que somos nosotros quienes estamos llevando a cabo o causando una acción. Apelar al sentido de agencia nos permite distinguir entre agentes humanos y agentes animales porque, como hemos visto, cuando perdemos este sentido surgen diferentes trastornos psiquiátricos y neuropsicológicos, los cuales no se hacen presente en los animales. Así, mientras que podemos distinguir a los agentes naturales de los artificiales, también podemos distinguir (dentro de los agentes naturales) entre los humanos y los animales (véase la **Figura 6**).

Una vez que hemos mostrado como distinguir entre los diferentes tipos de agentes, cobra sentido el hacer una división tripartita: agencia humana, agencia animal y agencia artificial.

⁹³ “A robot failing in its performance does not show any signs of preoccupation — failure or success do not affect its structure in any dangerous way, nor does any form of concern accompany its actions simply because the desired goal is not desired by the robot but by the designer” (Di Paolo 2010, p. 132).

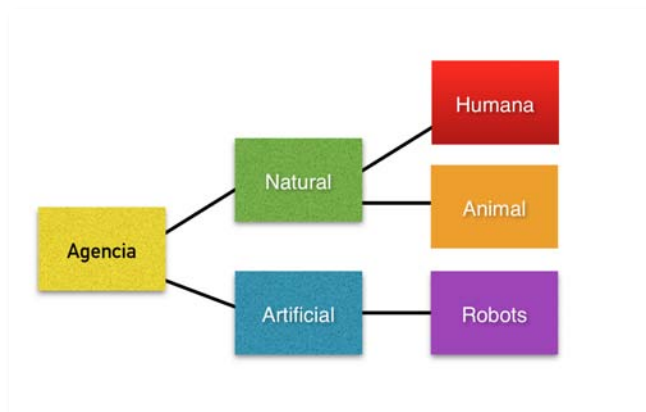


Figura 6. Esquema de la división tripartita de la agencia.

La necesidad de esta división se hace patente cuando nos damos cuenta de que mientras el enfoque enactivista de la agencia (Barandiaran et al., 2009) nos permite trasladar esta propiedad en el diseño de agentes artificiales, esta definición no nos sirve para analizar trastornos psiquiátricos y neurofisiológicos a los cuales nos enfrentamos los humanos. Por otro lado, dicha división nos permite generar una definición de agencia que pueda ajustarse solo al caso de los animales, y esta sería útil, por ejemplo, en el campo de la robótica evolutiva. Veamos algunas características principales de estos tres tipos de agencia.

4.2.2 Agencia humana

La agencia humana no se reduce a la capacidad de actuar y tener intenciones, la agencia humana involucra el sentido de pertenencia corporal y el sentido de agencia. Mientras que el concepto general de agencia puede reducirse y explicarse con la capacidad de actuar, en el caso de agencia humana esa capacidad de actuar involucra un sin fin de otras características como lo son la consciencia de objetivos, la intención de actuar, la iniciación de una acción, el sentido de actividad y de tener control mental y corporal, saber que somos nosotros los autores de la acción, etc. (Balconi 2010c).

La agencia humana involucra saber que somos nosotros quienes estamos llevando a cabo dichas acciones pero también implica el uso de deseos e intenciones. Además, podemos aventurar junto con Balconi (200c) que en el caso específico de los agentes humanos, la agencia debe ser pensada no solo como la capacidad de actuar sino también como tener la capacidad de generar pensamientos conscientes.

4.2.3 Agencia animal

Jonas (1966/2001) ha señalado que lo que distingue la vida animal de la vida vegetal es la movilidad, la percepción y las emociones. La principal característica de la agencia animal es que carece del sentido de agencia, pero comparte junto con los agentes humanos los puntos señalados por Jonas. En este tipo de agencia tenemos intenciones y por tanto podemos atribuir este tipo de agencia a los animales o acciones que podemos llevar a cabo sin ser consciente de que somos los agentes. Pero al comparar este tipo de agencia con la agencia artificial descubrimos que el grado de agencia de los agentes artificiales no puede ser compartido con el de los animales por la sencilla razón de que los animales sienten la amenaza de peligro. Además en el caso de los agentes humanos y animales las acciones son algo más que mero movimiento: involucran emociones.

4.2.4 Agencia artificial

En este tipo de agencia encasillaremos a los robots o cualquier simulación artificial de un organismo. A pesar de que los agentes de esta sección pueden llevar a cabo movimiento, éstos dependen completamente de otro agente. Aunque se puede argumentar que los agentes artificiales también tienen intenciones claras y determinadas, ya hemos visto que éstos vienen impuestas por los creadores de dichos agentes. Así, en la agencia artificial no tenemos ni emociones, ni el sentido de agencia y mucho menos el carácter subjetivo de la experiencia.

4.3 Aplicaciones del marco tripartito

4.3.1 Compatibilidad de los diferentes enfoques de la agencia

La división tripartita de la agencia nos permite resolver el problema acerca de las múltiples definiciones de la agencia que tenemos y el poco avance que permiten:

Muchas de las definiciones disponibles de la agencia o son muy flojas o imprecisas para permitir un progresivo programa científico. Implícitamente y problemáticamente asumen las características que caracterizan a los agentes, por lo tanto oscurecen el potencial completo y el desafío de modelar la agencia⁹⁴ (Barandiaran et al., 2009, p. 1).

⁹⁴ “[m]ost of the available definitions of agency are either too loose or unspecific to allow for a progressive scientific program. They implicitly and unproblematically assume the features that characterize agents, thus obscuring the full potential and challenge of modeling agency” (Barandiaran et al., 2009, p. 1).

Al tener un esquema en donde agrupar las principales definiciones útiles de la agencia estamos dando un marco desde el cual iniciar nuevos proyectos de investigación y saber bien desde que hipótesis estamos partiendo. No es que las diferentes definiciones sean incompatibles sino que apuntan a diferentes casos de agencia, partiendo del marco tripartito la definición de Gallagher queda entonces encasillada como *agencia natural humana*. Teniendo esto en mente nos queda claro por qué desde esta idea de la agencia podemos trabajar en la psicología, la fenomenología, la epistemología, la psiquiatría y las neurociencias, pues esta definición nos dice que los humanos somos agentes pero que además tenemos un *sentido de agencia* y que es precisamente la falta de este sentido lo que provoca desordenes mentales.

El marco tripartito nos da la ventaja de que lejos de hacer incompatibles las diferentes nociones de agencia (y por lo tanto descartar algunos enfoques), las agrupa y nos permite aprovechar sus diferentes ventajas metodológicas. Pero además, nos ayuda a entender qué problemas específicos están tratando los diferentes definiciones de la agencia. Por ejemplo, [Moreno y Etxeberria \(2005\)](#) dudan que la agencia humana pueda ser modelada, pese a ello [Weber & Varela \(2002\)](#) y [Froese & Ziemke \(2009\)](#) buscan cada vez más un aproximación a la agencia natural y de este modo, proveer una base biológica a los agentes artificiales. Sin embargo, una vez que hemos llevado a cabo la división de la agencia nos damos cuenta de que estos esfuerzos estarían aproximándose a la agencia animal más que a la agencia humana.

Por otro lado, cuando a través del marco tripartito analizamos el enfoque enactivista de la agencia ([Barandiaran et al., 2009](#)) nos damos cuenta de que ganamos un marco naturalista para poder desarrollar agentes artificiales, sin embargo perdemos los aspectos psicológicos y emocionales de la agencia natural. Por ello, una de las ventajas del marco tripartito es que hace compatible las diferentes definiciones de la agencia; mostrando con ello que analizar la agencia resulta tan complejo que es mejor descomponerla en diferentes problemas. Y será en estos diferentes planteamientos en donde se necesiten los distintos enfoques de la agencia.

4.3.2 Respuesta al problema normativo de la agencia

Anteriormente señalamos que una de los problemas de la agencia es saber a qué entidades atribuirles agencia, razón por lo cual [Barandiaran et al., \(2009\)](#) preguntan:

¿Es el robot Khepera un agente independientemente de su arquitectura del control o de su cuerpo, o solo en virtud de su capacidad de moverse alrededor la arena? ¿qué hay de un sistema protocelular bombeando iones fuera de su membrana? ¿un pájaro volando sobre las corrientes de aire es un agente? ¿Los temblores afectando a un paciente con el mal de parkinson cuentan como agencia? ¿cómo podemos justificar la negación o atribución de agencia sobre los casos de arriba en una manera responsable para el escrutinio científico?⁹⁵ (p. 1).

Con la división tripartita de la agencia podemos responder a las preguntas normativas porque, a pesar de que podemos aplicar el término agente a organismos humanos y artificiales, podemos seguir manteniendo diferencias significativas entre los agentes. El robot Khepera puede ser considerado un agente si es capaz de ejecutar acciones en el mundo independientemente de su capacidad de moverse, pues su capacidad de moverse puede venir dada por un agente externo.⁹⁶

Mientras que, en el caso de los pacientes con parkinson, podemos negar que los temblores indiquen agencia, pues al haber dividido la agencia en tres tipos diferentes tenemos una manera de hacer un reporte detallado y más elaborado de la agencia en humanos. Por otro lado, podemos por ejemplo afirmar que la agencia en humanos no solo consiste en la capacidad de actuar sino de generar ciertos pensamientos en nuestra consciencia (Balconi 2010c). Con esta definición de agencia incluimos como agentes a las personas con parálisis, algo que no podemos hacer cuando entendemos a la agencia en términos generales como la capacidad de actuar. Además, con la definición de Barandiaran et al., (2009) podemos considerar agentes a otros sistemas artificiales; así, el problema de saber a qué entidades aplicarles agencia se va volviendo menos complejo.

4.3.3 Construcción de robots

El esquema que he presentado nos permite mantener la idea de que tenemos diferentes tipos de agentes pero distinguiendo sus principales características entre ellos. Por ello, otras de las ventajas del marco tripartito es que resulta útil a la hora de diseñar robots, pues nos permite tener una idea

⁹⁵ “Is a Khepera robot an agent, independently of its control architecture or its body, just in virtue of its capacity to move around an arena? What about a protocellular system pumping ions outside its membrane? Is a bird gliding on wind currents an agent? Do the tremors affecting a Parkinson disease patient count as agency? How can we justify the negation or attribution of agency to the above cases in manner amenable to scientific scrutiny?” (Barandiaran et al., 2009, p. 1).

⁹⁶ Nicoletis y Chapin (2002) han llevado a cabo un experimento en donde un mono en un laboratorio es capaz de mover un brazo artificial localizado en el laboratorio de otro ciudad. Mientras que Hochberg et al., (2012) ha llevado a cabo experimentos en donde gente con parálisis es capaz de manipular un brazo artificial, todo con el objetivo de poder beber agua.

clara de qué comparaciones necesitamos hacer para diseñar determinados robots: si queremos emular la vida y robots que nos ayuden a entender cómo razonamos, seguramente necesitaremos estudiar los tres tipos de agencia. Mientras que si queremos robots que simplemente cumplan tareas básicas, bastará con mirar la agencia animal y la agencia artificial, pues como bien señala [Pfeifer \(1996\)](#):

Lo que consideramos que es un buen diseño de un agente autónomo o una teoría interesante y valiosa, depende de los objetivos que tenemos en mente. Si queremos construir robots que recojan basura estamos buscando algo muy diferente que el biólogo que está tratando de entender la evolución. O el computólogo que está interesado en el poder los algoritmos evolutivos hace algo diferente que el psicólogo tratando de entender la cognición⁹⁷ (p. 4).

Con esto podemos ver cómo el esquema tripartito nos permite tener en mente cuáles son los objetivos que buscan los diferentes campos de estudio, en dónde están y hacia donde se están encaminando. Además, nos permite tener objetivos claros y poner las debidas pautas en nuestros proyectos de investigación, pues si no hemos logrado simular la agencia de los insectos ¿cómo podríamos, entonces, querer modelar ya la agencia humana?

4.3.4 Futuras investigaciones

Otras de las ventajas del marco tripartito es que nos permite acomodar futuras investigaciones en torno a los diferentes aspectos de la agencia que apenas se están llevando a cabo. Si como mencionamos anteriormente, la agencia humana la definimos como la capacidad no solo de actuar en el mundo sino de generar pensamientos conscientes, en un futuro podremos dar un reporte de experimentos que se están llevando a cabo y que cambiaran nuestra manera de entender las acciones y los movimientos. Trabajos como los de [Nicolelis & Chapin \(2002\)](#), [Velliste et al., \(2008\)](#) y [Hochberg et al., \(2012\)](#) se han centrado en investigaciones que se enfocan en manipular objetos a través de los pensamientos.

⁹⁷ “What we consider to be a good design of an autonomous agent or an interesting and valuable theory, depends on the goals we have in mind. If we want to build robots that collect garbage we are looking for something very different than the biologist who is trying to understand evolution. Or the computer scientist who is interested in the power of evolutionary algorithms is after something else than the psychologist trying to understand cognition” ([Pfeifer 1996, p. 4](#)).

Si dichos proyectos resultan exitosos entonces nuestra noción de agencia humana tendría que ampliarse, por ello desde ahora es viable que la agencia en humanos sea entendida no solo como la capacidad de actuar sino también como la capacidad de generar pensamientos conscientes. Por otro lado, mientras que la mayoría de los filósofos cazan la idea de agencia con el hecho de tener acciones intencionales, poco se ha dicho del papel que en ella desempeñan las emociones. Si queremos analizar a fondo la agencia que hemos designado con el nombre de agencia natural humana, no podemos dejar de lado el rol que juegan las emociones en las acciones intencionales ya que muchas de nuestras intenciones (sino es que todas) viene determinada por nuestras emociones.

El marco tripartito que hasta aquí he presentado deja abierta un sin fin de cuestiones, pero dado que objetivo principal de mi investigación es enfocarme en el problema externista de la agencia, lo hasta aquí presentado sirve para entrar de lleno a este último problema.

4.4 El problema externista de la agencia

Algo que hemos remarcado desde el capítulo 1 es la idea de que no podemos separar el estudio del hombre del mundo, los últimos trabajos en inteligencia artificial y en robótica parten de un punto en común: la necesidad de estudiar a los organismos como anclados en el mundo. Así, llegados a este punto de la tesis, queda por preguntarse qué pasa con la agencia una vez que nos tomamos en serio los casos de cognición extendida. Dado que ya tenemos una noción más clara de la agencia, veamos qué pasa con ella una vez que tomamos como punto de partida entender los fenómenos mentales como un proceso extendido hacia el mundo.

Me concentraré en dos argumentos que partiendo de ésta idea llegan a conclusiones diferentes: a) a pesar de tener casos de cognición extendida la agencia sigue siendo una propiedad enteramente humana e interna, es decir ni es situado ni tampoco extendido y 2) la agencia es un fenómeno que emerge de la interacción entre el cerebro, el cuerpo y el mundo; o en otras palabras: la agencia lejos de ser solo una propiedad enteramente humana es un fenómeno emergente que surge cuando interactuamos con el mundo. Antes de continuar, es preciso remarcar que el tipo de agencia que analizaremos en este apartado es lo que hemos llamado, en la sección pasada, agencia humana.

4.4.1 Argumento internista antropomórfico

Giere (2006, 2012) tiene una tesis externista de la cognición, pero rechaza la idea de que tengamos que atribuir estados cognitivos a los sistemas cognitivos extendidos. Desde su punto de vista los componentes a quienes debemos atribuir estados cognitivos y agencia es solo a los componentes humanos. Para defender este punto presenta el caso de un sistema cognitivo que nos da una creencia U (acerca del origen del universo) donde queda patente que la atribución de esa creencia solo puede darse a los componentes humanos. Veamos su línea argumentativa:

- i) Tenemos la creencia U de que el universo se generó hace más de 14 mil millones de años.
- ii) Obtuvimos U gracias a un sistema cognitivo en donde tenemos a los científicos y al telescopio Hubble.
- iii) Mediante el telescopio Hubble los científicos obtuvieron la evidencia para U , dicha evidencia es información de galaxias que están a más de 2 mil millones de años luz.
- iv) Dado que sin el telescopio Hubble no hubiéramos podido obtener la información para la creencia de U , queda patente porque el telescopio jugó un papel importante en el sistema cognitivo que obtuvo la creencia U . Pero, ¿podemos atribuirle también la creencia U ? En otras palabras: ¿la creencia U está distribuida entre el telescopio Hubble, entre las galaxias de las cuales se extrajo la información y los científicos?
- v) La respuesta de Giere a las preguntas de iv) es un “no”.

Para Giere la atribución de estados cognitivos y de agencia solo puede dársele a los componentes humanos, en el caso del sistema cognitivo que generó la creencia U la adscripción de la creencia se le daría a los científicos. Para Giere este es un ejemplo claro para trazar una distinción entre un sistema cognitivo extendido *produciendo* una creencia y el sistema al cual le *atribuimos* la creencia.

Al igual que las preocupaciones de los internistas que vimos en el capítulo 2, Giere también se pregunta: “una vez que permitimos que un sistema cognitivo pueda incluir partes fuera de un cuerpo humano, ¿dónde nos detenemos?”⁹⁸ (2012, p. 201). Su respuesta es que cualquier cosa que impacta directamente en el comportamiento del sistema es parte de dicho sistema cognitivo, sin embargo, la atribución de estados cognitivos solo puede dársele a los componentes humanos.

⁹⁸ “[o]nce one allows that a cognitive system can include parts outside a human body, where does one stop?” (Giere 2012, p. 201).

Tomar el caso de la creencia U no es algo fortuito, pues en este ejemplo queda patente por qué parece erróneo atribuir una parte de la creencia al telescopio. Aún peor, cambiar nuestra noción de lo que significa conocimiento y atribuir la creencia a todo el sistema implicaría atribuir la creencia U a las galaxias (que están a millones de años luz) que dieron la información para llevar al conocimiento de dicha creencia. Ahora bien, en sentido estricto Giere no ha dado una razón de por qué no podemos atribuir la creencia U a todo el sistema, ya notamos que las implicaciones son raras y bizarras, pero aún así podríamos aceptar que dicha creencia pertenece a todo el sistema cognitivo que la originó, pues cosas más raras que estas se han aceptado y defendido en filosofía.

Para desmantelar este punto de vista Giere apela a la responsabilidad y para ello da un pequeño ejemplo: si reviso mi iPhone y me dice que mañana va a llover, entonces genero la creencia C de que mañana va a llover (note que esta creencia es producida por el sistema que mi iPhone y yo constituimos). Pero, si mañana resulta que no llueve, ¿de quién sería la culpa? Giere afirma que la culpa no sería del iPhone, sino nuestra. Cuando algo sale mal no queremos atribuir *responsabilidad moral* a los sistemas extendidos. Si tuviéramos que corregir los hechos científicos que originaron la creencia U , no podríamos atribuir responsabilidad al telescopio Hubble o a las galaxias que están a más de dos mil millones de años luz. La responsabilidad caería solo sobre los científicos, por ello es que, desde este punto de vista, queda patente por qué las creencias solo pueden ser atribuidas a los componentes humanos.

Aunque el argumento apela a un estado cognitivo (en este caso *creer*), dicho argumento sirve para defender que no podemos ni distribuir ni extender la agencia; la idea de fondo, claro está, es que [Giere \(2012\)](#) equipara la agencia como un estado cognitivo. Giere finaliza mostrando las dos caras de los casos de cognición extendida:

Entender los procesos cognitivos como usualmente distribuidos entre el cerebro humano y otros objetos naturales, o artefactos, parece un avance en ciencias cognitivas. Por el otro lado, atribuir estados cognitivos o agencia a sistemas extendidos amenaza con crear un hueco confuso entre las ciencias cognitivas y su objeto de estudio, la cognición humana⁹⁹ (p. 204).

⁹⁹ “Understanding cognitive processing as often distributed among human brains and other natural objects or artifacts seems an advance in cognitive science. On the other hand, imputing cognitive states or agency to extended systems threatens to create a confusing gap between cognitive science and its subject matter, human cognition” ([Giere 2012](#), p. 204).

La conclusión es que lo mejor es aceptar que aunque un sistema cognitivo extendido (o distribuido) es el generador de un estado cognitivo (en este caso el de la creencia U) solo le atribuimos tal estado a los componentes humanos, pues si tal creencia resulta errónea los culpables serían los científicos.

Algunas de las consecuencias del punto de vista de Giere es que entonces la agencia es una propiedad interna de los sujetos; así, no importa que tengamos casos de cognición extendida. A pesar de que los estados cognitivos puedan ser generados por elementos externos al sujeto, se sigue manteniendo una característica interna de los sujetos que no es posible extender: la agencia. El punto de vista de [Malafouris \(2008, 2013\)](#) irá totalmente en contra de esta tesis.

4.4.2 Argumento externista material

[Knappett y Malafouris \(2008\)](#) desde el campo de la arqueología proponen una nueva manera de entender la agencia. Ambos sostienen que la agencia es pensada desde un punto de vista totalmente antropocéntrico y que la tradición parte de una suposición que es generalmente aceptada: que la agencia es una propiedad enteramente humana y no solo eso, sino también una propiedad interna del sujeto. Por ello, el objetivo de [Malafouris \(2008, 2013\)](#) es defender que la agencia es un *enredo material* que surge en el espacio en donde se interceptan el cerebro, el cuerpo y la cultura.

Malafouris pretende cambiar lo que nosotros hemos llamado *agencia humana*, uniéndola a lo que él llama *agencia material*. Al hacer este cambio no se trata simplemente de cambiar de nombre a la agencia humana sino de entender de un modo distinto a este tipo de agencia, ni Knappett ni Malafouris defienden que la agencia humana es material, lo que defienden es que la agencia es un fenómeno que no puede ser considerado solo humano ya que la agencia está enraizada con la cultura. Dicho de otra manera: que la agencia humana no puede ser sólo humana debe entenderse como que este tipo de agencia no puede ser una propiedad o capacidad interna del sujeto sino un proceso que emerge solo cuando interactúan el cerebro, el cuerpo y el aspecto material del mundo. Por ello, la hipótesis de partida de Malafouris está expresada de la siguiente manera:

Si la agencia humana es entonces también la agencia material es, no hay manera de que la agencia humana y la agencia material puedan ser desenredadas. O de otro modo, mientras que la agencia y la intencionalidad pueden no ser propiedades de las cosas, no son propiedades de los humanos tampoco:

son las propiedades del involucramiento material, es decir, de la zona gris donde el cerebro, el cuerpo y la cultura se mezclan¹⁰⁰ (Malafouris, 2008 p. 55).

Notemos que, desde este punto de vista, si preguntamos en dónde está la agencia cuando realizamos una acción la respuesta será que no está ni en las cosas ni en nosotros mismos, pero si la agencia no están ni en el mundo ni en nosotros ¿entonces en dónde está? La respuesta de Malafouris es que la agencia emerge cuando estamos interactuando con en el mundo. Así, cuando Malafouris expresa que la agencia no es una propiedad humana lo que quiere decir es que la agencia no es una propiedad interna de los humanos; y por lo tanto, si no es una propiedad humana interna, entonces dicha capacidad emerge cuando interactuamos con el mundo. Con esto Malafouris desea dar un mayor papel al aspecto material del mundo:

El punto que estoy tratando de hacer aquí es que, a pesar de que el concepto de agencia es controversial, es controvertido dentro de los márgenes teóricos de una estrecha perspectiva antropocéntrica. Este punto de vista antropocéntrico de la agencia, aunque incorpora una variedad de matices, está basado en un acuerdo general sobre un solo hecho indisputable: que la agencia, en el real sentido de la palabra, es una propiedad humana, y que “los únicos verdaderos agentes en la historia son los individuos humanos”¹⁰¹ (Giddens and Pierson 1998, p. 89) (Malafouris 2013, p. 121-122).

Ya hemos visto que la agencia es la capacidad de actuar, pero específicamente en el caso de la agencia humana la asociamos a las intenciones, pues antes llevar a cabo una acción tenemos la intención de llevarla a cabo. Pero, dado que las intenciones también son estados mentales con intencionalidad (Jacob 2014), la agencia humana está ligada no solo a la conciencia sino también a la intencionalidad (Gallagher 2007).

Tradicionalmente la intencionalidad, en tanto que es la capacidad que tiene la mente para estar dirigida (o representar) a objetos del mundo, ha sido definida como una característica

¹⁰⁰ “If human agency is then material is, there is no way that human and material agency can be disentangled. Or else, while agency and intentionality may not be properties of things, they are not properties of humans either: they are the properties of material engagement, that is, of the grey zone where brain, body and culture conflate” (Malafouris, 2008 p. 55).

¹⁰¹ “The point I am trying to make here is that, although the concept of agency is much contested, it is contested within the theoretical margins of a narrow anthropocentric perspective. This anthropocentric view of agency, though it incorporates a variety of nuances, is based on a general agreement about a single undisputable fact: that agency, in the real sense of the word, is a human property, and “the only true agents in history are human individuals” (Giddens and Pierson 1998, p. 89)” (Malafouris 2013, p. 121-122).

fundamental de los estados mentales humanos; pero además, como una propiedad interna de los sujetos (Searle 1983). Por lo tanto, si aceptamos que la intencionalidad es una propiedad mental interna, entonces no hay modo de sugerir que la agencia humana no es también una propiedad interna.

Así, si Malafouris quiere negar el antropocentrismo alrededor de la agencia entonces le quedan dos caminos: 1) negar que la agencia esté conectada con la intencionalidad, o 2) aceptar 1 pero negar que la intencionalidad sea una propiedad interna. Malafouris elegirá la opción 2. Veamos una manera de presentar su argumento:

1. Si la intencionalidad es una propiedad mental humana interna, entonces la agencia también es una propiedad humana interna.
 2. Si 1 es el caso, entonces la agencia humana no puede ser un fenómeno emergente.
 3. No es el caso que 1 porque el antecedente es falso.
- ∴ De 1) y de 3) se sigue que la agencia no es una propiedad humana interna.

Si la agencia no es una propiedad interna, Malafouris dirá que entonces la agencia es una propiedad material, es decir que surge cuando interactuamos con el mundo. Una vez presentado el argumento Malafouris tiene que mostrar por qué la intencionalidad no es una propiedad interna; o dicho en otras palabras: Malafouris tiene que demostrar por qué el antecedente, del condicional, de la premisa 1 es falso.

4.4.2.1 Intencionalidad en la acción

Como hemos visto, la intencionalidad [*intentionality*] es una de las características principales de la agencia, por ello Malafouris no negará que la agencia esté ligada a ella sino que defenderá que “[c]omo muchas otras de las dimensiones de la mente humana, la intencionalidad debería ser entendida como un fenómeno distribuido, emergente e interactivo más que como un estado mental subjetivo [...]”¹⁰² (Malafouris 2008, p. 33). Así, parte de la argumentación se centrará en rechazar una parte de la tesis de la intencionalidad de Searle (1983) y defender que la intencionalidad no está antes de la acción sino que la intencionalidad *está en la acción*. Por esta razón, la ruta argumentativa

¹⁰² “[...] [a]s with many other dimensions of the human mind, intentionality should be understood as a distributed, emergent and interactive phenomenon rather than as a subjective mental state” (Malafouris 2008, p. 33).

de Malafouris se abocará a presentar las tesis filosóficas de Searle sobre la intencionalidad y la acción.

Para Searle la acción tiene dos partes, 1) un *estado intencional* en la cabeza y 2) un *movimiento externo* en el mundo. Sin embargo, dentro de los estados intencionales distingue entre dos tipos de intenciones: a) estados intencionales con *intenciones previas* [prior intention] y b) estados intencionales con *intencionalidad en la acción*¹⁰³ [intention-in-action]. Mientras que en a) lo que tenemos son acciones que van precedidas por intenciones, en b) lo que tenemos son el tipo de acciones en donde las intenciones se dan al mismo tiempo que las acciones.

Estos dos tipos de intencionalidad son para Searle fenómenos representacionales e internos del sujeto, pues en ambos casos algo interno (la intención) causa movimiento externo en el mundo (el movimiento del agente). Para Searle la diferencia entre las intenciones previas y la intencionalidad en la acción es que en este último tipo el *estado intencional* interno y el *movimiento externo* son indistinguibles. Para ilustrar esta distinción piense en las dos situaciones siguientes:

- A. Mientras escribo noto que mi café se ha acabado, por lo tanto (y obedeciendo a mi deseo de beber más café) pienso que necesito ir a la cocina a preparar más café. Acto seguido lo que hago es pararme e ir a la cocina a preparar más café, en este caso mi acción de ir a preparar café está precedida por una intención. Ahora piense en la siguiente situación:
- B. Cuando vuelvo de la cocina con la taza de mi café en mano, me siento nuevamente a escribir en mi computadora. Minutos después mientras sigo concentrado en escribir, sin mirar tomo la taza de café pero a causa de un mal agarré el café se me derrama en la camisa y parte de él cae en la computadora. Una vez que el café se me derrama lo que hago es pararme instintivamente y moverme porque el café está caliente, acto seguido también alzo mi computadora por miedo a que se descomponga.

En la situación B, a diferencia de la situación A, mi acto de pararme se dio porque el café está caliente, por ello mi acto no tuvo intención previa, o al menos no conscientemente. Mi acción estuvo

¹⁰³ Searle (1983, p. 84) nos dice que podemos distinguir lingüísticamente entre estos dos tipos de intenciones con las oraciones i) “Haré X acción” y ii) “estoy haciendo X acción”. Mientras que i) expresa intenciones previas, ii) expresa intención en la acción.

determinada por la sensación de la quemadura del café, mientras que al quitar mi computadora tampoco tuve tiempo de analizar y meditar la situación. Ambas acciones fueron llevadas a cabo sin una intención previa, en el caso de pararme fue instintivo pues fue la sensación de dolor lo que hizo que me levantara en seguida; mientras que en el caso de salvar mi computadora tal vez fue un acto un poco más consiente. Sin embargo, en ambos casos tanto las intenciones como las acciones van al mismo tiempo, por ello es que se hacen indistinguibles.

Ahora bien, dado que para Malafouris la agencia tiene que ver con acciones en el mundo más que con procesos mentales internos del sujeto; él cree que nuestra idea de agencia tienen que ir de la mano con la intencionalidad, pero con la que se da en los estados intencionales que tienen intenciones en la acción:

Aceptando que la agencia es sobre eventos causales en el mundo físico más que sobre eventos representacionales en nuestro mundo mental, se sigue que si una asociación entre agencia e intencionalidad puede ser hecha, tiene que ser con el tipo de intencionalidad aquí llamado intencionalidad en la acción¹⁰⁴ (2008, p. 30).

A pesar de que el tema del externismo se hace presente en la propuesta de Malafouris por la influencia de [Clark y Chalmers \(1998\)](#), este solo le sirve como punto de partida pues una vez que analiza el tema de la agencia sus conclusiones y tesis se vuelven más enactivas. Una vez que Malafouris da un enfoque enactivo a la intencionalidad podría apoyarse en el trabajo de [Noë \(2004\)](#), [Thompson \(2007\)](#) o el enactivismo radical de [Hutto y Myin \(2013\)](#). A pesar de esto, bien podríamos defender que a pesar de que en algunos casos las intenciones están en la acción en la gran mayoría estas están antes de la acción; además, como habíamos señalado anteriormente, una característica de la agencia en humanos es su conexión con la consciencia. De este modo, aunque las intenciones estén en la acción aún tenemos un referente para determinar que la agencia humana no sea material como sugiere Malafouris.

Para poner en jaque estas críticas Malafouris presenta un experimento llevado a cabo por [Libet \(1983, 1985, 1999\)](#), el cual es uno de los experimentos más famosos hechos con la agencia, el

¹⁰⁴ “Accepting that agency is about causal events in the physical world rather than about representational events in our mental world, it follows that if an association between agency and intentionality can be made, it has to be with the type of intentionality here called intention-in-action” ([Malafouris 2008, p. 30](#)).

libre albedrío y la consciencia. Los trabajos de Libet muestran que antes de que seamos conscientes de que queremos llevar a cabo una acción ya hay cambios eléctricos en nuestro cerebro.

Malafouris tomará los pasos descritos de Libet para analizarlos en una línea temporal (véase la **Figura 7**) y mostrar dos puntos: 1) que los cambios eléctricos en el cerebro empiezan 550 milisegundos antes de la acción y 2) los sujetos llegan a ser conscientes de la intención de actuar 350-400 milisegundos después del paso 1. De estos dos reportes se concluye que los sujetos llegan a ser conscientes de la intención de actuar solo 200 milisegundos antes del acto motor.

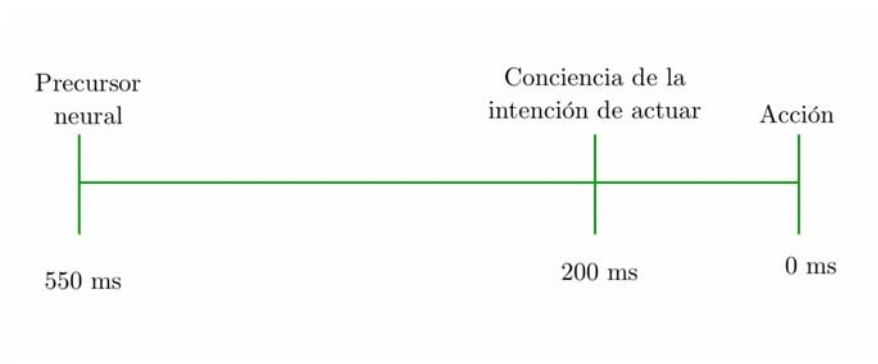


Figura 7. *Análisis temporal de los experimentos de Libet (1983).*

Una vez que apela a estos experimentos, Malafouris concluye que la consciencia, la cual había sido piedra angular para defender la agencia como una propiedad enteramente humana, tampoco puede servir de referente para adscribir agencia. Si antes de tener la intención consciente de actuar ya hay un precursor neuronal, entonces el cerebro decide iniciar (o preparar) un acto antes incluso de que seamos conscientes. Por ello es que la consciencia tampoco puede ser una cláusula para prescribir agencia, de este modo la agencia más que una propiedad humana es un proceso emergente que surge cuando interactúan el cerebro, el cuerpo y el mundo.

Este enfoque de la agencia va muy bien de la mano con las tesis enactivistas de la cognición, por ello es que uno de los ejemplos que Malafouris pone para ejemplificar a la mente activa es el caso de un alfarero creando un vaso en un torno. En este ejemplo, Malafouris señala que queda patente como la agencia es una propiedad emergente pues en el simple acto de crear un vaso se utiliza el cerebro, el cuerpo y el torno. O dicho de otro modo, para crear un vaso de barro necesitamos recursos físicos, mentales, biológicos y materiales. Cuando un alfarero está haciendo un vaso en el torno, ¿cómo funcionan sus manos, el torno y su cerebro? Para Malafouris lo que hay en

ese acto de creación es agencia, pero está se da solo en el momento en que el alfarero construye el vaso pues en ese momento es cuando todos los recursos convergen en una sola acción: los dedos agarrando la arcilla precisa, las manos dándole la forma en el torno, etc. En este acto, más que un proceso pensado lo que hay es acción.

4.5 Crítica al argumento internista

El argumento de Giere en contra de extender la agencia nos lleva a la conclusión de que la agencia no es un fenómeno extendido ni distribuido, sino una propiedad interna de los humanos. El argumento internista de la agencia demuestra que la agencia no puede ser un proceso compartido entre nosotros y las herramientas, pero dicha demostración parte del nivel ético y brinca al nivel ontológico. Una estrategia más fuerte de este argumento sería apelar al sentido de agencia y al sentido de pertenencia corporal, es decir defender que los humanos además de ser agentes tenemos la noción de que llevamos a cabo las acciones con nuestros cuerpos. De este modo Giere podría defender que:

A) Los humanos tenemos algunas características (el sentido de agencia y el sentido de pertenencia corporal) que no podemos compartir con los demás componentes de un sistema cognitivo extendido.

Si aceptamos la tesis A), inmediatamente nos vemos compelidos a aceptar la conclusión siguiente: B) Si dichas propiedades son solo humanas, entonces no pueden ser distribuidas.

Pero mientras que dichas propiedades pueden no estar distribuidas entre los diferentes componentes del sistema, no se descarta que se originen solo en la interacción de nuestra parte biológica con la parte material del mundo. Y precisamente el argumento de [Malafouris \(2008\)](#) nos muestra por qué la agencia es una propiedad emergente. La idea errónea que subyace en la postura de Giere es que podemos descomponer un sistema cognitivo en partes y que al hacer esto, la agencia quedaría solo en el componente humano de este sistema. Sin embargo, hemos ya dado argumentos de por qué no podemos pensar a la mente como separada del mundo; y mucho menos, descomponer a un sistema cognitivo en partes.

4.6 Implicaciones del argumento externista material

Respecto a la postura de [Malafouris \(2008, 2013\)](#), otra de sus ventajas es que puede alinearse con las tesis de [Silberstein y Chemero \(2011\)](#), pues ambos defienden un tipo de agencia anti-representacionista y externista: “[n]osotros rechazamos todas las suposiciones en favor de un reporte dinámico de las acciones intencionales y la agencia; un reporte que nos permite evitar los extremos del dualismo y el eliminativismo sobre las acciones intencionales y la agencia”¹⁰⁵ ([Silberstein y Chemero 2011, p. 2](#)). Si unimos este punto de vista junto con el argumento presentado por [Malafouris \(2008\)](#), entonces la agencia tiene que entenderse en términos enactivistas y sobre todo desde un punto de vista dinámico. Así, mirando atentamente todo lo que hasta aquí se ha presentado podemos ver por qué la idea de que podemos separar al hombre de su nicho ecológico (y de las herramientas que utiliza) resulta cada vez más una posición dudable. No podemos trazar una línea divisoria entre el hombre y el mundo. Así como Immanuel Kant defendió (en una vena trascendental) que no podemos pensar a los objetos fuera del espacio y del tiempo, de manera similar nos sucede a nosotros si entendemos por tiempo y espacio al mundo: no podemos separarnos o situarnos fuera de él.

La idea de un cerebro desnudo separado del cuerpo y del mundo es una hipótesis errónea, pues la cognición se da por la interacción dinámica entre el cerebro, el cuerpo y el mundo. Por este y otros motivos debemos entender a la agencia como un proceso emergente que surge cuando interactúan estos tres componentes. Una vez que miramos la agencia de este modo podemos concluir los siguientes puntos:

1) La agencia es un proceso dinámico y no necesita de representaciones para su explicación. Además, a pesar de que el argumento de Malafouris se centra en lo que llamamos la agencia humana, sus implicaciones van para los tres tipos de agencia que habíamos enunciado en el esquema tripartito. Pues si entendemos a la agencia como una propiedad emergente podemos aplicar su noción de agencia no solo a casos humanos (como el caso del alfarero) sino también a agentes artificiales como los de [Beer \(2003\)](#).

2) Entender la agencia como un proceso emergente da puerta a que las investigaciones en robótica e inteligencia artificial puedan trabajar con la agencia. Una vez que se elimina la tesis de

¹⁰⁵ “We reject all these assumptions in favor of a dynamical account of intentional action and agency; an account that allows us to avoid the extremes of dualism and eliminativism about intentional action and agency” ([Silberstein y Chemero 2011, p. 2](#)).

que la agencia es una propiedad enteramente humana e interna, tenemos un punto de partida teórico para trabajar en modelar la agencia en sistemas artificiales. Además, con este punto de vista de la agencia, proyectos históricos del siglo XX como las tortugas de [Walter \(1950\)](#) tienen que ser consideradas en sentido estricto agentes.

3) Finalmente, el entender la agencia como un proceso emergente nos permite empezar a estudiar si la idea de agencia colectiva o agencia social tienen sentido ([véase Stapleton y Froese 2015](#)).

Conclusión

Dos han sido los temas claves que he tratado en esta tesis: los casos de cognición extendida y la agencia. Mi pregunta de partida ha sido un condicional simple: si nos tomamos en serio los casos de cognición extendida, ¿cómo debemos entender la agencia? Poniendo dicho condicional en su notación formal tenemos la forma $p \rightarrow q$.

A pesar de que ha sido esta simple pregunta el hilo conductor de esta tesis, la ruta que he seguido ha sido minuciosa y por ello me he tenido que centrar en lo que de otro modo pude haber tomado como suposiciones. Al darme cuenta de que en la epistemología los casos de cognición extendida (a pesar de la enorme influencia y de las implicaciones de los casos de cognición extendida no solo en la filosofía de la mente sino en la biología, la psicología, la arqueología, la inteligencia artificial, etc.) no han sido del todo analizados, en lugar de solo obviar los casos de cognición extendida he hecho antes de proseguir en mi investigación de la agencia una apología general del externismo de la mente.

Así, no he tomado a “p” como casos plenamente aceptados sino que la primera parte de este trabajo ha sido mostrar por qué los casos de cognición extendida tienen que ser tomados en serio. Por ello, la ruta que he seguido ha sido presentar la tesis tradicional de la mente extendida tal y como fue presentada por [Clark y Chalmers \(1998\)](#), pues ha sido esta tesis la que ha desencadenado que el externismo de la cognición y la mente haya sido planteado en filosofía. Una vez que he hecho una exposición de dicha tesis, también he presentado algunas objeciones; la mas importante de ella ha sido la llamada *falacia constitución-acoplamiento*. Dicha crítica, esgrimida por [Adams y Aizawa \(2001\)](#), afirma que del hecho de que algún elemento externo a la mente juegue un rol causal en algún estado mental, no se sigue que dicho elemento sea parte constitutiva de la mente. En sus diferentes versiones esta ha sido la principal crítica a la tesis tradicional de la mente de Clark y Chalmers, por ello he destinado todo un capítulo a dar una respuesta a esta crítica.

Mis respuestas fueron tomadas desde tres posturas diferentes: Menary ([2007](#)), Chemero ([2009](#)) y Hutto & Myin ([2013](#)). Algunas de estas posturas son más radicales que otras pero las tres tienen un punto en común, cada una (a su manera) tiene una respuesta a las críticas internistas de Adams y Aizawa. Posteriormente, después de dar tres respuestas a la *falacia constitución-acoplamiento*, y mostrar con ello que los ataques internistas pueden responderse de diferentes

maneras, he concluido que tenemos razones más que suficientes para tomarnos en serio los casos de cognición extendida.

Así, una vez mostrado que “p” es el caso y que, por lo tanto, debe ser tomado en serio, he pasado a analizar al consecuente del condicional. El problema al empezar hablar de agencia es que a pesar de que en términos generales podemos definirla como la capacidad de actuar en el mundo (Schlosser 2015), una vez que entramos en una análisis profundo nos damos cuenta de que tanto en psicología, robótica, inteligencia artificial, fenomenología, neurociencias, psiquiatría, vida artificial etc., tenemos un sin fin de definiciones de la agencia y en algunos casos hasta incompatibles. Pero, además de las múltiples definiciones de la agencia, Barandiaran et al., (2009) ha señalado que la agencia tiene un problema metodológico: la idea de que a diferencia de los términos de racionalidad o mente, la agencia no es un término controversial.

Así, por estrategia metodológica he descompuesto el problema de la agencia en tres problemas: i) el problema metodológico, ii) el problema ontológico y iii) el problema externista. Una vez hecho esto presento las diferentes maneras en que la agencia es utilizada en diferentes áreas, con esto se doy una respuesta a i) y al mismo tiempo queda patente el problema ii). Con el objetivo de dar una respuesta al problema ontológico presento un marco tripartito que divide a la agencia en tres tipos de agencia: humana, animal y artificial. Este marco tripartito nos permite mantener las diferencias entre las diferentes concepciones de la agencia pero al mismo tiempo agruparlas bajo un solo fenómeno.

Además, presento la idea de que la agencia tiene que ser definida como la capacidad de generar acciones pero también, en el caso de humanos, de generar pensamientos conscientes. Otra característica es que nuevamente en el caso de humanos tenemos un sentido llamado el *sentido de agencia*, el cual nos permite distinguir entre los agentes humanos, animales y artificiales. Así, una de las conclusiones del problema ii) es que mientras que podemos tener agentes naturales (humanos y animales) y artificiales, solo los humanos tenemos un sentido de agencia. Finalmente, una vez que tenemos una noción clara de la agencia, y al mismo tiempo buenas razones par tomarnos en serio los casos de cognición extendida, paso entonces a analizar lo que sucede con el condicional $p \rightarrow q$.

Así, en la parte final de la tesis presento dos posiciones que, a pesar de que ambas sostienen un externismo de los estados mentales, llevan a conclusiones diferentes. Giere (2006, 2012) concluye que la agencia solo puede ser atribuida a los componentes humanos de un sistema

cognitivo extendido; por ello, algunas de las consecuencias principales de esta tesis es que la agencia es una propiedad enteramente humana e interna.

Por otro lado, [Malafouris \(2008, 2013\)](#) concluye que la agencia no es una propiedad, ni es interna y tampoco solo humana. Malafouris defiende que la agencia es una propiedad emergente que surge solo cuando interactúan el cerebro, el cuerpo y el mundo. Una vez analizados ambos argumentos tomo partido por Malafouris y concluyo que la agencia (y los estados mentales) tienen que ser entendidos de una manera más enactiva, es decir sin representaciones y con un punto de vista dinámico. Así, con estas conclusiones propongo entender a la agencia no como una sustancia interna humana sino como una proceso emergente que surge solo cuando interactuamos con el mundo.

El proyecto de emular la agencia humana tiene un largo camino por recorrer ya que aún se necesitan muchos años de investigación (en torno a nuestra psicología, biología y fenomenología) para poder implementarla en agentes artificiales. Sin embargo, empezar a entender la agencia como un fenómeno emergente más que como una propiedad interna y humana es el primer paso. Por ello, espero que esta investigación sea tomada como un pequeño paso en el largo camino que queda por recorrer para descubrir qué clase de agente somos y qué es lo que nos hace humanos.

Bibliografía

- Adams, F. R. and K. Aizawa (2001). The Bounds of Cognition. *Philosophical Psychology* 14(1): 43-64.
- Adams, F. R. and K. Aizawa (2010). “Defending the bounds of cognition,” en R. Menary (ed.), *The Extended Mind*. MIT Press.
- Aguilar, W., G. Santamaría-Bonfil, T. Froese and C. Gershenson (2014). The past, present, and future of artificial life. *Frontiers in Robotics and AI* 1: 1-15.
- Ashby, W. R. (1952/1960). *Design for a Brain*. Chapman & Hall.
- Balconi, M. (2010a). “The Sense of Agency in Psychology and Neuropsychology”, en M. Balconi (ed.), *Neuropsychology of the Sense of Agency. From Consciousness to Action*. Italia: Spring.
- Balconi, M. (2010b). “The Neuropsychology of Senses of Agency: Theoretical and Empirical Contributions”, en M. Balconi (ed.), *Neuropsychology of the Sense of Agency. From Consciousness to Action*. Italia: Spring.
- Balconi, M. (2010c). Preface. “Neuropsychology of the Sense of Agency. From Consciousness to Action”, en M. Balconi (ed.), *Neuropsychology of the Sense of Agency. From Consciousness to Action*. Italia: Spring.
- Balconi, M. and A. Bortolotti (2010). “Body and Self-awareness: Functional and Dysfunctional Mechanisms”, en M. Balconi (ed.), *Neuropsychology of the Sense of Agency. From Consciousness to Action*. Italia: Spring.
- Barandiaran, X. (2008). *Mental Life: A naturalized approach to the autonomy of cognitive agents*. Tesis de doctorado, University of Basque Country.
- Barandiaran, X., E. Di Paolo and M. Rohde (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior* 17(5): 367-386.
- Barrett, L. (2011). *Beyond the Brain: How Body and Environment Shape Animal and Human Minds*. Princeton University Press.
- Beer, R. (1995). A dynamical systems perspective on agent-environment interactions. *Artificial Intelligence* 72: 173-215.
- Beer, R. (2003). The Dynamics of Active Categorical Perception in an Evolved Model Agent. *Adaptive Behavior* 11: 209-243.
- Brentano, F. (1874/1995). *Psychology from an Empirical Standpoint*. Routledge.
- Brooks, R. (1990). Elephants don't play chess. *Robot. Auton. Syst.* 6(1-2): 3-15.

- Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence* 47: 139-159.
- Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy* 4(1): 73-122.
- Chaminade, T. and J. Decety (2002). Leader or Follower? Involvement of the Inferior Parietal Lobule in Agency. *Neuroreport* 13(1528): 1975-1978.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, Massachusetts: MIT Press.
- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. MIT Press.
- Clark, A. (2008). “Where Brain, Body and World Collide”, en C. Knappett and L. Malafouris (ed.), *Material Agency - Towards a Non-Anthropocentric Approach*. Springer.
- Clark, A. (2010a). “Memento's Revenge : The Extended Mind, Extended”, en R. Menary (ed.), *The Extended Mind*. MIT Press.
- Clark, A. (2010b). “Coupling, Constitution and the Cognitive Kind: A Reply to Adams and Aizawa”, en R. Menary (ed.), *The Extended Mind*. MIT Press.
- Clark, A. and D. J. Chalmers (1998). The extended mind. *Analysis* 58(1): 7-19.
- Clark, A., D. Pritchard and K. Vaesen (2012). Introduction: Extended cognition and epistemology. *Philosophical Explorations* 15(2): 87 - 90.
- Crane, T. (2001). *Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford University Press.
- Dawkins, R. (1982). *The Extended Phenotype*. Oxford University Press.
- Dennett, D. C. (1989). *The Intentional Stance*. Bradford Book.
- Descartes, R. (1641/2005). *Meditaciones metafísicas*. Alianza Editorial.
- Di Paolo, E. (2009). Extended Life. *Topoi-an International Review of Philosophy* 28(1): 9-21.
- Di Paolo, E. (2010). Robotics Inspired in the Organism. *Intellectica* 1-2(53-54): 129-162.
- Dreyfus, H. (1979). *What Computers Can't Do*. Harper and Row.
- Dreyfus, H. (1991). *Being-in-the-world: A Commentary on Heidegger's Being and Time*. MIT Press.
- Dreyfus, H. L. (2013). “The myth of the pervasiveness of the mental”, en J. K. Schear (ed.), *Mind, Reason, and Being-in-the-world: The McDowell-Dreyfus Debate*. Routledge.
- Farrer, C. and D. Frith (2001). Experiencing Oneself vs. Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency. *NeuroImage* 15: 596-603.

- Fodor, J. (2009). Where is my mind? *London Review of Books* 31(3).
- Froese, T., N. Virgo and E. Izquierdo (2007). “Autonomy: a review and a reappraisal”, en F. A. e. C. et. al. (eds.), *Proceedings of the 9th European Conference on Artificial Life*. Berlin, Springer-Verlag.
- Froese, T. and T. Ziemke (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence* 173: 466-500.
- Froese T. & Stewart J. (2010) Life after Ashby: Ultrastability and the autopoietic foundations of biological individuality. *Cybernetics & Human Knowing* 17(4): 83–106.
- Froese, T., C. Gershenson and D. Rosenblueth (2013). The Dynamically Extended Mind: A Minimal Modeling Case Study. [arXiv preprint http://arxiv.org/abs/1305.1958](http://arxiv.org/abs/1305.1958).
- Gallagher, S. (2000). Philosophical conceptions of the self. *Trends in Cognitive Sciences* 4(1): 14-21.
- Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass* 2(2): 347–357.
- Gallagher, S. and D. Zahavi (2008). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. Routledge.
- Gardner, H. (1985). *The Mind's New Science: A History of the Cognitive Revolution*. Basic Books.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.
- Giere, R. N. (2006). *Scientific Perspectivism*. Chicago, University of Chicago Press.
- Giere, R. N. (2012). Scientific cognition: human centered but not human bound. *Philosophical Explorations* 15(2): 199 - 206.
- Gigerenzer, G. (1998). “Psychological Challenges For a Normative Models”, en D. Gabbay and P. Smet, *Quantified Representation of Uncertainty and Imprecision*. Dordrecht: Kluwer.
- Gigerenzer, G. (2005). I think, therefore I err. *Social Research: An International Quarterly* 72(1): 1-24.
- Grey, W. (1950). An Imitation of Life. *Scientific American* 182: 42-54.
- Grünbaum, T. (2010). “Action and Agency”, en S. Gallagher and D. Schmicking (eds.), *Handbook of Phenomenology and Cognitive Science*. New York: Springer.
- Harvey, I., E. Di Paolo, R. Wood and M. Quinn (2005). Evolutionary Robotics: A New Scientific Tool for Studying Cognition. *Artificial Life* 11: 79-98.
- Heidegger, M. ((1927/1996)). *Being and Time*. State University of New York Press.

- Hernandez, J. P. (2015). Racionalidad y acción no reflexiva. El debate Dreyfus-McDowell. *Crítica* 47(140): 43-63.
- Hochberg, L., D. Bacher, B. Jarosiewicz, N. Masse, J. Simeral, J. Vogel, S. Haddadin, J. Liu, S. Cash, P. van der Smagt and J. Donoghue (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485(7398): 372-375.
- Holland, O. (2003). The first biologically inspired robots. *Robotica* 21: 351-363.
- Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.
- Hutto, D. D. and E. Myin (2013). *Radicalizing Enactivism: Basic Minds Without Content*. MIT Press.
- Jacob, P. (2014). Intentionality. *The Stanford Encyclopedia of Philosophy*. Winter Edition.
- Jonas, H. (1966). *The Phenomenon of Life: Toward a Philosophical Biology*. Northwestern University Press.
- Kant, I. (1781/2006). *Crítica de la razón pura*. Alfaguara.
- Kirsh, D. and P. Maglio (1994). "On distinguishing epistemic from pragmatic action." *Cognitive Science* 18: 513-549.
- Kiverstein, J. and M. Farina (2011). "Embraining Culture: Leaky Minds and Spongy Brains." *Teorema* 32(2): 35-53.
- Knappett, C. and L. Malafouris (2008). "Material and Nonhuman Agency: An Introduction", en C. Knappett and L. Malafouris (eds), *Material Agency: Towards a Non-Anthropocentric Approach*. Springer.
- Laland, K. and M. O'Brien (2011). Cultural Niche Construction: An Introduction. *Biol Theory* 6: 191-202.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. New York, Oxford University Press.
- Lewontin, R. (1982). "Organism and environment", en P. EC. (ed.), *Learning, development and culture*. New York: Wiley.
- Lewontin, R. (1983). "Gene, organism, and environment", en B. DS. (eds.), *Evolution from molecules to men*. Cambridge: Cambridge University Press.
- Libet, B., C. Gleason, E. Wright and D. Pearl (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* 106: 623--664.
- Libet, B. W. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8(4): 529-566.

- Libet, B. W. (1999). Do we have free will? *Journal of Consciousness Studies* 6(8-9): 47-57.
- Malafouris, L. (2008). “At the Potter's Wheel: An Argument for Material Agency” en C. Knappett and L. Malafouris (eds.), *Material Agency: Towards a Non-Anthropocentric Approach*.
- Malafouris, L. (2013). *How Things Shape the Mind*. MIT Press.
- Martínez, A. (2013). La tesis de la mente extendida: nuevas orientaciones. *Ciencia cognitiva* 7(2): 37-39.
- McDowell, J. (2013). “The myth of the mind as detached”, en J. K. Schear (ed.), *Mind, Reason, and Being-in-the-world: The McDowell-Dreyfus Debate*. Routledge.
- Menary, R. (2006). *Radical Enactivism: Intentionality, Phenomenology, and Narrative : Focus on the Philosophy of Daniel D. Hutto*. J. Benjamins Publishing Company.
- Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Great Britain: Palgrave Macmillan.
- Menary, R. (2010a). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences* 9(4): 459-463.
- Menary, R. (2010b). “Introduction: The Extended Mind in Focus”, en R. Menary (ed.), *The Extended Mind*. MIT Press.
- Menary, R. (2012). Cognitive practices and cognitive character. *Philosophical Explorations* 15(2): 147 - 164.
- Merleau-Ponty, M. (1942/1976). *La Estructura del comportamiento*. Librería Hachette.
- Merleau-Ponty, M. (1945). *Fenomenología de la percepción*. Planeta-Agostini.
- Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice*. Bradford Books / MIT Press.
- Moreno, A. and A. Etxeberria (2005). Agency in natural and artificial systems. *Artificial Life* 11(1-2): 161-175.
- Nicoletis, M. and J. Chapin (2002). Controlling Robots with the Mind. *Scientific American* 287(4): 46-53.
- Noë, A. (2004). *Action in Perception*. Bradford book.
- Pfeifer, R. (1996). “Building "Fungus Eaters": Design Principles of Autonomous Agents”, en P. Maes, M. Mataric, J. Meyer, J. Pollack and S. Wilson (eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour*. MIT Press.

- Pickering, A. (2002). Cybernetics And The Mangle: Ashby, Beer And Pask. *Social Studies of Science* 32(3): 413-437.
- Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7: 131-193.
- Rosch, E., E. Thompson and F. J. Varela (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
- Ross, D. (2007). "Introduction: Science Catches the Will", en D. Ross, D. Spurrett, H. Kincaid and G. Lyn (eds.), *Distributed Cognition and the Will*. Cambridge, Massachusetts: MIT Press.
- Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. Cambridge: Cambridge University Press.
- Rowlands, M. (2010). *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. A Bradford Book.
- Rupert, R. D. (2009). *Cognitive Systems and the Extended Mind*. Oxford University Press, USA.
- Sacks, O. (1985). *The Man Who Mistook His Wife for a Hat and Other Clinical Tales*. New York, Summit Books.
- Shear, J. K. (2013). *Mind, Reason, and Being-in-the-world: The McDowell-Dreyfus Debate*. Routledge.
- Schlosser, M. (2015). Agency. *The Stanford Encyclopedia of Philosophy*. E. N. Zalta.
- Searle, J. (1983). *Intentionality*. Oxford University Press.
- Sheldon, P. D. (2007). "What Kind of Agent Are We? A Naturalistic Framework for the Study of Human Agency", en D. Ross, D. Spurrett, H. Kincaid and G. Lyn (eds.) *Distributed Cognition and the Will*. Cambridge, Massachusetts: MIT Press.
- Silberstein, M. and A. Chemero (2011). Dynamics, agency and intentional action. *Humana.Mente* 15: 1-19.
- Silberstein, M. and A. Chemero (2012). Complexity and Extended Phenomenological-Cognitive Systems. *Topics in Cognitive Science* 4(1): 35-50.
- Steels, L. (1995). "Building Agents out of Autonomous Behavior Systems", en L. Steels and R. Brooks, L. Erlbaum (eds.), *The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated Agents*. Associates.
- Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Wiley-Blackwell.

- Sterelny, K. (2004). “Externalism, epistemic artefacts and the extended mind”, en R. Schantz (ed.), *The Externalist Challenge*. De Gruyter.
- Sutton, J. (2010). “Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process”, en R. Menary (ed.), *The Extended Mind*. MIT Press.
- Synofzik, M., G. Vosgerau and A. Newen (2008). I move, therefore I am: A new theoretical framework to investigate agency and ownership. *Consciousness and Cognition* 17(2): 411 - 424.
- Synofzik, M. and M. Voss (2010). “Disturbances of the Sense of Agency in Schizophrenia”, en M. Balconi (ed.), *Neuropsychology of the Sense of Agency. From Consciousness to Action*. Italia, Spring.
- Theiner, G. (2011). *Res Cogitans Extensa: A Philosophical Defense of the Extended Mind Thesis*. Peter Lang.
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Belknap Press of Harvard University Press.
- Toda, M. (1962). The design of a fungus-eater: A model of human behavior in an unsophisticated environment. *Behavioral Science*(2): 164-183.
- van Orden, G., J. Holden and M. Turvey (2003). Self- organization of cognitive performance. *Journal of Experimental Psychology: General* 132: 331-350.
- van Orden, G., J. Holden and M. Turvey (2005). Human cognition and 1/f scaling. *Journal of Experimental Psychology: General* 134(117-123).
- Varela, F. (1988/2005). *Conocer. Las ciencias cognitivas: tendencias y perspectivas. Cartografía de las ideas actuales*. Barcelona: Gedisa.
- Vega, J. (2005). Mentes híbridas: cognición representaciones externas y artefactos epistémicos. *Revista de Antropología Iberoamericana* (Especial): 1-36.
- Velliste, M., S. Perel, M. C. Spalding, A. S. Whitford and A. B. Schwartz (2008). Cortical control of a prosthetic arm for self-feeding. *Nature* 453: 1098-1101.
- Weber, A. and F. J. Varela (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences* 1(2): 97-125.
- Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*. New York: Cambridge University Press.
- Wittgenstein, L. (1921/2012). *Tractatus logico-philosophicus*. Grupo Anaya Comercial.
- Wittgenstein, L. (1967). *Zettel*. Londres: Basil Blackwell.