



UNIVERSIDAD DE CHILE
FACULTAD DE FILOSOFÍA Y HUMANIDADES
ESCUELA DE POSTGRADO

**Relectura de la polémica simbolismo-conexionismo
desde una perspectiva pluralista de las ciencias y el
fenómeno cognitivo**

Tesis para optar al grado de Magíster en Estudios Cognitivos

Alumno:
Gonzalo Luis Aguilera Retamal

Profesor Guía:
Manuel Rodríguez Tudor

Santiago de Chile
2016

Resumen

La presente investigación es el desarrollo de una relectura pluralista explicativa de la relación interteórica entre simbolismo y conexionismo, y se organiza en dos partes bien diferenciadas. La primera parte corresponde a la presentación ortodoxa de los contenidos de la polémica entre el simbolismo y el conexionismo a fines de la década de los 80', una exposición sucinta de la evolución de algunos de tales contenidos y, finalmente, la estructuración del punto de partida teórico que contiene, por un lado, tanto supuestos básicos sobre algunos de los problemas de la polémica, como por otro lado, las reflexiones y preguntas que servirán como metas de la investigación en la segunda parte de la tesis.

La segunda parte, partiendo de una búsqueda frustrada para hallar algún terreno común o acuerdo transversal en ciencias cognitivas, desembocará en la defensa y exposición del pluralismo, postura filosófica de las ciencias que servirá como trasfondo para la creación de un medio conceptual descriptivo que funcione como soporte abstracto para explicitar y detallar cualquier relación interteórica posible. Este medio conceptual, que intenta ir en línea con una epistemología y filosofía de las ciencias actualizada, será denominado como "la tesis de los Instrumentos Conceptuales" y servirá como un modo plausible, no sólo de lidiar con un caso concreto de relación entre modelos en ciencias cognitivas, sino que, específicamente, para re-comprender y reorganizar la polémica entre el simbolismo cognitivista y el conexionismo subsimbolista.

A mi familia,
por su incondicionalidad
e infinito cariño.

Índice

Introducción.....	1
I. Las ramificaciones de la polémica Simbolismo-Conexionismo	5
1.1. Presentación de la polémica	5
1.1.1. Primer momento – Las ventajas del enfoque PDP según los propios conexionistas.....	6
1.1.2. Segundo momento – La crítica de los cognitivistas desde el simbolismo	18
1.1.3. Tercer momento – El (<i>innecesariamente</i>) prolongado debate de Paul Smolensky y Jerry Fodor	30
1.2. Una posible organización de los contenidos de la polémica a lo largo del tiempo	44
1.2.1. Considerando la polémica externamente.....	45
1.2.2. Considerando la polémica internamente	54
1.3. Desde las reflexiones en torno a la polémica a la síntesis de un punto de partida teórico	70
1.3.1. Reflexiones y cuestionamientos en torno a la polémica simbolismo-conexionismo y su evolución..	71
1.3.2. Digresión 1 – Causalidad, instrumentos formales versus empíricos y el cognitvismo.	76
1.3.3. Digresión 2 – Sobre la interpretación de la Paradoja de Smolensky y sus consecuencias	82
1.3.4. Síntesis del punto de partida teórico	85
II. El Pluralismo Explicativo, la abstracción de los Instrumentos Conceptuales y su aplicación	87
2.1. La búsqueda de un <i>terreno común</i> en filosofía de las Ciencias Cognitivas	88
2.2. Presentación del Pluralismo Explicativo.....	101
2.3. La tesis de los Instrumentos Conceptuales (IC) y sus consecuencias.....	117
2.3.1. Las metas de la Ciencia.....	117
2.3.2. La abstracción de los IC y sus principios de construcción.....	121
2.3.2.1. Definiciones sobre la estructura IC y funcionamiento.....	125
2.3.2.2. Identidad, dinámica y categorías de los IC.....	135
2.3.2.3. Revoluciones y la unidad de las ciencias.....	145
2.4. Aplicación de la tesis de los IC al análisis de algunos IC aislados en Ciencias Cognitivas.....	149
2.4.1. La neurona de McCulloch-Pitts	151
2.4.2. El perceptrón de Rosenblatt y la neurona de M&P.....	159
2.5. La relación simbolismo-conexionismo a la luz de la tesis de los IC.....	166
2.5.i. Pluralidad en las ciencias cognitivas	167

2.5.ii. Respecto a una posible definición de “simbolismo”	172
2.5.iii. Respecto a una posible definición de “conexionismo”	176
2.5.iv. ¿Competencia o colaboración?	178
2.5.v. Posibles objeciones a la conclusión anterior y sus respuestas	185
2.5.vi. Re-comprensión y re-organización de la polémica simbolismo-conexionismo	187
Conclusiones	192
3.1. Ventajas y limitaciones e inconsistencias de la tesis de los IC	192
3.2. Perspectivas de desarrollo e investigación futura	197
Bibliografía	201

Introducción¹

Con el descubrimiento de que, histológicamente, la corteza cerebral no estaba conformada por una masa de células nerviosas continuas (teoría reticular), sino que está compuesta por células discretas contiguas conectadas entre ellas mediante axones y dendritas (teoría de la neurona), se dio comienzo a la visión del cerebro, y en general del sistema nervioso, como configurado por redes neuronales conformadas de unidades discretas, aunque densamente conectadas, cuya aferencia anatómica estaría segregadamente distribuida a lo largo de todo el cuerpo al igual que su eferencia. Este primer hito preliminar, la teoría de la neurona, fue gracias a la obra combinada del histólogo español Santiago Ramón y Cajal en 1888 (López-Muñoz et al., 2006) y del citólogo alemán Heinrich von Waldeyer-Hartz. A este par de científicos decimonónicos se agrega la figura del fisiólogo inglés Sir Charles Scott Sherrington, quien postula la noción de la sinapsis, permitiéndose así configurar el paisaje de toda la neurofisiología posterior (Bennett, 1999, 114-6).

Entrando en el siglo XX, el estudio de estas redes neuronales se dio, como era de esperar, en el contexto de las discusiones académicas de la neurobiología y neuropsicología de la época, que seguía enfrascada en sofisticaciones de la antigua disputa entre los bandos del localizacionismo – postura cuyo exponente más conocido, Joseph Gall, ya lo defendía en el siglo XVIII (Finger, 2000, 119-124) – y el holismo – teoría defendida por Pierre Flourens, contemporáneo de Gall, quien impugnaría la frenología de su oponente (op. cit., 132-5). Por consiguiente, en una primera instancia, lo medular dentro de tal disputa fue saber dónde en tales redes neuronales estaban localizadas las diferentes funciones cognitivas tales como de recepción, almacenamiento, integración y producción. Posteriores trabajos teóricos, como el de Donald Hebb (1949), y experimentales, como el de su pupilo Karl Lashley (1950), llevados a cabo en el contexto del debate holismo-localizacionismo, permitieron asentar definitivamente que la memoria (en concreto, el constructo teórico del “engrama”) y, en general, las funciones cognitivas consideradas “superiores” estarían distribuidas y no localizadas específicamente en ningún sector de la corteza.

En efecto, gracias a tales investigaciones pioneras que permitieron guiar la investigación ulterior en las funciones cognitivas y de percepción, se puede afirmar a estas alturas del debate que las posiciones antagónicas se han suavizado hasta el punto de existir *terrenos comunes* de discusión en neurociencias. Ya es aceptado que – y quien no lo hace simplemente niega la evidencia neurocientífica acumulada hasta hoy – la memoria y funciones cognitivas estarían distribuidas, aunque no con ello se niega la existencia de modularidad y especificidad de las zonas encargadas del procesamiento de información primaria.

De manera provisional por “terreno común” (TC) entenderé aquellos (i) consensos fundamentales sobre hechos básicos, los cuales permiten tanto (ii) una superación de posturas conceptuales antagónicas, como (iii) la existencia de cierta metodología común. Por ejemplo, con respecto al estado actual del mapeo de las funciones cognitivas en neurociencias, existe el consenso de que las funciones se sitúan en redes – cumplimiento condición (i) – lo cual permite una superación del modularismo y el holismo – condición (ii) – y la existencia, en este caso aun en evolución, de metodologías que se van

¹ La presente tesis es parte del proyecto “El principio de composicionalidad en el contexto de los debates sobre la relación entre la mente y el lenguaje: un problema de elección de arquitecturas cognitivas” financiado por el Fondecyt Iniciación 11121604 a cargo del profesor Manuel Rodríguez Tudor.

complementando entre sí – condición (iii) –, por ejemplo del enfoque observacional e intervencionista. (Nazarova & Blagovechtchenski, 2015). Este supuesto, a saber, de la existencia de un TC establecido en neurociencias es capital para la presente tesis. Sin embargo, puede reemplazarse las neurociencias por cualquier disciplina cuya “madurez” permita un TC definido por los puntos anteriores (v. gr. la física), aunque hacerlo con otra disciplina sería poco natural, dada la conexión de subsunción de las neurociencias a las ciencias cognitivas².

Sin embargo, en el campo conjunto de las ciencias cognitivas tal TC sobre el cual asentar el debate inter-teórico, considero que, apenas estaría siendo insinuado en una serie de trabajos relativamente contemporáneos; relativamente reciente, el 2008, hubo un edición especial en la *Journal of Experimental & Theoretical Artificial Intelligence*, cuyo tópico a tratar era *Pluralism and the future of cognitive science*. De manera extremadamente sintética, el objetivo de esta edición especial retomaba un tema que ya existía al menos cuatro décadas de anticipación en los trabajos de filosofía de las ciencias (Suppes, 1978) y lo aplicaba a ciencias cognitivas como una forma tentativa que justificara y permitiera dar cabida a una *interacción constructiva de posturas conceptuales antagónicas*.

La necesidad de un TC en ciencias cognitivas es auto-evidente *prima facie* cuando se observa la oposición teórica de los elementos del cognitivismo³ y los avances en las técnicas de neuroimagen (Zubicaray, 2006; 272-3). En efecto, considero que cualquier progreso en la construcción de un TC, posibilitaría una verdadera colaboración entre las diferentes disciplinas y posturas teóricas que convergen en el entramado de las ciencias cognitivas. No obstante, para poder construir se hace imperativo enfrentarse a los resabios filosóficos de la ortodoxia cognitiva o cognitivismo, ya que son éstos los que impiden terminar de reconocer que el cognitivismo es insuficiente por sí solo para ocupar el puesto de “director de orquesta” del avance investigativo *interdisciplinar* que debiese constituir las ciencias cognitivas, “interdisciplinarietà” apuntada por Gardner (1985, 5-7, 42), Von Eckardt (1995)⁴ y también presente en introducciones al campo (Stillings et al., 1995; Kolak et al., 2006, 7-8).

Ahora bien, la propuesta central de esta tesis es que – dando por superados ciertos tópicos sobre los cuales tiene poco valor discutir dada la evidencia en neurofisiología, como, por ejemplo, la modularidad estrictamente fodoriana (Fodor, 1983) – **la polémica simbolismo-conexionismo debe ser re-**

² El lector avisado ya se habrá percatado que mi uso de “ciencias cognitivas” se remite a una bastante general de Gardner (1985, 5-7), donde por ciencias cognitivas estoy apuntando a la suma simple de las disciplinas tales como filosofía, psicología, inteligencia artificial, lingüística, antropología y neurociencias, y no a una perspectiva que las trascienda. Mi definición amplia de ciencias cognitivas, que contrasta con la reconstruida por Von Eckardt a base de la unidad de análisis de los marcos investigativos (Von Eckardt, 1995, 27-31), tiene su justificación, como se podrá prever, en mi defensa del pluralismo explicativo y una nueva unidad de análisis (segunda parte de la tesis). Este punto específico, a saber, de la definición del campo-conjunto de las ciencias cognitivas no será expuesto en el presente trabajo, ya que es parte de los tópicos que quedaron por desarrollar. Por otra parte, la evaluación crítica del concepto de “terreno común” estará en (2.1) y tentativamente redefinida en (3.2).

³ “Cognitivismo” tiene diversos significados dependiendo el contexto y el autor. Para algunos autores “cognitivismo” es sinónimo de *computacionalismo* (Chemero & Silberstein, 2008, 15). Por otro lado, los mismos conexionistas también serían computacionalistas (Piccinini & Bahar, 2013, 479-480). Empero, esta cuestión estaría lejos de ser dirimida de manera sencilla, al existir una distinción entre un conexionismo clásico versus uno no-clásico (Calvo, 2003). Una definición más exacta será desarrollada en la sección (2.5.ii).

⁴ Aunque en este caso más que “interdisciplinario” se habla de “transdisciplinario”.

comprendida a partir de un medio filosófico fundamentalmente descriptivo de trasfondo pluralista explicativo que posibilite comprender de manera más detallada los variados niveles posibles de relación interteórica, de modo de responder a una pregunta que es básica respecto de la misma polémica, a saber, hasta qué punto existe una competencia entre ambas “teorías” y hasta qué punto ésta puede ser dirimida en favor de uno u otro bando. Este medio filosófico será lo que denominaré “*la tesis de los Instrumentos Conceptuales (IC)*”. Empero, esta propuesta capital es un primer paso necesario para poder transitar a una segunda fase, la cual consistiría en averiguar de qué modo sería posible afirmar que el simbolismo posee *limitaciones* que conllevan una menor *profundidad explicativa-predictiva relativa* que el conexionismo (esto último será ensayado brevísimamente en la sección (3.2.)). La tesis, además de obviamente en una primera parte exponer la polémica simbolismo-conexionismo y algunas de sus consecuencias, se limitará exclusivamente a desarrollar cómo sería la forma de este medio filosófico enfocado en resolver la cuestión de las relaciones interteóricas, mientras que dejará bosquejado cómo es posible mostrar las limitaciones relativas del simbolismo respecto al conexionismo.

Por lo tanto, la presente tesis tiene como disciplina o campo de fondo, *primariamente*, y de manera general, la *filosofía de las ciencias*, mientras que, más particularmente, la *filosofía de las ciencias cognitivas*. *Secundariamente*, y nuevamente de modo general, tiene como disciplina de fondo, *las ciencias cognitivas* propiamente tales, y más particularmente *la historia de los primeros modelos cognitivos*. Esto es necesario aclararlo desde el comienzo, ya que en varios tramos de lo que se ha escrito he supuesto que el lector ya posee cierta familiaridad con los campos apuntados.

Originalmente mi intención era defender una tesis que postulase algo así como “en el campo del modelamiento cognitivo los modelos simbólicos tienen, en el mejor de los casos, un valor descriptivo, mientras que los modelos conexionistas tienen, en el peor de los casos, un valor explicativo”. No obstante, mi investigación me llevó a desechar el propio vocabulario de tal tesis, ya que, al interiorizarme en el campo de la filosofía de las ciencias y su evolución temática hasta la actualidad, me percaté que no tiene sentido aseverar algo así sin caer, tácitamente, en alguna variante del *monismo explicativo*, una de las características clave del cognitivismo – e inclusive de otros trasfondos filosóficos ligados a las ciencias cognitivas. Tomando en cuenta ello, mi plan partirá con una tesis que tenga a la vista la relectura de la relación entre simbolismo y conexionismo dentro de una estructura conceptual de inspiración pluralista.

Por lo anterior, dado que el *quid* es la búsqueda de una sistematización panorámica del intercambio de ideas, modelos y teorías en ciencias cognitivas, entonces el argumento principal va en línea con el esquema comparativo de IC, no haciéndose pertinente ir al pormenor de las ventajas y desventajas que ya han sido defendidos por el propio grupo PDP en su momento (Rumelhart & McClelland, 1986; tratados superficialmente en la sección (1.1.)) y refinados como caballito de batalla contra el cognitivismo por autores posteriores (por ejemplo, Churchland (1989)). Esto vale la pena recalcarlo, ya que, usualmente, el debate entre el simbolismo y el conexionismo tiene como principio y fin la discusión sobre esta lista de ventajas y desventajas ya más o menos establecida implícitamente. Al contrario, mi meta va más allá de acortar o ampliar tal lista, sino que va por el lado *de responder por el sentido y la pertinencia de tales discusiones*.

En mundo ideal el presente trabajo debiese haber ido acompañado de anexos que trataran con alguna profundidad ciertos tópicos útiles para el desarrollo de la tesis, tales como conocimientos básicos en neurofisiología (teoría de la neurona, histología del sistema nervioso, organización y principales estructuras, así como también los métodos de exploración del cerebro), filosofía de la mente (los problemas clásicos y las diferentes posturas en competencia) y filosofía de las ciencias (su historia desde el positivismo, pasando por el historicismo hasta desembocar en su estado actual), pero por cuestiones prácticas tales anexos no han sido hechos.

La tesis puede ser malinterpretada de diversas maneras: por ejemplo, considerándola como un manifiesto contra toda forma de modelado simbólico y como una reivindicación de que el enfoque emergentista conexionista es el único válido sin necesidad de complementario. También malinterpretada como una defensa enrevesada de un anarquismo epistemológico solapado. O, también malentendida como un ataque contra toda la filosofía de la mente o de las ciencias anteriores a este punto. No obstante, mi credo de fondo es más humilde y minimalista: *parto de la conjetura que toda idea encarnada en un modelo, una teoría o cualquier otro instrumento conceptual tiene su valor intelectual dentro de su propio contexto temporal y disciplinario bien delimitado*, y que sostener esto no implica necesariamente caer en un anarquismo. El texto total de la tesis es el resultado al que he desembocado a partir de tal credo de fondo.

I.- Las ramificaciones de la polémica simbolismo-conexionismo

Esta primera parte tiene como objetivo presentar el tema que da título a la tesis y mostrar cómo una revisión del contenido de la polémica y de la literatura ulterior que trata los tópicos derivados de la polémica, conduce a la necesidad de un replanteamiento de la misma. Esta reorganización se hará en la segunda parte de la tesis y dejará de lado buena parte de los detalles que a continuación se expondrán. Así pues, el objetivo de la segunda parte será entender los contenidos del debate desde una óptica que tenga en cuenta, tanto los avances contemporáneos en el campo de filosofía de las ciencias, como una historia un poco más detallada de las ciencias cognitivas, y las primeras redes artificiales (o “Artificial Neural Networks” – ANN) específicamente la neurona de McCulloch-Pitts y el perceptrón de Rosenblatt.

1.1. Presentación de la polémica

La aparición de un modelo o una teoría nueva explicando un fenómeno más o menos ya conocido dentro de una disciplina, nunca va exenta de críticas por parte de los defensores de teorías preexistentes. El caso del perceptrón de Rosenblatt (1958) – precursor directo del conexionismo y considerado como esencial para la comprensión del mismo en, prácticamente, todas sus introducciones más básicas – es bastante especial, no sólo porque no es criticado por un defensor de otro “paradigma”⁵, al contrario, es rebatido por un investigador famoso por sus aportes al propio campo de las nacientes ANN⁶, sino que porque también la obra *Perceptrons* (Minsky & Papert, 1969), centrada en criticar el perceptrón de Rosenblatt, según *la historia estándar del conexionismo*, causará una total caída en la investigación centrada en las ANN. Será sólo después de la aparición de los trabajos del grupo PDP, compuesto por personajes tales como David Rumelhart, James McClelland, Geoffrey Hinton, Paul Smolensky, Terrence Sejnowski, Michael I. Jordan, Jerome Feldman entre otros, que sobrevendrá el resurgimiento de la popularidad de las ANN aplicadas al campo de la explicación psicológica y la predicción conductual.

Pues bien, a pesar de la importancia histórica del impasse de *Perceptrons* en el campo de las ANN para el conexionismo (Dreyfus & Dreyfus, 1988), la polémica entre el simbolismo y el conexionismo será historiada bastante posteriormente, con tres momentos fundamentales: el primero, que instigaría la respuesta del simbolismo, es la exposición teórica de las ventajas del conexionismo presentados en los volúmenes del grupo PDP (Rumelhart & McClelland, 1986; McClelland & Rumelhart, 1986). El segundo momento, se da con la publicación de la edición especial titulada *Symbols and Connections* de la revista *Cognition*, específicamente se presentará la crítica de Fodor y Pylyshyn (1988). El tercero, es el debate entre Fodor y Smolensky que se extiende desde una primera respuesta de Smolensky a Fodor en 1987 hasta la última réplica de Fodor en 1997. Considero que el primer paso lógico a seguir es descomponer los momentos centrales de la polémica simbolismo-conexionismo para, posteriormente, mostrar su evolución (sección (1.2.)). Ya sintetizado un punto de partida teórico en la sección (1.3.), lo primero dentro de la segunda parte será, o bien *encontrar*, o bien *construir* un *terreno común* en cs. cognitivas.

⁵ Este entrecomillado no es antojadizo; su explicación en (2.5.i).

⁶ La tesis de doctorado en matemáticas de Marvin Minsky publicada en 1954, *Neural Nets and the Brain Model Problem* trataba específicamente el tema, por otra parte su trabajo inmediatamente posterior fue colaborativo con otros investigadores afines al tópico, por ejemplo, *Learning in random nets* (1961) con Oliver Selfridge (Véase la bibliografía del sintético artículo de Minsky, 1963).

1.1.1. Primer momento – Las ventajas del enfoque PDP⁷ según los propios conexionistas.

Los dos volúmenes del grupo PDP tienen como gran objetivo popularizar el “conexionismo”⁸, una nueva *teoría* que es expuesta partiendo desde su marco general, pasando por los elementos psicológicos de los cuales se inspiraron, pasado también por el análisis formal de sus propios modelos, hasta la ejemplificación concreta de esta “meta-teoría” (Rumelhart & McClelland, 1986, 144-5). El conexionismo asumirá, contra el simbolismo, que la mente está compuesta por la actividad excitatoria e inhibitoria de una cantidad variable de unidades conectadas entre sí, formando una red de funcionamiento paralelo. Así pues, el conexionismo se auto-definirá como suponiendo que la memoria no está localizada, ni es en sí misma una estructura de datos estática, además de que el procesamiento no se dará secuencialmente.

Si tomamos el prefacio del libro y su organización veremos que, en lo que respecta a su cotejo con el simbolismo, será la parte uno, enfocada en la perspectiva general, donde se concentrarán las *comparaciones generales* entre ambos⁹. Estas comparaciones se encuentran desperdigadas a lo largo de los cuatro capítulos que componen esta primera parte del primer volumen PDP y van acompañadas de una presentación del marco general que subsume todo modelo que pertenecería a esta nueva teoría del conexionismo. Luego, antes de entrar en las comparaciones, las cuales constituyen el detonante del intercambio entre defensores del simbolismo y los conexionistas, vale la pena remitirse antes a la presentación del marco general conexionista, el cual descompone un modelo conexionista genérico en ocho elementos constituyentes los cuales procederé a detallar (op.cit., 45-54):

1.- *Un conjunto de unidades de procesamiento o “neuronas”*: de funcionamiento paralelo y sin una supervisión centralizada de su operación. Dada la *arquitectura* de la red diferenciada por capas o estratos, usualmente, estas “neuronas” se dividen en tres tipos: de entrada o *input*, ocultas – cuya aferencia y eferencia no viene desde ni va hacia el ambiente – y de salida o *output*. La “semántica” de ellas, es decir, cómo se interpretan en un modelo específico, viene dada por la construcción misma de la red y el fenómeno que se quiere explicar, ergo, varían desde una representación localista donde cada neurona representa, por ejemplo, un concepto, a otra donde cada unidad de procesamiento representa

⁷ “Parallel Distributed Processing” o *procesamiento en paralelo distribuido*. Para acortar se usará siempre “PDP”.

⁸ El origen del término aplicado a la psicología se debe, al parecer, por primera vez a Edward Thorndike (Medler, 1998). Por otra parte William James, más de tres décadas antes de Thorndike, es considerado el primero en tener las ideas y esquematizarlas (Anderson & Rosenfeld, 1988), pero ello es algo que no puede ser asentado definitivamente en primera instancia y requeriría una investigación concienzuda separada.

⁹ Las comparaciones específicas entre ambos se desarrollan cuidadosa y selectivamente a lo largo de los dos volúmenes. Considérense, por ejemplo, el ataque contra los esquemas, al menos a sus nociones que implican que es una estructura no emergente y preexistente en la cognición (McClelland & Rumelhart, 1986, cap. 14). También las objeciones contra Chomsky al atacar la existencia de un dispositivo específico de adquisición del lenguaje (LAD) que sería innecesario según los modelos PDP (McClelland & Rumelhart, 1986, cap. 18). Y, abogando por la existencia memoria distribuida, se seguiría de tal la no necesidad de etiquetas innatas para las categorías lo cual también iría contra el simbolismo (McClelland & Rumelhart, 1986, 189). De ambos tipos de comparaciones, solo me haré cargo de las no específicas, las cuales serán expuestas en la presente sección (1.1.).

cierto “microrrasgo” (Hinton, 1981). Tales unidades podrían ser consideradas como un conjunto de $N = \{u_1, \dots, u_i\}$ unidades de procesamiento simple, *nodos* o *neuronas artificiales*¹⁰.

2.- *Un estado de activación*: presente en cada unidad en un momento t dado, de modo que para u_1 su estado de activación sea $a_1(t)$, para u_2 sea $a_2(t)$ y así sucesivamente hasta $a_i(t)$. El estado de activación del sistema se representará por un vector de notación $a(t)$ de largo N , siendo ello de profunda importancia puesto que el procesamiento del sistema puede verse como igual a la evolución a través del tiempo de los vectores de actividad del conjunto de neuronas N . Los valores del vector pueden ser continuos o discretos, y pertenecer a un intervalo abierto o cerrado.

3.- *Una función de salida para cada neurona*: al igual que el estado de activación se representa como un valor correspondiente a cada unidad. También tendrá la forma de un vector para representar las funciones de salida de todas las unidades del sistema en un t dado. Simbolizadas como $f_1(a_1(t))$, ..., $f_i(a_i(t))$ cada una de ellas corresponde a la salida u output de cada neurona, ergo para cada $a_i(t)$ existe un $o_i(t) = f_i(a_i(t))$. El vector correspondiente se notará como $o(t)$. Esta función de salida es frecuentemente entendida como una *función umbral* que determinará cuándo la unidad se considerará activa de manera de afectar a sus vecinas excitándolas o inhibiéndolas.

4.- *Un patrón de conexión*: la entrada de una neurona se definirá como la *sumatoria de las entradas multiplicadas por cierto peso*, permitiendo esto que el patrón de conexión total pueda ser representado mediante la especificación de todos los pesos correspondientes a cada conexión existente en la red. Este patrón se representa por una matriz de pesos W , donde w_{ij} será la notación para una conexión que va desde la neurona u_j a la u_i y donde el valor absoluto de w_{ij} será la intensidad o fuerza de conexión. Por otra parte, si w_{ij} es positivo se subentenderá que la conexión será excitatoria, en caso contrario, de ser negativo se considerará como inhibitoria. Las matrices de pesos, usualmente, se asumen como separadas, es decir, una matriz inhibitoria y otra excitatoria, de modo que existirían matrices diferentes para distintos tipos de conexiones. Es mediante la matriz de pesos que pueden ser diferenciadas las diferentes arquitecturas de redes¹¹ (Rumelhart & McClelland, 1986, 57-60).

¹⁰ Tanto el término de “nodo” como de “neurona artificial” no son propios de la exposición que se hace en el capítulo 2 del primer volumen del grupo PDP. En particular, nótese que la “neurona artificial” se puede utilizar como elemento definitorio de una sub-clase de modelo neural-inspirado: las ANN. Las neuronas “biológicamente inspiradas” constituirán el componente básico de una sub-clase de modelamiento paralela a las ANN. El desarrollo de esta idea que atraviesa tanto implícita como explícitamente la extensa literatura de neurociencia computacional se dará en la sección (2.4.1.), empero no se apuntará más que muy superficialmente en el comienzo de tal.

¹¹ En el capítulo donde se exponen los elementos constituyentes básicos de una red conexionista no aparece explícitamente el término “arquitectura” para referirse al modo en que fluye la información dentro de una red neuronal artificial. Sin embargo, dentro de contextos expositivos posteriores en el tópico de las ANN, a las diferentes organizaciones jerarquizadas (o no) de las múltiples capas de unidades de procesamiento se les denomina como “arquitecturas” (Mehrotra et al, 1997, sección 1.3). Esta acepción de “arquitectura” debe ser cuidadosamente distinguida de la otra más abstracta de “arquitectura cognitiva o funcional” (Fodor & Pylyshyn, 1988, 4-5), ya que su relación puede ser, o bien de subsunción, o bien de disyunción exclusiva dependiendo si se toma una postura implementacionista del conexionismo o no (las diferentes relaciones posibles simbolismo-conexionismo son tratadas en Smolensky, 1988 y son expuestas dentro del tercer momento del debate).

5.- *Una regla de propagación:* será aquella regla que tome el vector de salida $o(t)$ y lo combine con la matriz de conexión para producir una entrada neta para cada tipo de entrada. Por ejemplo, net_{ij} será la entrada neta de tipo i en la unidad u_j . Así pues, tomándose como guía el ejemplo donde existen dos tipos de matrices de pesos, tendremos una $net_i(t)$ como el vector entrada neta “ i ” en t y una $net_e(t)$ como el vector de entrada neta tipo “ e ” en t . Luego, la regla de propagación será $net_e = W_e o(t)$, $net_i = W_i o(t)$. La complejidad de las reglas de propagación irá en directa relación con la complejidad de los patrones de pesos – i.e. matrices de pesos.

6.- *Una regla de activación:* es la regla que toma las entradas netas de cada tipo que afectan a cada unidad de procesamiento, las combina con otras y produce un nuevo estado de activación. En los casos más simples con una regla de activación F definida como función identidad se tendrá que $a(t+1) = W o(t) = net(t)$. No obstante, las más veces F es la función umbral, por tanto la entrada neta debe sobrepasar un cierto valor para contribuir a un nuevo $a(t)$. Puede ser determinista, es decir, se asume que $a_i(t+1) = 1$ si pasa el umbral y 0 si no lo pasa – en este ejemplo sería discreto y cerrado –, o bien puede ser estocástico, es decir, a cada valor de activación la regla determina ciertas probabilidades de activación. Cuando $a(t)$ toma valores continuos entonces F es un tipo de *función sigmoide*, en este caso la unidad puede *saturarse* alcanzando un mínimo o un máximo de activación¹². La clase más común de función de activación es la cuasi-lineal, por ejemplo: $a_i(t+1) = F(net(t)) = F(\sum w_{ij} o_j)$.

7.- *Una regla de aprendizaje:* dentro de una red conexionista el aprendizaje es la modificación de los patrones de conexión en función de la experiencia, siendo tanto la formación de nuevas conexiones como su pérdida casos especiales de modificaciones de las fuerzas de conexión existentes. Prácticamente todas las reglas de aprendizaje son variantes de la propuesta de Hebb en su libro *Organization of Behavior* (1949), cuya idea básica era que si una neurona u_i recibe una entrada de u_j , entonces, si ambas están fuertemente activadas, la fuerza de la conexión w_{ij} debería aumentar. Expresada más formalmente esta regla se podría traducir como:

$$\Delta w_{ij} = g(a_i(t), t_i(t)) h(o_j(t), w_{ij}),$$

donde $t_i(t)$ es una entrada que *enseña* a u_i . Esta ecuación dice que el cambio en la conexión desde u_j a u_i viene dada por el producto de una función $g(\)$ dependiente de la activación de u_i y la entrada que enseña t_i y otra función $h(\)$ dependiente del valor de salida de u_j y la fuerza de conexión w_{ij} . Las versiones más simples del aprendizaje hebbiano *no tienen enseñante* y las funciones g y h son tan solo proporcionales a sus primeros argumentos:

$$\Delta w_{ij} = \eta a_i o_j,$$

¹² Hay diferentes “formas” que toman las funciones de umbral dependiendo la especificación de la regla de activación que no serán tratadas en profundidad en el capítulo 2 que estoy exponiendo lo más ortodoxamente. Una fuente complementaria puede ser Anderson (2007, 74-7) donde se presentan esquemáticamente otras funciones de activación. La sigmoide frente a otras funciones de activación presenta ventajas como ser derivable en todos sus puntos, además de tener cierto parecido al modo en que los potenciales de acción se presentan (véase la figura 2.14 Anderson (1997, 77)). Por otra parte, se dice que “satura” en el sentido de ser una función aplastadora, es decir, transforma una entrada de valores que van desde un intervalo abierto a uno que va a un intervalo cerrado.

donde η es una constante de proporcionalidad representando tasa de aprendizaje. Sin entrar en más detalles, otras reglas de aprendizaje serían:

Regla delta o de Widrow-Hoff: $\Delta w_{ij} = \eta(t_i(t) - a_i(t)) o_j(t)$

Regla de Grossberg: $\Delta w_{ij} = \eta a_i(t) (o_j(t) - w_{ij})$

8.- *Un ambiente*: Comúnmente, el ambiente se define como una distribución estable de probabilidad del conjunto de posibles patrones o vectores de entrada independientes de las entradas pasadas y las respuestas pasadas del sistema. Se puede imaginar el listado del conjunto de posibles entradas al sistema como yendo desde 1 a M, luego el ambiente es el conjunto de probabilidades p_i con $i = 1, \dots, M$. Estos vectores pueden ser *ortogonales*, *linealmente independientes* o simplemente arbitrarios dependiendo de lo que el modelo PDP es capaz de aprender. Quizás vale la pena agregar que, dado que es el microrrasgo la semántica de los nodos en varios de los trabajos expuestos en los volúmenes PDP¹³, entonces la información del estímulo representada por un vector¹⁴ implica que donde el componente es positivo se interpretará que el microrrasgo se encuentra presente, siendo lo contrario cuando el input es negativo (McClelland & Rumelhart, 1986, 8-9).

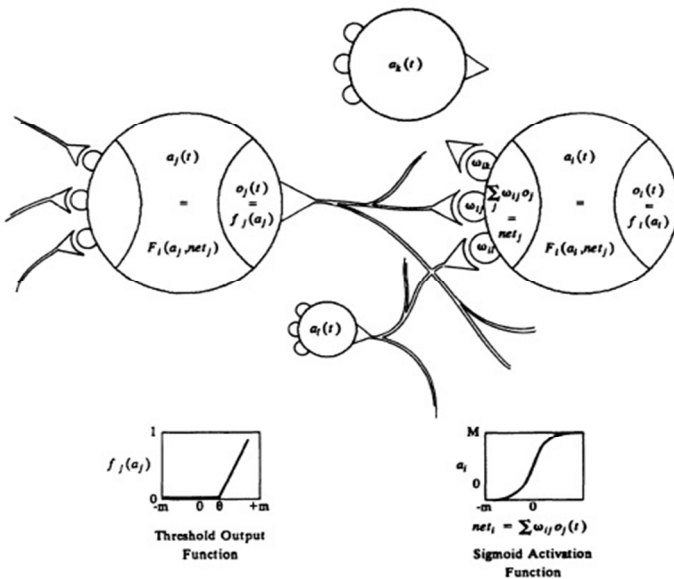


FIGURA 1; Aquí se pueden observar elementos tales como los diferentes estados de activación de varias unidades ($a_i(t)$, $a_j(t)$, $a_k(t)$ y $a_l(t)$), la sumatoria ($\sum w_{ij} o_j$) – o “entrada neta” – que se hace previamente a la aplicación de la regla de activación o la función de salida de la neurona $o(t)$. Sin embargo, esta es una representación pedagógica, ya que una representación bastante más acertada de lo que constituye un “estado instantáneo” en el tiempo del funcionamiento de un modelo cualquiera puede encontrarse en Rumelhart & McClelland, 1986, p. 50 donde es posible observar una matriz de pesos en un instante dado.

¹³ Digo “varios” y no “todos” porque hay modelos que usan representaciones locales como los presentados en (McClelland & Rumelhart, cap 15-16). No obstante es necesario agregar un matiz que el propio grupo PDP da a la distinción local/distribuida, a saber, que un mismo modelo es “local” y “distribuido” dependiendo del punto de vista, es decir, es local en cuanto cada unidad individúa un concepto particular, o una característica sensorial única, y a la vez es distribuida en cuanto que cualquier estímulo generará un patrón de activación sobre un número indeterminado de unidades.

¹⁴ El vector como entidad matemática carece de la semántica que lo caracteriza en la física, por ejemplo, dándose así también el caso en ciencias cognitivas, ya que un vector puede ser entendido tanto como una lista de números, un punto en el espacio, como una flecha, etc. (Rumelhart & McClelland, 1986, 374). Al final su significado puede ser distinguido exclusivamente en el contexto de su uso – y por extensión de toda el álgebra lineal que se esté ocupando – y es esencial para distinguir entre la utilización explicativa-predictiva de las ANN versus el uso meramente técnico al que se les reduce en campos como *machine learning* (por ejemplo, Shalev-Shwartz & Ben-David, 2014).

Los anteriores ocho elementos van complementados, posteriormente en el mismo capítulo, con una exposición sobre las clases de modelos PDP según diferentes *paradigmas de aprendizaje*, diferentes *organizaciones jerárquicas del procesamiento* y diferentes *momentos de actualización de la información*. Finalmente, se cerrará el capítulo con una muestra representativa de trabajos hechos hasta ese entonces en el campo de modelos PDP por investigadores conexionistas, articulándose tales modelos PDP como casos ejemplares de los conceptos expuestos a lo largo del mentado capítulo 2 del primer volumen.

Ahora bien, una manera de entender – bastante simplificada – la estructura argumental donde se introduce la contrastación entre el conexionismo y su predecesor, el simbolismo, es mediante una división en etapas: en una primera etapa, el grupo PDP aporta un par de *intuiciones cognitivas y fisiológicas esenciales* que, dado el contexto, se considerarán como primitivas. Segundo, muestran cómo estas intuiciones pasarán a ser *directrices de la generación* de los modelos PDP. En tercer lugar se presentan las *diferencias y contrastes* entre ambos tipos de modelamiento y, finalmente, en cuarto lugar, se presentan *posibles objeciones y sus respuestas*.

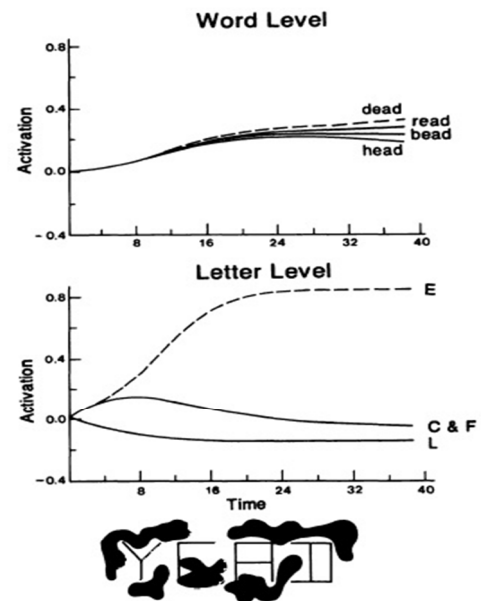
1.- *Intuiciones cognitivas y fisiológicas*: La primera intuición es de carácter fisiológico, y tiene que ver con que la divergencia entre la inteligencia mostrada entre el computador y los seres humanos, divergencia que iría, muy probablemente, más allá del “software”, enraizándose a nivel de “hardware” en arquitecturas computacionales totalmente opuestas (Rumelhart & McClelland, 1986, 3-4). Esta diferencia de arquitecturas quedaría evidenciada basándonos en el hecho intuitivo cognitivo de que somos expertos en ciertas tareas cotidianas, donde claramente no es el caso de los computadores, siendo tales tareas las que se caracterizarían por la existencia de *restricciones múltiples simultáneas*. Esta clase de tareas se ejemplifican en acciones como el alcanzar y agarrar un objeto sobre un escritorio entre varias cosas que impiden tomarlo fácilmente, en fenómenos de influencia mutua entre la sintaxis y la semántica para la comprensión de frases o en el reconocimiento de estímulos ambiguos, concretamente en el reconocimiento de palabras (op. cit., 4-9)

2.- *Directrices de la generación* de los modelos PDP: dadas las intuiciones anteriores, entonces se considerará que la forma en que diferentes fuentes de datos se conjuguen entre sí, más allá de las estructuras de datos rígidas, sea apelando a la existencia de *interacción* entre unidades de procesamiento simples, las cuales serán la base de la arquitectura computacional de un red de procesamiento en paralelo y distribuida (op. cit., 10). Por otra parte, dado que el soporte material es demasiado lento en comparación al sustrato electrónico de los computadores, y que además se trataría de unidades que actuarían paralelamente, entonces la manera de entender el pensamiento secuencial que aparece bajo la introspección, no debe partir suponiendo que tales estados mentales tienen una contraparte cognitiva que se ordenaría en serie, sino que serían una manifestación que *emerge* de la interacción. Con esto no se niega la “macroestructura” de la cognición, sino que se les transforma en aproximaciones al fenómeno “real” (op. cit., 12-13), el cual debiese ser analizado en una conjunción de las estrategias *bottom-up* y *top-down*, con prioridad de la primera¹⁵.

¹⁵ Esto quedará un poco más explicado en “Posibles objeciones al conexionismo y sus respuestas” dentro de esta misma sección.

Ejemplos un poco más concretos, aunque sin entrar en los detalles técnicos, serían casos de tareas con restricciones múltiples simultáneas (op. cit., 13-25): dentro del grupo de tareas de control motor se modela el escribir a máquina y el que una persona recoja un objeto sin caerse, siendo ambos, además de tareas tipo restricciones múltiples simultáneas, solucionados mediante una red donde interactúan unidades simples de procesamiento. También, dentro de esta clase de tareas, se encuentra el completamiento perceptivo de patrones familiares, específicamente se toma el caso del reconocimiento de palabras, versus el completamiento de patrones nuevos, mostrándose que en este segundo caso existe "(...) the possibility that PDP models may provide more accurate accounts of the details of human performance than models based on of rules representing human competence – at least in some domains." (op. cit., p. 24-5). La ejemplificación general con estos modelos lleva al tercer momento donde se contrastarán de manera general simbolismo y conexionismo.

FIGURA 2 (Rumelhart & McClelland, 1986, 25); Este caso representa una tarea donde se presenta un patrón nuevo, "YEAD", que activará, sin predominar ninguno, patrones que tomen _EAD y se acerquen a palabras conocidas como HEAD, DEAD o READ. Además, dado que no hay palabras que calcen con una "F" en el segundo lugar del patrón nuevo, entonces se suprime la interpretación que en un comienzo compite con "E". El hecho que el modelo PDP de percepción de las palabras exhiban facilitación perceptual a palabras no existentes, así como a aquellas existentes, ilustraría cómo la conducta de acuerdo a principios generales puede emerger a partir de interacciones de elementos simples de procesamiento, contrastando ello con explicaciones a partir de reglas específicas al lenguaje como tarea modelada separadamente de la percepción.



3.- Diferencias y contrastes entre simbolismo y conexionismo:

- *Degradación elegante en la recuperación de memorias* (op. cit. 25-9):

No hay duda que cuando tratamos de recordar algo, tal búsqueda es guiada a partir de ciertos atributos de lo que estamos tratando de recordar, aunque un atributo por sí solo puede ser insuficiente para llegar a encontrar lo que buscamos en nuestra memoria. Es por ello que será gracias a la conjunción de más y mejores pistas que, finalmente, llegaremos a recordar lo que esperábamos. En un computador estándar es posible implementar el rasgo de *ser guiado por el contenido* de la memoria mediante algún indexado, el cual será una lista de las características de cada uno de los recuerdos contenidos en la memoria. Sin embargo, tal forma de recuperar "recuerdos" generará errores si la pista ingresada es errónea y, aunque existen maneras de sobrellevar tales problemas, estos a la larga podrían acarrear explosiones combinatorias. Una red PDP puede ser alimentada con pistas parciales o defectuosas y, aun así, recuperará el contenido que se estaba buscando, característica que lo separa de los modelos simbólicos, no sólo en cuánto supera los problemas de un indexado, sino que logra capturar el fenómeno de la memoria humana guiada por el contenido.

- *Generalización espontánea* (op. cit., 30-1):

En un sistema simbólico puede darse que se almacenen generalizaciones explícitamente acerca del contenido que está siendo especificado mediante una lista, no obstante un modelo PDP tiene la propiedad de recuperar aquello que es común a todos los “recuerdos” – o a cierto conjunto de ellos – según alguna pista de recuperación demasiado general para indicar un ejemplar específico. En estos casos la ventaja del sistema PDP es que no necesita ser programado para producir generalizaciones sobre el contenido, sino que tales generalizaciones se producen automáticamente gracias a la misma naturaleza de una pista o input correcto y suficientemente general.

- *Representación de la información* (op. cit., 31-3):

Ya se ha apuntado la existencia de estructuras rígidas representando la memoria en un modelo simbólico, pero falta distinguir la existencia de dos tipos de memoria: una operativa o de trabajo y otra de largo plazo, desde la cual se extrae la información hacia la primera, mediante un proceso de copiado temporal del “recuerdo”. En el caso de una red conexionista no existe un almacenamiento estático, ya que la memoria son los cambios que ocurren en la matriz de pesos debido a la introducción de nueva información codificada vectorialmente, por lo tanto no existe una diferencia entre el procesamiento y la memoria, como sí la hay en un modelo simbólico. Por consiguiente, la memoria es activa al punto donde nunca se recuerda la copia exacta de lo sucedido, sino que meramente se *reconstruye* el pasado¹⁶.

- *Definición del aprendizaje* (op. cit., 31-3):

Por aprendizaje, de modo amplio, se entiende el proceso de entrenamiento que atraviesa cualquier red conexionista hasta lograr ajustar su matriz de pesos de modo de dar la respuesta que se está buscando. Este proceso puede darse dentro de diferentes paradigmas de aprendizaje¹⁷, los cuales serán formalmente bastante disimilares entre sí – i. e. diferentes reglas o algoritmos de aprendizaje, diferentes arquitecturas, etc. – y, por ende, dependerá tanto de la manera en que se ha diseñado el sistema como del *proceso variable de interacción de la red con el ambiente*¹⁸. En cambio, por lo general, un modelo

¹⁶ Esta última idea de que las representaciones distribuidas reproducen el recuerdo de un modo compatible con una fabulación, o “reconstrucción plausible”, no pertenece al capítulo que se está analizando, sino que es parte del capítulo 3 de la obra (Rumelhart & McClelland, 1986, 79-81)

¹⁷ El grupo PDPD distingue entre: (i) Aprendizaje asociativo, en el cual un conjunto de unidades cualquiera aprende a producir un patrón particular de activación siempre que otro patrón de activación cualquiera ocurre en otro conjunto de unidades, es decir, la asociación de un vector arbitrario con otro vector. Y (ii) descubrimiento de regularidades, donde las unidades aprenden a responder a patrones “interesantes” de input. En tal paradigma de aprendizaje no requiere “entrada enseñante” como en el anterior, sino que es una función de enseñanza interna que permite que sin supervisión la red aprenda a detectar rasgos (Rumelhart & McClelland, 1986, 54-7).

¹⁸ Esta segunda idea no es especialmente destacada en este capítulo, pero puede vislumbrarse fácilmente en la discusión sobre la relación entre los modelos PDP con el debate empirismo-innatismo (op.cit., p. 139-142) y, además, en algunos ejemplos que se dan a lo largo del trabajo del grupo PDP. Considérese el modelo sobre la tarea de interpretación del cubo de Necker (McClelland & Rumelhart, 1986, 9-13), donde se muestran la evolución de tres simulaciones diferentes, dos de las cuales pueden dar cuenta del fenómeno de la ambigüedad de interpretación del cubo y una tercera, anómala psicológicamente hablando, donde se superponen a la vez dos caras frontales en el cubo sin predominar ninguna. Lo interesante aquí es que las diferentes interpretaciones del cubo son dependientes de los nodos donde se produjeron las entradas, ergo del *ambiente*. Artículos posteriores de

simbólico se limita a tener como meta la formulación de reglas explícitas (proposiciones, producciones, etc.) que sean capaces de capturar generalizaciones, siendo estas últimas el contenido mismo del aprendizaje, como por ejemplo hace ACT*. Además, nótese que tales modelos para su funcionamiento requieren bastante más elementos innatos que las redes conexionistas.

4.- Posibles objeciones al conexionismo y sus respuestas:

- Los modelos PDP son muy débiles (op. cit., 111-3):

Normalmente, se considera que los modelos “tipo perceptrón” fueron refutados definitivamente con la crítica presentada por Minsky y Papert en *Perceptrons* (1969), sin embargo para Rumelhart y McClelland esto no es así, ya que la crítica se centraba sobre las limitaciones del perceptrón de un estrato e, históricamente, la crítica confluyó con algunos éxitos tempranos en el procesamiento simbólico, lo que significó rápidamente el descrédito de las nacientes ANN. Al no haber tratado con sistemas multiestrato ni con bucles de realimentación, puede considerarse de inmediato que no aplica la crítica de Minsky y Papert a los modelos usados por el grupo PDP. No obstante, se puede ir aún más lejos, como lo hacen los propios conexionistas, y exponer cómo la regla delta generalizada, en tanto algoritmo de aprendizaje, es capaz de llenar el vacío que existía para los sistemas multiestrato en la época del perceptrón de Rosenblatt (op. cit., p. 321-2)¹⁹ y ser capaz de completar los problemas postulados en *Perceptrons*.

- El problema de la equivalencia del estímulo (op. cit., 113-6):

Una letra siempre es reconocida a pesar de estar en diferentes orientaciones, esto se denomina “equivalencia del estímulo”. En el caso de los modelos simbólicos este problema se supera mediante la normalización de la entrada, sin embargo tal proceso no es muy claro en una red PDP. Ante esta situación se postula con detalle una red que normaliza ciertas entradas visuales demostrándose la posibilidad de normalización del input. Tal red, que sirve a modo de réplica para esta posible objeción, tendría dos conjuntos de detectores de rasgos, uno primero que serviría como dominio del mapeo de un segundo conjunto canónico de referencia, para luego producir el reconocimiento en la salida.

- Recursión (op. cit., 119-120):

Es una característica importante de los computadores la posibilidad de analizar recursivamente estructuras bien definidas, como por ejemplo son las oraciones, donde el anidamiento produce relaciones entre ciertos elementos que se encuentran a cada vez mayores distancias entre sí a nivel de la superficie a medida que la recursividad aumenta. Es una objeción común el decir que los modelos PDP no pueden tratar con estructuras recursivas. Los conexionistas responden con dos puntos importantes a

investigadores pertenecientes también al grupo PDP, como Jeffrey Elman, ponen especial atención a *la importancia del ambiente en el desempeño de la red conexionista* (Elman, 1993).

¹⁹ Este asunto es explicado con lujo de detalles en el capítulo 8 del primer volumen del grupo PDP. Empero, vale la pena apuntar un poco más sobre las dos etapas que conlleva la aplicación de la regla delta generalizada: en la primera etapa la entrada se presenta y se propaga hacia delante de la red, calculándose así el valor del output de cada unidad. La segunda etapa consiste en comparar el output con el vector que se deseaba como salida de modo de generar una señal de error (o delta) para cada unidad de salida. Esta señal de error se propaga hacia atrás en la red generándose de este modo los cambios en la matriz de pesos. (Rumelhart & McClelland, 1986, 127)

tener en cuenta: primero, es posible construir un sistema computacional arbitrario únicamente acudiendo a unidades de umbral lineales, por lo cual es posible construir un dispositivo conexionista que implemente una máquina de Turing – obviamente sin una memoria infinita al ser físicamente imposible – de lo que se sigue que posea la característica de recursividad. Segundo, en sus palabras: “We have not dwelt on PDP implementations of Turing machines and recursive processing engines because we do not agree with those who would argue that such capabilities are of the essence of human computation.” (op. cit., p. 120). Los conexionistas no estarían interesados en lidiar con la recursividad, ya que tal característica se basa en el análisis exclusivamente sintáctico del lenguaje, dejando de lado las claves contextuales, semánticas y pragmáticas del lenguaje. Para ellos lo que sería interesante sería construir un dispositivo que partiera con todas estas esferas interactuando bajo mutuas restricciones.

- Los modelos PDP no son cognitivos (op. cit., 120-1):

La reticencia de usar los modelos cognitivos basados en reglas explícitas ha sido considerada como una debilidad de los modelos PDP, ya que tal proceder les quitaría la categoría de “cognitivos”. No obstante, se responde que es necesario distinguir entre “regla” y “regularidad”. Las “reglas” estarían ahí para explicar las regularidades y las regularidades observables por sí solas no implican que haya aplicación de una regla, siendo tal aplicación ni menos ni más “cognitiva” que la explicación emergentista que apela a la interacción de unidades simples. Los modelos conexionistas pertenecen a las ciencias cognitivas en tanto: primero, tratan de explicar fenómenos mentales a través de los mecanismos que les subyacen. Segundo, se ocupan del problema de la representación interna, el cual va esencialmente contra la corriente conductista. Y, tercero, junto con la representación, en general se postula la existencia de un procesamiento mental, lo cual es exactamente lo que aspira explicar cualquier tipo de ciencia cognitiva.

- Los modelos PDP están en el nivel erróneo de análisis (op. cit., 121-7):

Se ha dicho que el conexionismo puede reducirse al nivel de implementación de acuerdo al esquema de Marr de los niveles de análisis²⁰ y que el nivel válido de la teorización cognitiva es el computacional. Para los conexionistas esto puede ser discutido, solamente, bajo la luz del juicio científico y no puede ser discutido *a priori* de modo puramente prescriptivo. No obstante, debe ser apuntado que aquí es difícil establecer la postura unívoca del conexionismo frente al simbolismo respecto a los niveles de análisis, ya que, por un lado, se declara que “as psychologists, we are committed to an elucidation of the algorithmic level” (op. cit., p. 123), ante lo cual la única variación con respecto al simbolismo sería el proceder desde una estrategia abajo-arriba complementaria. Al contrario, y por el otro lado, se considera que a veces el fenómeno cognitivo puede ser tan complejo que no pueda ser representado por una red PDP, ante lo cual el simbolismo – con sus conceptos de símbolo y regla, y su diferenciación explícita entre procesamiento y almacenamiento – sería muy útil para salir del paso, subentendiéndose que *queda al*

²⁰ Parte de aquello que dejaré en el tintero en este punto expositivo es una exposición más detallada de los niveles de análisis (Marr, 2010 [1982]) en este punto de la exposición de la tesis. En esta nota dejaré consignados cuáles son tales tres niveles de análisis, desde arriba hacia abajo: *Nivel Computacional*, se pregunta por cuál es la función cognitiva a ser computada, el *Nivel representacional-algorítmico*, que se pregunta por la representación de la entrada y salida, además del algoritmo que da lugar a la transformación de tal representación, y el *Nivel de implementación o instrumentación*, donde se ve en la práctica cómo se lleva a cabo la representación. Me detendré un poco más en cómo introduce David Marr sus niveles de análisis en la sección (2.1.).

arbitrio del fenómeno si es el caso que simbolismo y conexionismo entran en una relación de competencia o de complementariedad. Inclusive, es posible encontrar esta complementariedad supuesta explícitamente cuando se dice que (cito extensamente debido a la importancia del contenido):

Before examining the detailed arguments in favor of distributed representations, it is important to be clear about their status within an overall theory of human information processing. It would be wrong to view distributed representations as an *alternative* to representational schemes like semantic networks or production systems that have been found useful in cognitive psychology and artificial intelligence. It is more fruitful to view them as one way of implementing these more abstract schemes in parallel networks, but with one proviso: Distributed representations give rise to some powerful and unexpected emergent properties. (...)There we will see clearly that schemata and other higher - level constructs provide only approximate characterizations of mechanisms which rely on distributed representations. Thus, the contribution that an analysis of distributed representations can make to these higher-level formalisms is to legitimize certain powerful, primitive operations which would otherwise appear to be an appeal to magic. (op. cit., 78-9)

Lo único que parece ser seguro – dado que no es clara la postura que varía desde un cuasi-implemencionismo (una lectura del párrafo anterior basta), pasa por la “complementariedad” y acaba en una postura de “teoría en competencia” –, y gracias a que es reiterado en más de una ocasión a lo largo de las exposiciones del grupo PDP, es la afirmación de que las restricciones tipo abajo-arriba han sido subestimadas debido a que la noción de nivel de análisis ha sido inadecuada. Para ello se presentan dos analogías: la primera de la relación entre lenguaje de programación de nivel inferior versus los de nivel superior y la segunda de la relación entre mecánica newtoniana y la teoría cuántica. Presentando aquí solo la segunda analogía, la idea es que la mecánica clásica es una aproximación a fenómenos esencialmente microscópicos, a partir de los cuales emergen los fenómenos macroscópicos, siendo finalmente la teoría cuántica adonde se acude en caso de fallos explicativos de la teoría newtoniana.

- Son reduccionistas (op. cit., 127-9):

Se afirma como objeción que la psicología requiere su propio lenguaje, y que el enfoque PDP lo estaría reduciendo al renunciar al lenguaje propio del simbolismo al adoptar terminología cercana a la neurociencia. Contra esto se replica que el proyector PDP no es reduccionista, ya que presentaría la cognición como una propiedad que emerge de las interacciones de unidades conectadas en red. Se presenta una analogía para defender este punto: el diamante no puede ser entendido mediante el estudio de los átomos aislados, sino que se requiere un vocabulario de orden superior que permita la comprensión de los fenómenos mayores, por tanto la propiedad de la dureza del diamante requiere un vocabulario que apele más allá de los átomos de carbono, es decir a las estructuras que forman entre sí.

- No se sabe suficiente neurociencia para constreñir las teorías cognitivas (op. cit., 129-136):

Tradicionalmente, se supone que los hechos de neurociencias poca o ninguna restricción imponen sobre las teorías cognitivas – i. e. teorías que son definidas por el nivel computacional. Dado este estado de cosas, lo único que resta sería el análisis arriba-abajo que guiará la investigación al descubrimiento de los

mecanismos de la cognición de forma autónoma a sea lo que sea descubierto sobre el cerebro y su funcionamiento. Rumelhart y McClelland consideran que se han *inspirado neuronalmente* y que *no debe reducirse su enfoque en el modelamiento exacto neuronal*, sino que como un complemento abajo-arriba de la estrategia usual que parte desde el nivel computacional. Esta inspiración deviene de diversas *características neuronales* que son tomadas más o menos en cuenta dependiendo el modelo PDP que se trate: Primero, las neuronas son relativamente lentas, ya que su funcionamiento de milisegundos palidece comparado con el de un computador medido en nanosegundos. Segundo, existe muchísimas neuronas, lo que conduce a conjeturar que es el extremo paralelismo del cerebro lo que explica la sofisticación cognitiva humana. Tercero, las neuronas tienen grandes abanicos tanto de entrada como de salida de información con estimaciones desde las 1.000 a las 100.000 sinapsis, tanto de aferencia, como eferencia. Cuarto, el aprendizaje va ligado a la modificación de conexiones, de lo que se sigue que *el conocimiento está almacenado no en las neuronas, sino que en las mismas conexiones*. Quinto, existen dos modos de interactuar entre las neuronas: la excitación y la inhibición. Sexto, las conexiones en el cerebro parecer tener una geometría y topología claras, esto significa considerar como punto de partida un localizacionismo moderado como una visión de las diferentes redes especializadas. Séptimo, las neuronas están continuamente proporcionando un output, es decir, el estado de activación de una neurona refleja constantemente las entradas que se están produciendo en ellas, traduciéndose esto en que los nodos son pensados como proporcionando la evaluación continua de la plausibilidad de una hipótesis – o microrrasgo – y chocando esto con el enfoque habitual donde el procesamiento se da por fases de decisión diferenciadas y secuenciales. Octavo, existe degradación elegante al daño y a la sobrecarga informativa, es decir, el funcionamiento de la red en su conjunto es afectado gradualmente según el daño – o sobrecarga de información – va aumentado en la red y nunca presentará detención crítica ante daño leve o moderado. Noveno, el control del flujo de la información no es centralizado, sino que distribuido, y esto ha sido confirmado cuando se observa las consecuencias de daños en el córtex²¹, donde los pacientes muestran una baja sensible en el rendimiento de las tareas donde se vea involucrada la parte dañada, pero nunca se observa un fallo crucial. Y, décimo, el cerebro como sistema *no debe ser concebido como calculando una solución, sino que buscando asentarse en cierta solución*.

- Los modelos PDP carecen de realismo neuronal (op. cit., 136-8):

Paradójicamente, existe otra forma de argumentar contra el grupo PDP, que es achacar a su modelos falta de detalles neurofisiológicos. Ante esto la respuesta es que ellos se mueven en un nivel entre los detalles de la neurofisiología y el nivel macroscópico cognitivo donde se movería el simbolismo, además de que su concepción del progreso del modelamiento va en línea *con aproximaciones y simplificaciones sucesivas de las funciones*. Se reconoce que se han dejado de lado al menos dos aspectos que muy posiblemente influyen en los rasgos del flujo de la información en el cerebro: primero, los pesos de conexión son asociados a *tasas medias de disparo*, pero la importancia de los propios potenciales de acción no se considera en los modelos PDP. Segundo, se deja de lado el fenómeno de neuromodulación, es decir, la acción de ciertas sustancias químicas no dispersadas por las neuronas, y que normalmente actúan por difusión, que cambian tanto el comportamiento presináptico como postsináptico de las

²¹ Sobra decir que se exceptúan las zonas del tronco encefálico, ya que cada una de las características que están siendo enlistadas son relevantes para las tareas cognitivas superiores (op. cit., 134-5), mas no las inferiores.

neuronas. No obstante, es claro que para el grupo PDP estas simplificaciones, aproximaciones o idealizaciones de las neuronas individuales, que servirían para acercarse a la emergencia de las funciones cognitivas, serían una clase de progreso análogo al que se hizo cuando se estudiaron las propiedades emergentes de los gases mediante idealizaciones de las moléculas individuales de gas.

- Son “tabula rasa” (op. cit., 139-142):

Otra crítica enfrentada es que se consideran los procesos de aprendizaje conexionista muy generales y dando poca o ninguna importancia a los mecanismos innatos cognitivos, por ejemplo, con respecto a los factores innatos del procesamiento del lenguaje. El contrargumento aquí es presentar a los modelos PDP como agnósticos en relación con al debate entre innatistas y empiristas. Un sistema conexionista puede ser adaptado para defender tanto un innatismo radical, como una empirista radical: en el primer caso imponiéndose al sistema una determinación rígida en las interconexiones entre los nodos desde el nacimiento, y en el segundo caso suponiendo que el patrón de interconexión no posee ninguna restricción en su nacimiento y todo puede ser aprendido. Es obvio para los conexionistas que la real naturaleza de lo cognitivo se mueve entre los extremos y que no existe asidero para considerar las redes PDP como “tabula rasa”. Ergo, la relación de un modelo PDP con su ambiente vendrá dada por la conjunción entre la arquitectura de la red – que no se debe confundir con una “programación” en el sentido simbólico – y la caracterización formal que se haga del ambiente específico o, como es en otros términos “To the degree that different individuals share the same genetics (start state) and to the degree that their environments are similar, they will pass through similar trajectories.” (op. cit., 142).

- No dan cuenta de la diferencia entre animales y humanos (op. cit., 143):

También se ha aducido críticamente que, puesto que tanto la conducta de ratas, animales con funciones cognitivas exclusivamente inferiores, como la de las personas puede ser descrita a partir de redes PDP, entonces se carece de explicación de las diferencias que separan a seres humanos de otras criaturas. Este argumento se contesta bastante sucintamente con dos puntos: primero, los conexionistas no afirman que las redes sean homogéneas para todas las especies, más bien lo contrario. Y, segundo, tanto fisiológica, como culturalmente hay una separación entre ratas y personas, lo cual incide de manera esencial en la distinción entre ambos.

- No dan cuenta del razonamiento explícito y la consciencia (op. cit., p. 143-4):

Muchas veces es concedido por los investigadores clásicos que ninguno de los argumentos hasta aquí apuntados contra el conexionismo sea decisivo, empero aun así existe algo que el conexionismo no da cuenta, a saber, nuestra capacidad de usar conocimiento explícito de modo de guía de la conducta. Pueden ser útiles para dar cuenta del control motriz, la percepción u otra capacidad cognitiva inferior, pero las redes PDP son simplemente inútiles para explicar el nivel superior cognitivo. Este es el único argumento que los conexionistas han dejado a la espera de una mayor investigación futura, ya que, en las propias palabras de Rumelhart y McClelland “(...) we cannot and do not expect PDP models to handle complex, extended, sequential reasoning processes as a single settling of a parallel network” (op. cit., p. 144). Por consiguiente, todo trabajo del grupo PDP hasta ese entonces debe considerarse tentativo, dada aun la dispersión de modelos y falta de unidad en el tratamiento de lo cognitivo (op. cit., 145)

1.1.2. Segundo momento – La crítica de los cognitivistas desde el simbolismo

Esta crítica se publica en una edición especial, correspondiente a Marzo de 1988, de la revista *Cognition* y consta de tres artículos escritos de manera independiente, pero que confluyen sobre un mismo tópico, el cual es intentar analizar sistemáticamente los supuestos centrales del enfoque conexionista comparándolos con aquellos del enfoque simbólico que estaría siendo tratado de reemplazar, además de contrastarlos con aquellos hechos más visibles de la cognición humana. En palabras de Pinker y Mehler, “Now that many connectionist efforts are well known, it may be time for a careful teasing apart of what is truly new and what is just a relabeling of old notions; of the empirical generalizations that are sound and those that are likely to be false; of the proposals that naturally belong together and those that are logically independent.” (Pinker & Mehler, 1988, p. 2).

Este segundo momento del debate entre conexionistas y simbolistas será expuesto detallando exclusivamente uno de los artículos que componen la edición especial de *Cognition*, a saber, el artículo de Fodor y Pylyshyn, *Connectionism and Cognitive Architecture: A Critical Analysis*. Los otros dos restantes, primero, el de Pinker y Prince, *On language and connectionism : Analysis of a parallel distributed processing model of language acquisition* y, segundo, el de Lachter y Bever, *The relation between linguistic structure and a ssociative theories of language learning - A constructive critique of some connectionist learning models* no serán tratados, aunque el lector interesado en la contrastación específica del tratamiento conexionista versus el simbolista sobre el lenguaje puede recurrir a ellos como punto de partida para una investigación enfocada exclusivamente en el lenguaje.

Connectionism and Cognitive Architecture: A Critical Analysis, de Fodor y Pyslyshyn (F&P)

- *Primera Parte*

En la primera parte del artículo F&P comienzan notando que todo aquel que de uno u otro modo está en desacuerdo con la psicología cognitiva basada en los modelos de procesamiento de información ha ido adoptando el conexionismo como su caballito de batalla, aunque ello no signifique acuerdos más allá de la adopción de tal formalismo como argumento contra el simbolismo²² (op. cit., 4-6). Esta primera observación llevará a F&P, en un primer momento, a buscar aquello que es común con el grupo PDP y que tiene que ver con aclarar sobre en qué niveles de explicación se mueve el conexionismo.

Nótese que antes de entrar en materia, ellos desde el comienzo clarifican que tanto los “modelos conexionistas” – los cuales son implementados por una multitud de modelos variables en sus detalles – como los “modelos clásicos de la mente” – los cuales también pueden ser implementados en una gama

²² El término “simbolismo” no es usado para hablar de la clase de modelos que están defendiendo F&P, sino que se habla de “modelos clásicos” o “tradicionales”, esto se debe, seguramente, a que “simbolismo” es un término propuesto por Smolensky (1987b), quien dirá “I will refer to the traditional approach to cognitive modeling as the *symbolic paradigm*, intending to emphasize that in this approach cognitive descriptions are built of entities that are *symbols* both in the semantic sense of referring to external objects and in the syntactic sense of being operated upon by ‘symbol manipulation’.” (Smolensky, 1987b, 1). En cualquier caso consideraré sinónimos “simbolismo” y “modelos tradicionales” o “clásicos”, aunque posteriormente sí haga distinción con “cognitivismo”, específicamente en la sección final 2.5. cuando haga el análisis respecto a la “relación interteórica” entre simbolismo y conexionismo.

variable de modelos y cuya raigambre se remonta a la máquina de Turing y la arquitectura Von Neumann – tienen como fin el modelamiento de la *arquitectura cognitiva*²³ (op. cit., 4-5) y tales son definidos introductoriamente de manera amplia *sin referencia a ningún modelo específico*.

Ahora bien, *¿qué tienen en común los modelos conexionistas y los clásicos? Que ambos pertenecen a la tradición representacionista de la teorización de la mente* (op. cit., 7). Esta tradición se opone a la eliminativa, la cual considera que una teoría psicológica no requiere nociones como “representación” y que el vocabulario adecuado para la psicología puede ser subsumido por el neurológico, conductual o, quizás, el sintáctico. En cambio, la representacionista postulará “estados representacionales” (sinónimamente “intencionales” o “semánticos”) esenciales para la teoría de la cognición y que *se definen según su función de codificar estados en el mundo*.

Sin embargo, según la perspectiva de F&P hay vacilación en el representacionismo conexionista, ya que se autodefinen también como “subsimbólicos”, es decir, como adhiriendo a una visión donde la cognición emerge desde lo subsimbólico, lo cual parece ser inconsistente con afirmar que la propia investigación tiene como meta el nivel cognitivo (op. cit., 8-9). Dado que, se supone, se comparte el realismo representacionista entre clásicos y conexionistas, entonces lo que se busca es encontrar el nivel en el cual el sistema codifica las propiedades del mundo, y que puede ser llamado el “nivel cognitivo”. No obstante, no es útil en lo absoluto, desde el punto de vista del psicólogo cognitivo, mostrar que estados no representacionales de un organismo constituyen una red conexionista, ya que tal cosa dejaría abierta la pregunta de si la mente es tal red al nivel psicológico²⁴ (op. cit., 10).

Pues bien, aunque la cuestión de los niveles de explicación es confusa en el conexionismo, F&P rescatan que tanto los modelos clásicos como los conexionistas están comprometidos con el realismo representacionista. Por otra parte, se deja de lado la cuestión del “seguimiento de las reglas”, dado que, como se especificará en la siguiente parte del artículo, lo que caracteriza a los modelos clásicos no va en línea con la idea de reglas explícitas mediando la etiología de la conducta, sino que va más allá de esto que parecía ser de vital importancia como distinción según los propios conexionistas (op. cit., 11).

²³ La definición de “arquitectura cognitiva” no es unívoca y parece tener diferentes matices según cada autor. En este artículo una referencia a su noción se encuentra en la frase siguiente: “Put differently, the architecture of the cognitive system consists of the set of basic operations, resources, functions, principles, etc. (generally the sorts of properties that would be described in a “user’s manual” for that architecture if it were available on a computer), whose domain and range are the *representational states* of the organism” (Fodor & Pylyshyn, 1988, 10). Esta observación será retomada en el contexto de la sección (2.5.iv).

²⁴ Lo que suponen F&P es que es la representación misma lo que distingue el nivel cognitivo del nivel no cognitivo y que los conexionistas, aun declarándose representacionistas, fallan en darse cuenta de esta distinción al declarar que lo que distingue a un sistema cognitivo de uno no cognitivo es la *complejidad* y no la *intencionalidad* (Fodor & Pylyshyn, 1988, 10 – y para ver la versión original Smolensky, 1987b, 15). Luego, es fácil concluir que, para que F&P afirmen que queda abierta la pregunta de si una red conexionista se mueve en el nivel psicológico, estarían comenzando de un principio que rezaría “de lo no representacional sin intencionalidad nada representacional se sigue”. Quizás, en defensa de Smolensky, puede aducirse que él sí era consciente de que los modelos simbólicos se definen a partir de procesos que tienen una semántica que se mueven en el mismo nivel conceptual y que, intencionalmente, decide seguir la clase de “reducción” o cambio de *dimensión* que es familiar en las ciencias naturales, es decir, buscar un nivel inferior que dé sentido al nivel superior, siendo tal nivel ni neuronal, ni conceptual (Smolensky, 1987b, 10-11).

- Segunda Parte

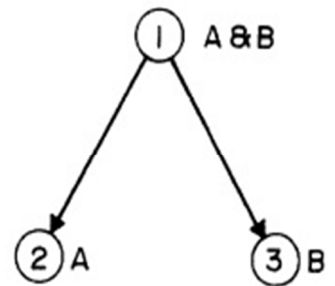
Si la diferencia entre arquitectura conexionista y clásica no tiene que ver con la existencia de reglas, ni con el estatus ontológico de los estados representacionales entonces ¿cuál es la naturaleza de la disputa? Según F&P el desacuerdo consiste en *adscribir diferentes tipos de relaciones primitivas entre las entidades portadoras de contenido*; por un lado las teorías conexionistas sólo consideran como relación primitiva la *conectividad causal* entre los nodos. Por el otro lado, las teorías clásicas no sólo consideran relaciones causales entre los objetos semánticamente evaluables postulados, sino que también relaciones *estructurales*, de las cuales la *constitutividad – constituency –* es paradigmática. Este desacuerdo se traduce en las siguientes diferencias arquitecturales (op. cit., 12-3):

[1] *Representaciones mentales formadas con sintaxis y semántica combinatoria*. Esto permite tres cosas; (a) Distinguir entre representaciones estructuralmente atómicas versus moleculares; (b) Las representaciones estructuralmente moleculares tendrían constituyentes sintácticos que, o bien son ellos mismo moleculares, o bien serían estructuralmente atómicos; (c) El contenido semántico de una representación molecular es una función del contenido semántico de sus partes sintácticas junto a su estructura constituyente. (a), (b) y (c) serán equivalentes a decir que una teoría clásica está comprometida con *representaciones mentales complejas* o con *estructuras simbólicas*.

[2] *Procesos con sensibilidad a la estructura*. Dado que las representaciones mentales clásicas tienen estructura combinatoria, entonces es posible para las *operaciones* mentales clásicas ser aplicadas por pura referencia a la *forma de las representaciones*. El resultado de esto es que un proceso mental opera sobre *cualquier representación* que satisface cierta *descripción estructural* y la transforma en otra representación mental que satisface otra descripción estructural. Por ejemplo, la operación que se aplica sobre representaciones de la forma “P&Q” para producir “P”, es satisfecha también por una expresión “(AvBvC)&(DvEvF)” de la cual se derivaría “(AvBvC)”.

F&P aseveran que [1] y [2] en conjunto definen todo modelo clásico constriñendo las implementaciones físicas de las estructuras de símbolos y que son estas características estructurales, más las contrapartes físicas de los símbolos, las que *causan que un sistema cognitivo se comporte como lo hace* (op. cit., 14). Asentado esto, entonces lo que sigue es mostrar que las representaciones conexionistas fallan como representaciones mentales complejas.

Para argumentar lo anterior, específicamente el punto [1], se considera una red conexionista muy trivial que muestre las limitaciones de basarse puramente en la conectividad causal como relación primitiva (**FIGURA 3**, a la derecha). Para que la red infiera de A&B que A, entonces debe existir una conexión causal entre el nodo 1 y el nodo 2, pero en tal inferencia no existe una relación estructural (parte/todo) entre tales nodos (op. cit., 15-6). Luego, se necesitaría la característica [1] para que la inferencia en la red fuese hecha mediante relaciones más que causales, pero si se hiciese así es evidente que entonces la red estaría implementando una arquitectura clásica. Ahora bien, F&P comentan que esta diferencia básica ha sido pasada por alto por al menos cuatro motivos:



(i) *El rol de las etiquetas en las teorías conexionistas*. En las redes PDP no es raro observar que a los nodos se les asignan etiquetas que representarían el contenido semántico de éstos. Sin embargo, estrictamente hablando, tales etiquetas no cumplen ningún rol en lo absoluto para la operación de las máquinas conexionistas, dado que simplemente no hay conexión sintáctica ni semántica entre las etiquetas de los nodos (op. cit., 17-8).

(ii) *Redes conexionistas y estructuras de grafos*. Las redes parecen grafos y es perfectamente compatible el uso de grafos para describir la estructura interna de un símbolo complejo. De hecho, eso es lo que hacen los lingüistas cuando usan “árboles” para exhibir estructura constituyente de oraciones. Empero, aunque se parecen superficialmente esta interpretación es inadecuada para los grafos conexionistas, ya que, nuevamente, la red conexionista sólo especifica relaciones causales entre sus nodos, pero nunca algo así como “X está gramáticamente conectado con Y”, como si sucede en un “árbol” (op. cit., 18-9)

(iii) *Las representaciones distribuidas*. Que algo tenga partes, como puede ser observado en una representación distribuida, no implica que tenga constituyentes sintácticos o semánticos. Además, aquí puede observarse el problema de los niveles de explicación, ya que la relación en una arquitectura clásica entre una representación molecular y una atómica es *dentro del mismo nivel* psicológico, mientras que en una arquitectura clásica la relación entre los nodos y las representaciones distribuidas es *entre niveles* (op. cit., 19).

(iv) *Representaciones como distribuidas sobre microrrasgos*. Esta crítica se visualiza mejor a través de un ejemplo: considérese el microrrasgo “tiene-un-asa” de la representación distribuida de “taza” y el nombre “María” como constituyente de la oración “Juan ama a Mary”. En el primer caso “tiene-un-asa” no está en una relación parte/todo como sí lo está “María” con respecto a la oración. La expresión “tiene-un-asa” es parte de la expresión “taza” en la misma medida que la frase “es un hombre no casado” es parte de otra oración como “es un soltero”, es decir no es constituyente, sino que a lo más pertenecería a una relación de tipo conjunto/subconjunto (op. cit., 21-2). El problema más grave para F&P de aceptar un tipo de relación conjunto/subconjunto, del conexionismo, como reemplazante de la relación parte/todo propio de las representaciones complejas mentales, es el explosivo aumento que tendría que tener el conjunto de microrrasgos primitivos. La alternativa de esta grotesca explosión sería precisamente una sintaxis y semántica combinatoria para los rasgos, pero claramente esto significaría instrumentar una arquitectura clásica (op. cit., 24).

Luego de detallar los problemas del conexionismo en implementar [1] se pasa a la característica arquitectural [2], a saber, la existencia de operaciones sensibles a la estructura. F&P afirman que para el conexionismo no está disponible la noción de descripción estructural o “forma lógico-sintáctica”, simplemente por el hecho de que sus modelos son incapaces de contener representaciones mentales compuestas (op. cit., 28). El tratamiento clásico de los procesos sobre representaciones descansa en dos ideas, las que en conjunto explican la existencia de tres niveles de organización computacional, no sólo dos como podría ser el nivel físico y el semántico, sino que agregando el sintáctico:

Primera idea (op. cit., 28-9): que es posible construir un lenguaje en el cual ciertas características de estructuras sintácticas de sus fórmulas se correspondan sistemáticamente con ciertas de sus

características semánticas, es decir, que la sintaxis de una fórmula codifique su significado, específicamente, aquellos aspectos del significado que determinan su rol en la inferencia. Este lenguaje sería cierto *lenguaje del pensamiento* (LOT)²⁵. Un ejemplo que ilustra esta característica de tal LOT sería:

(1.) Juan fue a la tienda y María fue a la tienda.

(2.) María fue a la tienda.

Por una parte, desde el punto de vista semántico, (1.) implica (2.) – de hecho la inferencia desde (1.) a (2.) preserva la verdad – y por otra parte, desde el punto de vista sintáctico, (2.) es constituyente de (1.). Estos dos hechos pueden ser conjugados aprovechándolos en un principio que especifique que las oraciones con la estructura sintáctica “(S1 y S2)” impliquen sus constituyentes oracionales. Nótese así que este principio conecta la sintaxis de estas expresiones con sus roles inferenciales.

Segunda idea (op. cit., 30): es posible diseñar e implementar máquinas cuya función sea la transformación de símbolos y con operaciones sensibles a la estructura sintáctica de los símbolos sobre los cuales están operando. Estas máquinas diseñadas clásicamente funcionarían de manera de satisfacer un criterio semántico de coherencia, aunque serían guiadas solamente sintácticamente.

Dado que estas dos ideas dan forma a las operaciones dentro de la arquitectura clásica debiese bastar mostrar cómo se distinguen de las operaciones en el conexionismo mediante ciertas tareas:

(i) Aprendizaje (op. cit., 30-1): en una red el aprendizaje depende del entrenamiento que reciba, lo que lleva a que las relaciones internas de la red recapitulen los eventos ambientales a los cuales se le sometió. Esto recuerda al viejo asociacionismo, y la única diferencia que tienen las redes PDP según F&P sería la presencia de unidades ocultas las cuales detectarían patrones estadísticos sobre las unidades “visibles”. Lo importante aquí es que todos estos procesos son sensibles a la frecuencia y esto las equipara al asociacionismo tradicional.

(ii) Razonamiento (op. cit., 31-2): La salida de una función es una especificación del patrón de excitación sobre los nodos de salida, lo que intuitivamente significa que la máquina elige el patrón de salida más altamente asociado a su entrada. Sin embargo, nuevamente, la estructura semántica/sintáctica de la representación de la entrada no se presume que sea un factor para la determinación de la salida determinada, ya que la estructura semántica/sintáctica de la representación no está definida.

- *Tercera Parte*

Los rasgos arquitecturales expuestos en la parte segunda del artículo de F&P tienen como fin explicar tres características interrelacionadas de la cognición: su *productividad*, su *sistematicidad-composicionalidad* y su *coherencia inferencial* (op. cit., 33). Estas tres características interrelacionadas se contrastarán con la propuesta conexionista, para posteriormente, en la siguiente parte, diseccionar el atractivo del conexionismo en relación con las arquitecturas clásicas.

²⁵ LOT (*Laguage Of Thought*) es diferente a cualquier lenguaje natural sencillamente por el hecho de que, por ejemplo, la conjunción (por extensión, también cualquier otra conectiva lógica) en LOT funciona de manera uniforme no siendo así el caso con un lenguaje natural.

-- Primera característica de la cognición, *Productividad del pensamiento.*--

En un sistema representacional sus capacidades representacionales se asumen que sean desencarnadas según una apropiada idealización, en particular, se postula que idealmente *existe una cantidad indefinida de proposiciones que el sistema sería capaz de codificar*. Esta capacidad de codificación potencialmente infinita se define como la productividad del sistema. Sin embargo, este poder expresivo desencarnado debe ser presumiblemente alcanzado por algún medio finito. Este medio finito sería un conjunto de representaciones que generaría las expresiones codificadoras, pero esta estrategia únicamente funcionará si las proposiciones son no atómicas. Un sistema simbólico será tal únicamente si y sólo si posee una sintaxis y una semántica combinatoria, en caso contrario, aunque un sistema contuviese estados representacionales, no sería suficiente tal característica para transformarlo en un sistema simbólico, el cual estaría constituido por estos elementos representacionales mínimos capaces de productividad (op. cit., 33).

La validez de la productividad como rasgo de la cognición humana puede ponerse en duda, pero F&P consideran que para demostrarla no es necesario recurrir exclusivamente a argumentos a priori a favor de capacidades productivas ideales, sino que se puede recurrir a una mezcla de consideraciones metodológicas y empíricas a su favor. Desde lo metodológico la idealización permite separar la especificación finita de un método para solucionar un problema computacional de los recursos reales del sistema físico (la persona) en estudio. Desde lo empírico existen pruebas relacionadas con la competencia lingüística, que muestran que cuando se extiende la memoria o el tiempo de procesamiento para alguna tarea el desempeño mejora, pero que ello no significa que el conocimiento de cómo la persona se representa el lenguaje aumenta, sino que precisamente mostraría la existencia de una competencia desligada de la actuación (op. cit., 34).

Para las arquitecturas clásicas estas consideraciones encajan perfectamente a la distinción funcional entre memoria y programa, como por ejemplo en una máquina de Turing donde el cambio en las capacidades de la memoria no afectaría la estructura computacional de la máquina. En cambio en una máquina conexionista, al ser un autómata de estados finitos, la adición de memoria (v. gr. mediante la adición de unidades a la red) altera las relaciones de conectividad entre los nodos y, por tanto, afecta la estructura computacional de la máquina²⁶ (op. cit., 34-5). Para F&P el test definitivo de la buena

²⁶ Aquí es importante advertir que es dudoso que las ANN puedan ser encajadas – homogeneizándolas formalmente – en una jerarquía que ponga a la máquina de Turing por sobre ellas, ello debido a que, aunque por ejemplo las redes de McCulloch-Pitts fueron probadas como equivalentes a un autómata finito (Piccinini, 2004, 197-9), ello no se puede aplicar a toda la familia de ANNs que han sido desarrolladas (Tabor, 2009; Siegelmann, 1999) las cuales varían ampliamente en sus poderes computacionales. Sin embargo, a pesar del gran interés para las ciencias formales que puede significar el trabajo sobre tal tópico (Zenil, 2005), en lo que reviste a ciencias cognitivas su valor es poco claro, relativo y muy discutible. En todo caso, en lo que concierne al argumento de F&P no es necesario que sea cierta su afirmación de que los modelos conexionistas son autómatas de estado finito para que se preserve la observación de que las ANN carecen de una distinción clara entre memoria y estructura computacional, ya que, como se puede derivar de un examen rápido, *no hay una estructura computacional mapeable exactamente desde una máquina de Turing a un marco conexionista general como el expuesto por los conexionistas*, de lo que se sigue que solamente basándose en este hecho sea posible afirmar que no existe

idealización es aquella que produce ciencia exitosa a largo plazo, y en la opinión de ellos, la idealización de la productividad ha sido útil en lingüística y teoría del razonamiento.

Sin embargo, al final, los argumentos a favor de la productividad los conceden como discutible, dado que se ha aceptado como una característica que requiere la idealización de una competencia desencarnada. Teniendo en cuenta que los conexionistas no consideran que la recursividad, derivada de la productividad, sea una esencial a la cognición humana, entonces F&P prefieren pasar a las otras características de la cognición (op. cit., 36-7).

-- Segunda característica, *Sistematicidad y composicionalidad de la representación cognitiva*²⁷.

Dado que sistematicidad y composicionalidad están tan cercanamente relacionadas, pueden visualizarse mejor como dos aspectos de un mismo solo fenómeno (op. cit., 41). Ahora bien, la principal manera de entender la sistematicidad es partiendo del dominio del lenguaje, aunque no suponiendo que la sistematicidad se circunscribe exclusivamente a tal dominio, sino que colándose por todos los rincones de la cognición humana. Por ejemplo, se dice que la capacidad lingüística es sistemática cuando *se tiene la habilidad para producir/entender algunas oraciones que están intrínsecamente conectadas a la habilidad de producir/entender ciertas otras*. El área del lenguaje es especialmente interesante para mostrar la diferencia entre una habilidad característicamente sistemática versus otra que no; considérese el aprendizaje de una nueva lengua, se dice que tal aprendizaje es sistemático cuando se es capaz de aprender cualquier trozo de un libro y con ello producir/entender cualquier nuevo trozo de texto en la nueva lengua, mientras que se dice que tal aprendizaje *no es sistemático* cuando, aun aprendiendo todo un libro de frases, se es incapaz de producir/entender otro novedoso (op. cit., 37).

La gran diferencia entre la productividad y la sistematicidad es que los argumentos para la primera están enfocadas en una capacidad desencarnada para entender oraciones arbitrariamente largas, mientras que los argumentos para la segunda están ligadas a experiencias muchísimo más cotidianas, como las que se observan en el uso del lenguaje natural, donde, por ejemplo, no existe un hablante nativo del español que produzca/entienda una frase que sea del tipo “Juan ama a la chica”, pero que sea incapaz de producir/entender la frase “La chica ama a Juan”. En palabras de F&P: “The assumption that linguistic capacities are productive “in principle ” is one that a Connectionist might refuse to grant. But that they are systematic *in fact* no one can plausibly deny” (op. cit., 39).

Puesto que en el conexionismo no existen representaciones mentales complejas, – como se mostró en la segunda parte del artículo con respecto a la característica arquitectural [1] – entonces la sistematicidad de las capacidades relacionadas al procesamiento del lenguaje pasan a ser un misterio, mientras que por

exactamente la misma distinción funcional de memoria y programa en una máquina de Turing que en una ANN cualquiera. Recuérdese que lo que permite que, por ejemplo una red McCulloch-Pitts sea considerada un autómata finito y así sea comparable a una máquina de Turing, es el *lenguaje* que procesan – o lo que es equivalente la complejidad de los problemas susceptibles de ser solucionados por ellas –, pero su organización interna no es tomada en cuenta (Alfonseca et al., 2006, 4).

²⁷ Aunque en el artículo originalmente van como secciones separadas la sistematicidad y la composicionalidad he preferido unir ambas secciones originales tanto por su extensión, como para mantener la coherencia de la propia introducción de F&P hecha para la tercera parte de su artículo.

el lado de las arquitecturas clásicas, dado que se postula una estructura constituyente de las representaciones, entonces la sistematicidad del lenguaje es un fenómeno fácilmente predecible (op. cit., 38). Luego, lo único que resta es mostrar cómo es característica pasa a ser parte de la cognición más allá de aparecer en el lenguaje, pero ello es rápidamente despachado considerando que, dado que se ha aceptado no se puede encontrar a alguien que produzca/entienda “Juan ama a la chica”, pero que no entienda “La chica ama a Juan”, entonces no puede existir alguien que pueda representarse mentalmente una situación compuesta en un modo lingüísticamente dado y que no pueda representarse la misma con los constituyentes en una estructura relacionada sistemáticamente (op. cit., 39).

Ahora bien, que las oraciones tipo “Juan ama a la chica” y “La chica ama a Juan” estén sistemáticamente relacionadas no es un hecho arbitrario desde el punto de vista semántico, y lo que explica esta no arbitrariedad es el principio de composicionalidad que puede ser definido como: en la medida que el lenguaje es sistemático, entonces *un ítem léxico debe tener aproximadamente la misma contribución semántica en cada expresión que éste ocurre* (op. cit., 41-2). Los modismos no invalidan el principio de la composicionalidad, ya que es concedido por F&P que los lenguajes naturales no son completamente composicionales y la composicionalidad se mantiene argumentativamente debido a que es imprescindible para explicar cómo un lenguaje finitamente representable puede contener infinitamente varias expresiones no sinónimas (op. cit., 42-3).

¿Cómo se llega de la composicionalidad del lenguaje a la composicionalidad del pensamiento? De un modo parecido a cómo se llega a decir lo mismo de la sistematicidad, es decir, las representaciones mentales deben tener una estructura interna, en el mismo grado que son expresadas lingüísticamente, luego hay composicionalidad en las representaciones mentales, cuyas partes constituyentes tendrían un valor semántico independiente del contexto. Se explicaría así que los pensamientos estén sistemáticamente relacionados (op. cit., 44).

Para F&P es obviamente cierto que la composicionalidad no es una característica de las representaciones conexionistas, simplemente por el hecho de que están comprometidos con representaciones mentales que no tienen estructura combinatoria y lo cual no los hace una buena alternativa al modelamiento clásico. Aun así, existen conexionistas, por ejemplo Smolensky, quienes afirmarían que los hechos de la composicionalidad serían falsos, porque el contenido es dependiente del contexto, y que lo único que existiría sería un “parecido de familia” entre los diferentes significados que tendría una misma representación en los diferentes contextos que ocurre (op. cit., 45). Esta noción suave o débil de composicionalidad es criticada con el siguiente ejemplo (op. cit., 46):

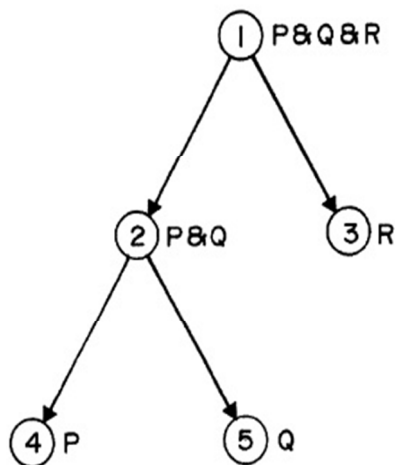
- (i.) Los caracoles son más lentos que los conejos
- (ii.) Los conejos son más lentos que los Ferraris
- (iii.) Los caracoles son más lentos que los Ferraris²⁸.

²⁸ He cambiado “turtles” por “los caracoles”, porque la traducción al español trae consigo la diferencia de géneros de sustantivos lo cual haría perder un poco la claridad del argumento. Por lo demás, hasta aquí he hecho una gran cantidad de acomodados desde el original que no han sido apuntados, pero que el lector podrá notar si se remite a los originales a los que he tratado de exponer lo más fielmente en su contenido teórico exacto.

La crítica consistiría en que habría una relación entre composicionalidad e inferencia que no sería explicada con una composicionalidad basada en un “parecido de familia”, ya que, como muestra el ejemplo, la validez desde (i.) y (ii.) a (iii.) depende de que el significado de la expresión “son más lentos que” sea justamente la misma en las tres proposiciones apuntadas.

-- Tercera característica de la cognición, *Sistematicidad de la inferencia*.

La idea esbozada con el ejemplo al finalizar la exposición anterior, a saber, que *los organismos exhibirán las mismas capacidades cognitivas con respecto a lógicamente similares inferencias*, es tan natural que parece inevitable. Esto no excluye la posibilidad de que existan modelos cognitivos en los cuales



inferencias similares lógicamente sean llevadas a cabo por mecanismo bastante diferentes, pero nótese que los modelos cognitivos propuestos por el conexionismo pueden llevar a una vida mental que sea capaz de llevar a cabo una de las inferencias, pero no la otra, por ejemplo que – siguiendo la **FIGURA 4** a la izquierda – pueda inferir de P&Q&R que P&Q, también que R, pero que selectivamente sea incapaz de inferir Q&R por el sencillo hecho de que las conexiones causales entre los nodos no hayan sido entrenadas antes, a pesar de que ya fue capaz de inferir P&Q y R. Esta clase de vidas mentales de capacidades cognitivas “puntuadas” o “con agujeros” simplemente son imposibles desde una óptica clásica, donde la sintaxis lógica impide tal opción (op. cit., 47-48).

En resumen; dada una arquitectura conexionista, no hay nada que prevenga que una vida mental sea arbitrariamente no sistemática, y partiendo del hecho de que las capacidades cognitivas vienen en cúmulos estructuralmente relacionados, entonces la sistematicidad impregna a la cognición en su conjunto. Toda la evidencia sugiere que las mentes puntuadas no existen y este argumento que sirvió contra el conexionismo de Hebb, Osgood y Hull debiese ser suficiente contra los conexionistas de hoy²⁹. Finalmente, F&P apuntan que nada detiene a un conexionista de estipular – como proposición aislada dentro de su teoría de lo mental – que todas las redes PDP implementadas neurológicamente son, de hecho, sistemáticas, pero tal cosa perdería el objetivo central de estipular la sistematicidad, que es explicar la naturaleza de los mecanismos que llevan a la estipulación de la sistematicidad (op. cit., 48-50).

²⁹ Es preciso advertir que, en contra de la versión simplificada de F&P sobre el conexionismo, no existe un enlace unívoco desde la serie de anécdotas que dan forma a la prehistoria de las ANN y el conexionismo, el cual desde la perspectiva histórica es más un “híbrido” que una continuación clara del asociacionismo de principio del siglo XX. Una defensa del punto “asociacionismo ≠ conexionismo” no será desarrollado más allá de esta de esta nota, ya que considero que anteriormente – y contemporáneamente al grupo PDP – se hizo suficientemente buena defensa de tal punto en Bechtel (1985). La dificultad de trazar una historia unívoca del conexionismo se subentenderá mediante el esquema en (2.5.i) que posibilita visualizar la gran distancia desde la neurona M&P al conexionismo. Esta gran distancia es más evidente si se consideran psicólogos como Clark Hull, figura justo anterior a M&P.

- *Cuarta Parte*

Existe una serie de argumentos que han sido esgrimidos por el conexionismo para subrayar las *limitaciones de los computadores convencionales como modelos de la mente* y que en conjunto constituyen el atractivo de la alternativa conexionista. A continuación F&P los enlistan (op. cit., 50-4):

- Demasiada rapidez del procesamiento respecto a la velocidad neural (restricción de los “cien pasos”).
- Dificultad de la arquitectura clásica de alcanzar una gran capacidad de reconocimiento de patrones y de recuperar información basada en el contenido. El ser humano hace uso de una cantidad sorprendente de información casi sin esfuerzo, lo que se demuestra en su capacidad de reconocimiento de una cantidad de patrones que va, fácilmente, de las decenas a centenas de miles.
- Los modelos convencionales están comprometidos con diferentes etiologías para el comportamiento; uno “basado en reglas” versus uno irregular o “excepcional”.
- Falta de progreso en el tratamiento con procesos que son no verbales o intuitivos.
- Fuerte sensibilidad de las arquitecturas convencionales al daño y el ruido.
- El almacenamiento es pasivo, y dependiente de un ciclo de “recuperación y ejecución” poco realista al ser contrastado con la memoria humana.
- Los sistemas basados en reglas representan la cognición de modo “todo o nada”, lo cual choca con habilidades cognitivas continuas; un caso sería la variación continua en el grado de aplicabilidad de diferentes principios, por ejemplo cuando tales principios o procedimientos están en competencia entre sí en un problema perceptivo. Otro caso sería el no determinismo de la conducta humana.
- Los modelos convencionales son dictados por las características técnicas actuales de los computadores y toman poco en cuenta los hechos de la neurociencia.

F&P consideran que estas razones usuales para preferir a la arquitectura conexionista por sobre la clásica son inválidas por uno u otro de los siguientes dos defectos (op. cit., 54):

Primero, las objeciones dependen de propiedades no intrínsecas a las arquitecturas clásicas, ya que es perfectamente posible la existencia de modelos clásicos que no exhiban los rasgos objetados (por ejemplo, que las reglas necesariamente deban ser explícitas o que las operaciones sean “todo o nada”).

Segundo, las objeciones son verdaderas en cuanto son implementadas en los actuales computadores, pero no necesariamente serán verdaderas cuando sea la arquitectura clásica implementada diferentemente. Estas objeciones son más bien dirigidas al nivel de la implementación que al cognitivo (esto se considera verdadero, por ejemplo, con los argumentos de velocidad o resistencia al daño)

Estos defectos son complementados con comentarios más o menos detallados por F&P:

- Computación paralela y velocidad (op. cit., 54-6): para despejar las dudas que se trata del nivel de implementación se recurre a señalar que, de hecho, el nivel físico de un computador tiene un funcionamiento paralelo diferenciándose así del procesamiento serial del algoritmo, y que esto puede diferenciarse claramente así como la velocidad de ambos procesos. Es más, específicamente el paralelismo no es un argumento en contra de las arquitecturas clásicas, ya que tales arquitecturas

*no excluyen la posibilidad de ejecución paralela de múltiples procesos simbólicos y que tales procesos interactúen unos con otros*³⁰.

- Resistencia al ruido y el daño físico (y el argumento de la representación distribuida) (op. cit., 56-7): Aunque si bien es cierto que la localización del contenido parece una causa clara de falta de resistencia al daño, se debe aclarar que la distribución de la representación en el substrato neuronal es compatible con la visión clásica, ya que lo único que se requeriría sería registros de memoria que distribuyan sus contenidos en el espacio físico. La arquitectura convencional requiere expresiones simbólicas distintas para representar estados de cosas diferentes y, dado que tales expresiones consisten en una estructura concatenada de partes, la relación de adyacencia debe ser instrumentada por *alguna* relación física cuando la arquitectura es implementada, sin embargo, que las relaciones a ser físicamente implementadas sean *funcionalmente adyacentes*, no significa que las implementaciones físicas de los símbolos adyacentes sean *espacialmente contiguas*. Mientras los conocimientos de cómo los procesos cognitivos se mapean en el tejido del cerebro sean bajos, entonces la importancia de la distribución de las representaciones para cs. cognitivas es discutible.
- Restricciones “suaves”, magnitudes continuas, mecanismos estocásticos y símbolos activos (op. cit., 57-9): la noción de que las restricciones “suaves” que varían de manera continua sean incompatibles con los sistemas simbólicos clásicos basados en reglas, es otro malentendido surgido de la falla de no mantener los niveles psicológicos y de implementación separados. Es posible tener un sistema clásico con reglas cuya decisión de disparar una regla resida en una arquitectura que dé magnitudes continuas variables. De hecho, tal cosa ha sido llevada a cabo con la aplicación de mecanismos bayesianos en sistemas expertos. Por otro lado, una regla no lleva necesariamente a un comportamiento “todo o nada”, si se diseñase una generación de conducta construida mediante interacciones o flujos paralelos de instrucciones. Con respecto a la degradación grácil no existe duda para F&P que las arquitecturas clásicas de entonces no están preparadas para enfrentar tal reto, aunque a futuro puedan estarlo. Finalmente, sobre la “pasividad” de los símbolos clásicos que están “a la espera” de ser procesados y no varían en el tiempo hasta su procesamiento, *es una visión ingenua suponer que un modelo clásico se va a comportar exactamente como un ordenador*, de lo que se sigue que los símbolos van cambiando, pero que tal variación puede ser achacada al nivel de instrumentación, o sea, de la “máquina protoplasmática” y no necesariamente al nivel psicológico.
- Explicitud de las reglas (op. cit., 59-61): Es simplemente falso que sea necesaria la explicitación de todos los programas en un modelo clásico. Esta confusión permea toda la literatura conexionista que no ha tomado en cuenta la existencia de una discusión dentro del campo clásico de décadas sobre la

³⁰ Esta clase de defensa ya es hecha por Marr contra quienes critican el carácter serial de su análisis, al decir que “anything programmed in parallel can be rewritten serially” (Marr, 2010 [1982], 27). Más allá de señalar este eco entre autores es importante notar que existen arquitecturas simbólicas, por ejemplo la arquitectura SOAR de Rosebloom et al., que siguen el principio apuntado por F&P, al constar en su diseño de grados de paralelismo (Rosebloom et al., 1991, 319). La arquitectura SOAR – al menos la versión descrita en el artículo citado – es fácilmente adherible al conjunto de modelos simbólicos en cuanto usa producciones y ocupa la noción de “espacio de problemas” (op. cit. 317-8), el cual deriva de la noción de “espacio de soluciones”, concepto utilizado en el desarrollo desde el Logic Theory Machine (Newell & Simon, 1956) al General Problem Solver (GPS) de Newell, Simon y Shaw (1958; 1959; sin embargo, a estos dos textos debe sumarse un tercero que permita ver la imagen completa de la evolución desde una noción a la otra, a saber, el artículo Simon & Newell (1971)).

necesidad de reglas explícitas para explicar algunos fenómenos. De hecho, no todas las funciones en un computador clásico pueden ser codificadas con la forma de un programa explícito, algunas de ellas deben estar cableadas en él, por ejemplo, a nivel de circuitos. En efecto, todos los programas que almacene pueden estar cableados en caso de que no se requiera modificarlos o examinarse a sí mismo. En estos casos tales máquinas clásicas serían de reglas implícitas respecto a sus programas y sus mecanismos de transición de estados serían totalmente subsimbólicos. *Lo único que necesita ser explícito en una máquina clásica es lo que se almacena o sus estructuras de datos, pero no las reglas de transformación.*

- Sobre el modelamiento “estilo cerebro” (op. cit., 62-4): *La relación entre conexionismo y neurociencias está abierto a las interpretaciones, variedad que se encuentra dentro del propio grupo PDP. Sin embargo, más allá de que existen muchos hechos biológicos que han tratado de ser usados para guiar los modelos conexionistas, esto no significa que tales hechos sean relevantes para inferir aspectos a nivel de la arquitectura cognitiva. En palabras de F&P, “The point is that the structure of "higher levels" of a system are rarely isomorphic, or even similar, to the structure of "lower levels" of a system. No one expects the theory of protons to look very much like the theory of rocks and rivers , even though , to be sure, it is protons and the like that rocks and rivers are 'implemented in '.”*

Las últimas páginas de la cuarta parte del artículo se abocan a la cuestión de si el conexionismo puede tratarse como teoría de la implementación. F&P consideran que, por ejemplo, un teórico que esté impresionado por las virtudes del conexionismo, le estaría permitido sin problemas considerarlos como parte de una teoría de la implementación, sin que por ello deba salir del enfoque cognitivo de procesamiento de información clásico, dado que existe un diferencia en principio entre un modelo – en este caso los modelos clásicos – y su implementación (op. cit., 65). *Esta distinción en principio se sostendría prácticamente sola una vez que se admite que los principios del nivel cognitivo son diferentes de los principios arquitecturales conexionistas; es un sinsentido preguntar si se debe hacer o no hacer ciencias cognitivas a partir de la “interacción de los niveles inferiores” como opuestos a los procesos de nivel cognitivo, cuando ciertamente se deben hacer ambos. Algunos científicos estudian principios geológicos, otros estudian el nivel de moléculas, pero dado que, de facto, ambos niveles son fenómenos genuinos, los principios autónomamente estables de la geología no están nunca en disputa como F&P aseguran, “people who build molecular level models do not claim to have invented a "new theory of geology" that will dispense with all that old fashioned " folk geological " talk about rocks , rivers and mountains!”* (op. cit., p. 66).

- *Quinta parte*

La última parte del artículo son algunas *posibles rutas de desarrollo teórico del conexionismo*, las cuales apenas toman un par de páginas de exposición (op. cit., 67-9):

1— Resistir con la existencia de representaciones mentales no estructuradas contra la visión clásica basada en representaciones estructuradas con semántica y sintaxis combinatoria. Empero, según F&P, tal opción no sería atractiva, dados los argumentos de productividad y sistematicidad aquí presentados.

2— Abandonar la arquitectura de red para extenderse en representaciones mentales estructurales, aunque basándose en una explicación asociacionista del procesamiento mental. El problema con esta

salida es que la asociación no es una relación sensible a la estructura y el problema es que con ella no se podría reconstruir la coherencia semántica del pensamiento.

3— Considerar al conexionismo como una teoría implementacional. No existen objeciones en principio de esta visión – aunque existen objeciones técnicas desde los propios conexionistas que han empezado a intentar implementar máquinas clásicas. Esta salida, obviamente, restaría seguidores al conexionismo, ya que muchos de ellos están atraídos por esta nueva metateoría porque, primero, promete quitar el nivel simbólico de análisis y, segundo, eleva a la neurociencia a la posición de cargar con tópicos cognitivos.

4— Rendirse con la idea que las redes PDP ofrecen una base razonable para modelar el procesamiento cognitivo en general. Podría, aun así, mantenerse que tales redes sostienen algún procesamiento cognitivo, sin embargo, desde la mirada racionalista de F&P hay fuertes dudas de que mucho del procesamiento cognitivo consista en análisis de relaciones estadísticas. Luego, lo que quedaría a tales redes sería modelar las inferencias exclusivamente estadísticas, lo cual es bastante modesto si se compara a las expectativas conexionistas.

1.1.3. Tercer momento – El (innecesariamente) prolongado debate de Paul Smolensky y Jerry Fodor.

Dejando de lado tanto la crítica de Pinker y Prince como la de Bever y Lachter, cuyo foco está constreñido (no exclusiva, aunque sí preferentemente) al modelo propuesto por Rumelhart y McClelland sobre la inflexión verbal del pretérito en inglés (McClelland & Rumelhart, 1986, cap. 18), Paul Smolensky, tomará el guante dejado por F&P y defenderá al conexionismo. Ergo, esta defensa debe ser considerada como una posición que no puede ser adjudicada al conjunto del grupo PDP, como sí se puede afirmar de la postura conexionista en el primer momento de la polémica simbolismo-conexionismo.

Esta primera respuesta a la crítica de F&P (1988) será compleja y requerirá señalar desde el comienzo que en ella confluyen al menos tres artículos para conformar sus distintas facetas³¹, las cuales ciertamente se complementan entre sí y que tomadas en conjunto pueden verse como redundantes en algunos de sus señalamientos constituyentes. Los tres artículos que usaré para reconstruir tal primera respuesta de Smolensky a F&P serán: *The Constituent Structure of Connectionist Mental States: A reply to Fodor and Pylyshyn* (1987a), *The Proper Treatment of Connectionism* (1987b) y *Connectionism, constituency and the language of thought* (1988a)³².

³¹ Digo “al menos”, ya que para esta primera parte del debate Fodor-Smolensky no considero artículos tales como *Putting together connectionism-again* (1988b) o *On variable binding and the representation of symbolic structures in connectionist systems* (1987c), los cuales son del mismo periodo que estoy considerando y que pueden verse, sin lugar a dudas, como agregando matices que permitirían una comprensión aún más completa y detallada de la primera respuesta de Smolensky a F&P y de la réplica de Fodor y McLaughlin. Por otra parte la postura de F&P puede ser expandida en sus detalles teniendo en cuenta los propios trabajos de McLaughlin (1993a; 1993b) paralelos al debate Fodor-Smolensky e independientemente redactados a Fodor, quien también escribirá en solitario otros textos relacionados (Fodor, 1989). Dada que mi idea no es ir al detalle de la evolución del pensamiento de Smolensky, ni a las sutiles distinciones que pueden construirse entre el pensamiento de Fodor y Pylyshyn versus Fodor y McLaughlin, entonces no consideraré tal bibliografía indicada, aunque queda al lector más entusiasta la oportunidad de ahondar en los textos señalados.

³² El lector más sagaz podrá notar que *The Constituent Structure of Connectionist Mental States: A reply to Fodor and Pylyshyn* tiene como fecha de publicación 1987, y su bibliografía cita explícitamente el artículo de F&P que

Posteriormente a tal respuesta, Jerry Fodor volverá nuevamente a la carga, pero esta vez no junto a Zenon Pylyshyn, sino que junto a Brian McLaughlin, en el artículo “contraataque” *Connectionism and the problema of systematicity: Why Smolensky’s solution doesn’t work*, que fue desarrollado y entregado en 1989, pero cuya fecha de publicación es de 1990. Bastante más sintético que el primero con F&P, la dupla F&M se centra en algunos detalles conceptuales que he usado para guiar y organizar la exposición de la producción de Smolensky en la primera parte del debate Fodor-Smolensky.

Finalmente, es preciso apuntar que existen una tercera y cuarta parte del debate que se remite, respectivamente, a los artículos de Smolensky (1995) – *Reply: Constituent structure and explanation in an integrated Connectionist/Symbolic Cognitive Architecture* – y Fodor (1997) – *Connectionism and the problem of systematicity (continued)- why Smolensky's solution still doesn't work* –, cada uno de los cuales pueden ser analizados dentro de un contexto aún más rico teóricamente³³ que los artículos que he seleccionado para constituir la primera y segunda parte, pero que en honor a la no dispersión y, además, a no ser relevantes para la segunda parte de la tesis, dejaré de lado. En síntesis, el debate entre Paul Smolensky y Jerry Fodor (y asociados – Pylyshyn y McLaughlin) puede ser dividido en cuatro partes bien definidas: una primera, donde Smolensky saca la voz por el grupo PDP y responde a F&P, una segunda donde Fodor junto a Brian McLaughlin vuelven a responder a la propuesta de Smolensky, una tercera como segunda respuesta de Smolensky y una cuarta, tardía en su publicación, de respuesta de Fodor. A continuación expondré las dos primeras partes que son las más relevantes para la presente exposición de la polémica entre conexionistas y simbolistas cognitivistas.

- *Primera parte del debate Fodor-Smolensky*

En el artículo *The Constituent Structure of Connectionist Mental States: A reply to Fodor and Pylyshyn* (1987), la respuesta crítica de Smolensky a F&P presupone un cierto contexto meta-teórico, el cual se reduce en una *paradoja cognitiva fundamental*. La importancia de esta paradoja radica en que es marco de referencia para la modelización cognitiva posterior, en tanto que la respuesta que se adopte sobre la paradoja implica una diferenciación meta-teórica evidente entre simbolismo y conexionismo, además de las variaciones teóricas intermedias entre tales extremos de modelización. La Paradoja es simple de decir: por un lado la cognición es *rígida*³⁴ caracterizada por las reglas de la lógica y el lenguaje, mientras que por el otro lado la cognición es *flexible*, puesto que, si uno escribe programas basados estrictamente en reglas lógicas, su comportamiento será falto de fluidez, sin generar un comportamiento que llamaríamos “inteligente”. Una manera alternativa de llamar a esta paradoja es la de “la *estructura versus la estadística*” (Smolensky, 1987a, 137-8).

Las diferentes respuestas a esta paradoja son las que darán lugar a los diferentes lugares teóricos relativos que puede ocupar el conexionismo respecto al simbolismo (Smolensky, 1987a, 138-143):

tiene como fecha oficial 1988. Esta anomalía temporal en las publicaciones se explica por el intercambio intelectual entre Smolensky y F&P en el marco de un *workshop* en torno al tema de los fundamentos de la inteligencia artificial en febrero de 1986, lo cual permitió a los autores conocer las posturas de cada lado (Smolensky, 1988a).

³³ Por ejemplo, un *paper* que sirve como profundización del lado conexionista podría ser Dolan & Smolensky, 1989.

³⁴ La traducción que aparece en Rabbossi (1997) traslada los términos “hard” y “soft” a “duro” y “blando”. Dado que “rígido” y “flexible” encaja mucho mejor con la crítica del grupo PDP a las estructuras de datos tradicionales (por ejemplo, véase la crítica a los esquemas en el capítulo 14 de PDP 1, 1986), entonces usaré tales términos.

- [I] “Negación de lo flexible de la cognición” o racionalismo³⁵ (o *cognitivismo*³⁶)

Considerada por Smolensky la elección más popular entre los modeladores cognitivos, esta postura tiene como directriz definitoria considerar la cognición como guiada por la lógica y el uso de reglas explícitas, de lo que se seguiría su negación de los aspectos flexibles de la cognición, los cuales serían considerados meramente “ruido”. Presumiblemente, su inspiración deviene del hecho que las reglas formuladas en el lenguaje natural proveen una formalización útil del conocimiento cultural, de lo que se seguiría que el uso consciente de reglas para llegar a conclusiones en las personas pueda ser mejor capturada mediante la interpretación secuencial de tales reglas mediante cierta máquina virtual. Estas reglas se formularían en términos de operaciones sobre conceptos/símbolos en el *nivel conceptual de análisis* (Smolensky, 1987b, 2-3).

- [II] “Cerebro dividido” o cohabitación

Esta opción rechaza la idea de eliminar los aspectos flexibles de la cognición mediante su reducción a “mero ruido”, por lo cual los aspectos flexibles de la cognición tendrían lugar en “módulos conexionistas”, los cuales estarían separados de los “módulos basados-en-reglas”, existiendo seguramente algún modo de comunicación entre ellos. No obstante, Smolensky apunta que si sigue la Paradoja hasta sus últimas consecuencias, entonces tal cohabitación entre módulos se diluiría, ya que la comunicación al entrelazarlos disolvería sus límites.

- [III] “Enfoque difuso” o emergentismo desde lo rígido

Más sofisticado que el anterior, se sintetiza como aquella postura donde el comportamiento flexible del sistema cognitivo surgiría de la confluencia de una cantidad tan alta de pequeñísimas reglas, que, cuando se observa desde un nivel superior la cognición, pareciese que tuviese un aspecto de flexibilidad, aunque sea totalmente rígida a nivel ínfimo.

- [IV] “Enfoque subsimbólico”, subsimbolismo o emergentismo desde lo flexible

El enfoque subsimbólico es propuesto y defendido en detalle en Smolensky (1987b). Este enfoque supone la existencia de un nivel de análisis inferior al simbólico o conceptual, un nivel subconceptual, el cual *no es equivalente al nivel neuronal*. Esto se justifica en nuestra ignorancia sobre qué

³⁵ Obviamente no se está pensando en el racionalismo que comprende la filosofía moderna, sino que este “racionalismo” es aquel que inspira la noción idealizada de competencia en lingüística, la cual para Chomsky (1967) presenta tres aspectos innatos cruciales (i) un aspecto creativo en el uso del lenguaje; (ii) estructuras profundas de naturaleza abstracta; (iii) universalidad de un sistema especial de mecanismos, que sería la gramática transformacional. En la misma fuente, Chomsky considera que tal racionalismo que permea su noción de competencia contrasta notoriamente con la noción empirista de aquello que puede ser considerado innato, a saber, (i) ciertos mecanismo elementales del procesamiento periférico (un sistema receptor) y (ii) ciertos mecanismos analíticos generales – que serían usados por toda función cognitiva y que no sería paradigmática de, por ejemplo, el lenguaje – o principios inductivos o alguna forma de mecanismos asociativos. Luego, en el empirismo habrían sólo dos clases de mecanismos diferenciables: el primario derivado del procesamiento fisiológico más básico y uno secundario dependiente de principios inductivos de alcance general.

³⁶ Smolensky no utiliza el término “cognitivismo”, sino que “simbolismo”. Empero, lo he agregado entre paréntesis, ya que considero el primero más acertado que el segundo por lo que señalaré en (2.5.ii)

arquitectura usa el cerebro para resolver la mayoría de las tareas cognitivas (pueden existir excepciones, como la visión o las tareas espaciales), especialmente para las funciones como la resolución de problemas, el lenguaje y muchos otras donde no necesariamente el modelamiento neuronal estricto funcionaría (Smolensky, 1987b, 5).

En términos formales, el subsimbolismo se basa en la hipótesis de que las redes PDP serían casos especiales de sistemas dinámicos, los cuales quedan sin apuntar por Smolensky. Empero, sí asevera que las redes PDP serían diferentes a la “hipótesis sintáctica simbólica” y que la pregunta de fondo es acerca de la *semántica* de tales sistemas dinámicos conexionistas. La respuesta a la pregunta “¿qué significa el valor de una unidad?” es que ellas carecen de semántica conceptual, ya que son *subconceptuales* y que su interacción a nivel patrones de activación no es directamente descrita por una definición formal dentro de un modelo subsimbólico; ellas deben ser computados por el analista (op. cit., 5-6). Ergo, el procesador intuitivo – definido como “la máquina virtual que corre aquellos programas que no son aplicaciones de reglas conscientes” (op. cit., 3) –, que en el caso simbolista es equivalente al “intérprete de reglas consciente”, en el caso conexionista es un sistema dinámico subconceptual que no admite una descripción precisa formal a nivel conceptual (op. cit., 6).

Esto último parece no responder a la pregunta, pero para Smolensky todo esto preparará el terreno para la respuesta – la cual tiene en cuenta que la incompatibilidad entre simbolismo y subsimbolismo radica en que los principios de la cognición para un modelo subsimbólico no descansan en el nivel conceptual (op. cit. 7) – que será que *la totalidad de las disposiciones de los sistemas cognitivos se explican en el paradigma subsimbólico como regularidades aproximadas de nivel superior que emergen de las leyes cuantitativas que operan a un nivel más fundamental , con una semántica diferente* (op. cit., p. 10-1).

Luego, la arquitectura cognitiva conexionista es de “niveles-divididos”, es decir, con un nivel – el de la activación de las unidades – donde sólo existe sintaxis y uno superior, el del análisis de los patrones emergentes en el tiempo, donde está la semántica. *La diferencia radical es que no existe como en los modelos clásicos un nivel donde se superpone la semántica y la sintaxis* (Smolensky, 1988a, 4-5).

- [V] “Negación de lo rígido de la cognición” (o *eliminacionismo*³⁷)

Esta visión se define por afirmar que la esencia de la inteligencia está en la evasión del seguimiento de reglas. El conexionismo se puede prestar a esta salida de la Paradoja, dado que como formalismo puede verse como compatible o no a las reglas, así como también a la psicología popular³⁸. En términos de Smolensky, “the strong eliminativist position states that even viewing the nonformal principles [tales como memoria, estructuras constitutivas y atención] as a starting point for thinking

³⁷ Paradigmáticamente esta postura es defendida por Paul Churchland (1984) y Patricia Churchland (1989).

³⁸ La relación entre conexionismo y la psicología *folk* es bastante discutida. Por un lado está la postura que muestra que existen modelos conexionistas incompatibles, en principio, con la modularidad proposicional de la psicología de sentido común (Ramsey, Stich & Garon, 1991). Por el otro lado, se les ve como naturalmente complementarias (Forster & Saidel, 1994). Más allá de tomar partido aseveraré, sin entrar en justificaciones por ahora, que la psicología popular puede verse tanto congruente como incongruente con el conexionismo (véase sección (2.5.iv))Paradigmáticamente esta postura es defendida por Paul Churchland (1984) y Patricia Churchland (1989).

about cognition is a mistake; e.g., that it is better to pursue a blind bottom-up strategy in which low-level connectionist principles are taken from neuroscience and we see where they lead us without being prejudiced by archaic prescientific notions” (Smolensky, 1987b, 5).

Pues bien, entre todas las posturas, habrían mejores razones para preferir la salida subsimbólica (Smolensky, 1987a, 140) a las otras: Primero, porque considera como soporte de la inteligencia a la percepción y no a la inferencia lógica. Segundo, porque el comportamiento rígido guiado por reglas parece más la excepción que la norma. Tercero, desde la perspectiva evolutiva, la percepción aparece más tempranamente que los procesos rígidos, y pareciera razonable asumir que la ontogenia cognitiva recapitule la filogenia. Cuarto, es más natural la relación sistema nervioso y subsimbolismo. Quinto, el análisis de un sistema conexionista no resulta extraño a la clase de análisis que se hace en los sistemas dinámicos físicos, lo cual implica facilidad para entender la emergencia cognitiva. Y, sexto, la posición que niega lo flexible y se afirma sólo en lo rígido de la cognición lleva tiempo siendo decepcionante.

Smolensky ve esperanzadoramente el futuro de la teoría conexionista, dado que si el enfoque subsimbólico es exitoso daría una solución unificada a la paradoja, aun cuando pareciera ser que el conexionismo tiene que pagar altos precios explicativos por no adherir al formalismo simbólico, como por ejemplo, apunta la crítica de F&P. Antes de entrar propiamente en la respuesta de Smolensky a F&P, vale la pena mencionar que la interpretación del conexionismo que defiende Smolensky no es una que sea adherida por el grupo heterogéneo que adscribe al formalismo del conexionismo. *El subsimbolismo debe ser distinguido así de los modelos de redes PDP* (op. cit., 141-2). A continuación, la reconstrucción que se hace de los puntos centrales de la crítica de F & P es la siguiente (op.cit. , 143):

1. Los pensamientos tienen una estructura compuesta – recordemos el rasgo arquitectural [1] de F&P.
2. Los procesos mentales son sensibles a tal estructura compuesta – rasgo arquitectural [2] de F&P.

Estos dos puntos, desde la perspectiva de Smolensky, son elevados a la categoría de ser definitorios de los modelos clásicos, siendo cuestionados por los conexionistas tales rasgos arquitecturales, por ende, primeramente, la respuesta a F&P considera cuál es la postura del subsimbolismo a tales rasgos arquitecturales. Smolensky no niega la necesidad de explicar tales rasgos arquitecturales, los cuales conceptualmente caracterizan el lado rígido de la cognición, sin embargo tampoco desea abandonar el lado flexible de la cognición. Empero, para F&P, el conexionismo fracasa en dar cuenta de tales rasgos arquitecturales, dado *que representa los pensamientos atómicamente* (estados mentales sin estructura), lo que viola [1], además de que *el procesamiento conexionista es asociacionista*, sensible solamente a la *estadística*, de modo que así también se viola el rasgo arquitectural [2].

Contra la suposición de F&P del atomismo conexionista, Smolensky responde que la versión presentada del conexionismo por F&P no se aplica a los modelos PDP que él está defendiendo. La red de F&P³⁹ es un caso de representación localista, y las redes PDP son de representación distribuida. Si se siguen las

³⁹ Véase las figuras 3 y 4 antes expuestas.

sugerencias del propio Pylyshyn, el análisis de una red conexionista distribuida implica que los estados mentales poseerían una estructura constitutiva de nivel “dentro de” y no “entre”⁴⁰ (op.cit. , 144).

El argumento central de Smolensky se basa en un caso paradigmático para mostrar la relación entre los microrrasgos, o microcaracterísticas, con las representaciones mentales complejas. Considérese la siguiente tabla:

	Taza con café	Taza sin café	Café
Recipiente vertical	●	●	○
Líquido caliente	●	○	●
Vidrio contactando madera	○	○	○
Superficie curva de loza	●	●	○
Olor a quemado	●	○	●
Líquido marrón contactando loza	●	○	●
Superficie curva de loza	●	●	○
Objeto de plata oblongo	○	○	○
Asa de tamaño del dedo	●	●	○
Líquido marrón con lados curvos y fondo	●	○	●

Si tomamos las listas de microcaracterísticas que corresponden a “taza sin café” y las de la lista de “café” observaremos una combinación muy extraña, dado que “taza con café” se ha obtenido de una representación de “café” que está extraída de la propia representación de taza con café. A su vez, que “taza sin café”, también está extraída de la misma situación de la que se extrajo “café”. Luego, aquí vale la pena citar extensamente como lo dice Smolensky:

“So the compositional structure is there, but it is there in an approximate sense. It is not equivalent to taking a context-independent representation of coffee and a context-independent representation of cup – and certainly not equivalent to taking a context-independent representation of the relationship in or with – and sticking them all together in a symbolic structure, concatenating them together to form the kinds of syntactic compositional structures that Fodor and Pylyshyn think connectionist nets should implement” (Smolensky, 1987a, 147).

Lo anterior significa que la representación de café que se extrae de una construcción tal como “taza con café” conduce a una representación totalmente distinta de construcciones que intuitivamente uno consideraría equivalentes – por ejemplo “café con leche”. Si se quisiese hablar de representación conexionista distribuida de “café”, lo máximo que obtendríamos es un parecido de familia entre patrones de activación distintos (op.cit., 148).

⁴⁰ Recuérdese que para F&P es relevante que los estados constitutivos de las representaciones mentales complejas estén en el mismo nivel conceptual o cognitivo. Esto se sigue de la propia definición cognitivista de lo que es una arquitectura cognitiva. Con esto Smolensky ya está preparando el terreno para mostrar que una representación distribuida en tanto se puede considerar como una representación mental, puede ser descompuesta en elementos que pueden considerarse dentro del mismo nivel cognitivo.

Luego, se responde al rasgo arquitectural [1] en tanto que los componentes de la representación compleja de, siguiendo el ejemplo, “taza con café”, se mueven en el mismo nivel de tal representación compleja. La primera moraleja de la historia del café para Smolensky es que no es relevante la crítica de F&P sobre la necesidad de relacionar un vector a sus microrrasgos constituyentes, *dado que la constitutividad es una relación, en el caso del conexionismo, que se puede encontrar entre los vectores y no entre un vector y sus elementos numéricos constituyentes.*

Una segunda enseñanza que se concluye es que las relaciones de constitutividad entre diferentes representaciones distribuidas aparecen *a posteriori* del análisis de los modelos PDP, pero no conforman un mecanismo causal dentro del modelo. De esto se sigue que uno puede partir por cualquier definición particular de constitutividad para analizar el modelo y tratar de responder sobre su comportamiento; las ambigüedades que puedan existir responden a nosotros al analizar el sistema conexionista, pero estas ambigüedades no existen dentro del procesamiento de la red. Luego, como una tercera moraleja que surge acerca del rasgo arquitectural [1], la constitutividad que es la esencia de tal rasgo arquitectural en el caso conexionista solamente podrá ser formalizado aproximadamente. No existe descomposición unívoca de los estados compuestos conexionistas. Finalmente, como cuarta cuestión, Smolensky recalca que la constitutividad aproximada no es precisamente compatible con una implementación directa de un lenguaje sintáctico del pensamiento (op.cit. , 149).

Con respecto al rasgo arquitectural [2], Smolensky considera que existen algunas claves que llevan hacia una respuesta sobre cuáles serían las operaciones sensibles a la estructura de las representaciones distribuidas conexionistas. En Smolensky (1987a) no hay más que una sucinta indicación de que tal individuación de la operación correspondiente existe, y los detalles técnicos de tal operación se encuentra en el artículo *On variable binding and the representation of symbolic structures* (1987c)⁴¹.

Por otra parte, existe una solución paralela a la presentada mediante el “cuento del café” para dar cuenta de la relación de constitutividad que vaya más allá de la basada en el “parecido de familia”. Smolensky se pone en el caso en el cual la codificación del vector de la representación distribuida de “taza con café”, con vectores constituyentes representando “taza” y “café”, sea demasiado débil para servir a todos los usos de la estructura constitutiva, en particular tal codificación demasiado débil para apoyar las inferencias formales, ya que el vector representado “taza” *no puede llenar múltiples roles estructurales*. Luego, un elemento constitutivo debiese poder moverse y llenar cualquier número de roles diferentes en diferentes estructuras, es decir, hacerse con vectores codificando representaciones distribuidas, y de manera que no meramente implemente constitutividad sintáctica simbólica. Estas restricciones darán a lugar a que Smolensky proponga una salida basada en un tipo de representaciones conexionistas que denominará *representaciones mediante producto tensorial* (Smolensky, 1988a, 11).

Una representación particular mediante producto tensorial estará construida en dos pasos (Op.cit., 12):

a) Especificar un proceso de descomposición donde sea que las estructuras discretas son explícitamente divididas como un conjunto de constituyentes, cada una relleno un rol particular en la estructura

⁴¹ El contenido de este reporte técnico es prácticamente en su totalidad el mismo que se presenta en el artículo *Tensor product variable binding and the Representation of symbolic structures in connectionist systems* (1990)

como un todo. Este paso no tiene mucho que ver con el conexionismo *per se*, solamente estará especificando la clase de estructura constituyente que se desea representar.

b) Especificar dos representaciones conexionistas: una para los roles estructurales y otra para sus rellenos – o *fillers* (los constituyentes). Luego, para cada relleno, se asignará un vector en el espacio de estados de alguna red para la representación de rellenos; equivalentemente se asignará a cada rol un vector en el espacio de estados de alguna red para representar roles.

Estos dos pasos indican los “parámetros” en el esquema general de representación mediante producto tensorial que deben ser especificados para individuar una representación particular. La representación del todo es construida desde la representación de sus partes constituyentes a través de la operación de superposición, la cual es simplemente adición de vectores: el vector representando al todo es la suma de los vectores representando las partes. El paso (a) anterior especifica exactamente qué constituyentes están involucrados en este proceso. El paso (b) especifica un conjunto de vectores que representan roles estructurales individuales y otro conjunto de vectores que representa rellenos individuales para tales roles independientes de cualquier rol. La operación de producción aquí es una operación vectorial denominada producto tensorial, la cual toma dos vectores y produce un nuevo vector; si dos vectores consisten en n y m valores de activación, entonces su producto tensorial es un vector de nm valores de activación, cada uno siendo un producto diferente de dos valores de activación, uno a partir de cada uno de los vectores originales (Op.cit.)⁴².

Para Smolensky la representación mediante producto tensorial de la estructura constitutiva fortalece considerablemente la noción de constitutividad que ha sido intuitivamente presentada mediante el cuento del café. En aquella se ve que la relación todo/parte entre “taza con café” y “café” está reflejada en una relación todo/parte entre sus respectivas representaciones: esta última relación no es una relación todo/parte entre estructuras simbólicas moleculares y sus constituyentes atómicos, como se da en el caso de LOT, sino más bien la relación entre un vector suma w y los vectores componentes que se adicionan a éste: $w = c_1 + c_2 + \dots$. Lo mismo es verdad aquí respecto a las representaciones mediante producto tensorial, pero ahora como agregado es posible identificar las representaciones de cada constituyente como una presentación rol-dependiente construida de manera sistemática (a través de la ligadura de variables mediante producto tensorial) a partir de representaciones rol –independientes del relleno y una representación relleno-independiente de su rol (Op.cit., 13).

Retornando a la crítica de F&P, Smolensky muestra cómo el esquema representacional mediante producto tensorial puede lidiar con problemas simples de inferencia aristotélica. Smolensky desciende a tal nivel de detalle para enfrentar los requerimientos formales de F&P para una representación de

⁴² Smolensky en este artículo es poco pedagógico respecto a la representación mediante producto tensorial. Un ejemplo adecuado para guiar la comprensión de su propuesta es el siguiente: Considérese el estado de hechos “Juan ama a María”. Los vectores f_1 , f_2 y f_3 representan respectivamente los *fillers* “Juan”, “ama a”, “María”, mientras que los vectores r_1 , r_2 y r_3 respectivamente representan los roles de sujeto, verbo y objeto. Luego, si el producto tensorial es representado por \otimes y la adición vectorial por $+$, entonces $(f_1 \otimes r_1)$ representará a “Juan” en el rol de sujeto, $(f_2 \otimes r_2)$ representará “ama a” en el rol de verbo y $(f_3 \otimes r_3)$ representará a “María”. Por consiguiente, $[(f_1 \otimes r_1) + (f_2 \otimes r_2) + (f_3 \otimes r_3)]$ representará el hecho “Juan ama a María”. Si se deseara representar lo contrario, es decir “María ama a Juan”, entonces podría usarse el vector $[(f_3 \otimes r_1) + (f_2 \otimes r_2) + (f_1 \otimes r_3)]$ (Guarini, 1998, 24-5).

estructura combinatoria (Fodor & Pylyshyn, 1988, 14). Así pues, según Smolensky, no se ha proveído meramente una implementación del simbólico LOT. Las razones para decir que no se trata de una implementación serían (solamente apuntadas) las siguientes (Smolensky, 1988a, 14):

- i. Inexistencia de unicidad respecto a roles o rellenos
- ii. Profundidad no restringida
- iii. Posibilidad de confusión en la memoria
- iv. Procesamiento no siempre independiente

Las anteriores propiedades potencialmente fallidas muestran que no se está tratando con una mera implementación de LOT, no obstante esta pérdida de algunas características superficialmente deseables de LOT, puede ser vista bajo el prisma de la ganancia de algunas características superficialmente deseables del procesamiento conexionista a cambio (op. cit., 14-5):

- i. Paralelismo masivo
- ii. Memoria dirigida al contenido
- iii. Inferencia estadística
- iv. Aprendizaje estadístico

Dado todo lo anterior, la conclusión es que un modelo conexionista distribuido, o bien mediante la aceptación de la propuesta presentada con la “historia del café”, o bien a través representaciones de producto tensorial, satisface [1] y [2]. Además, si bien el mecanismo de inferencia conexionista es diferente en sus detalles a el simbólico, *para Smolensky es suficiente el hecho de que un mecanismo conexionista es capaz de inferir A a partir de A y B, para mostrar que se satisface una relación de tipo todo/parte*, siendo en este sentido no diferente de un mecanismo de inferencia simbólico que se basa en la estructura sintáctica de A y B para extraer sintácticamente A (Smolensky, 1987a, 149-150).

Los rasgos [1] y [2] son aceptados por el subsimbolismo y el simbolismo, estando la diferencia entre los enfoques en cómo tales principios se instrumentan formalmente. La verdadera disputa está en el nivel de las instrumentaciones formales y no en los principios no formales tales como los rasgos arquitecturales, la relación de constitutividad, la composicionalidad, sistematicidad o la sistematicidad de la inferencia. Sintéticamente dicho por Smolensky:

The overall conclusion, then, is that *the classical and connectionist approaches differ not in whether they accept principles (1) and (2), but in how they formally instantiate them*. To confront the real classical/connectionist dispute, one has to be willing to descend to the level of the particular formal instantiations they give to these non formal principles. To fail to descend to this level of detail is to miss the issue. In the classical approach, principles (1) and (2) are formalized using syntactic structures for thoughts and symbol manipulation for mental processes. In the connectionist view (1) or (2) are formalized using distributed vectorial representations for mental states, and the corresponding notion of compositionality, together with association-based mental processes that derive their structure sensitivity from the structure sensitivity of the vectorial representations engaging in those processes. (op. cit., 151)

○ Segunda parte del debate Fodor-Smolensky

Fodor y McLaughlin (1990 – “F&M”) plantean el dilema al cual se enfrenta el conexionismo del siguiente modo: si el conexionismo no puede dar una explicación para la sistematicidad, entonces falla para proveer una base adecuada para una teoría de la cognición. Al contrario, si se diera el caso de dar cuenta de la sistematicidad mediante procesos mentales que son sensibles a la estructura constituyente de las representaciones mentales, entonces tal teoría cognitiva conexionista pasaría a ser parte de una arquitectura que implementaría un modelo clásico.

Dado que F&M consideran necesario hacer un recuento de cómo el problema de la sistematicidad es solucionado mediante el modelamiento clásico, sucintamente a continuación reconstruiré tal recuento. Se asumen sin argumentar cuatro puntos (Fodor & McLaughlin, 1990, 185):

- i. Que las capacidades cognitivas son generalmente sistemáticas en humanos y en muchos organismos infrahumanos⁴³.
- ii. Que lo anterior es nomológicamente necesario, luego apoyado contrafácticamente.
- iii. Luego, que deben existir alguna clase de mecanismos psicológicos en virtud del funcionamiento de los cuales las capacidades cognitivas sean sistemáticas.
- iv. Y que una adecuada teoría cognitiva debe exhibir tal mecanismo.

En este punto del debate es dado por hecho que Smolensky ha aceptado los puntos i al iv. F&M aclaran que dar por demostrada la sistematicidad deja la pregunta por la composicionalidad abierta; la sistematicidad de la cognición consiste en, por ejemplo, que del hecho que un organismo pueda pensar aRb, también pueda pensar bRa. *La composicionalidad propone una cierta explicación para la sistematicidad: que el contenido de los pensamientos está determinado uniformemente por el contenido de los conceptos contexto-independiente que son sus constituyentes.* La cuestión polémica es que si uno es un conexionista que acepte la sistematicidad, entonces uno debe argumentar, o bien que la sistematicidad puede ser explicada sin composicionalidad o que la arquitectura conexionista propuesta acomoda representaciones composicionales (op.cit.).

Esta aclaración es crucial porque a partir de F&P uno puede suponer que existe algo equivocado en principio desde el punto de vista clásico con el conexionismo, dado que falla en dar una noción descontextualizada de la representación mental. Sin embargo, la aclaración de F&M apunta precisamente a que no es necesario que el conexionista asuma representaciones mentales sensibles al contexto o no, sino más bien ofrezcan alguna explicación de la sistematicidad que no implique composicionalidad y, el fondo de lo que fue articulado por F&P es que los conexionistas no tienen una explicación de la sistematicidad que no implique composicionalidad – recuérdese la noción de composicionalidad basada en el parecido de familia y la representación del producto tensorial.

⁴³ Esta es una pequeña variación con respecto a lo presentado por F&P donde se dirá que es una cuestión empírica saber si ciertos organismos infrahumanos son cognitivamente sistemáticos, no obstante, en tal contexto también se afirmará que, muy presumiblemente, sean sistemáticos, y el que ciertos organismos respondan a aRb, pero no a bRa, serían más bien los casos etológicos donde “la excepción que confirma la regla” (Fodor & Pylyshyn, 1988, 41)

Entrando ya en el recuento, la visión clásica mantiene que la sintaxis de las representaciones mentales es como la sintaxis de las expresiones lingüística en el sentido que ambos incluye símbolos complejos. Luego, se estipula que, para cualquier par de expresiones tipo E1, E2, la primera será constituyente clásica de la segunda si y solo si la primera es implementada siempre y cuando la segunda lo es (por ejemplo “Juan” es constituyente de la expresión – o de una “representación mental compleja”, dado el alto paralelismo entre un lenguaje natural y LOT – “Juan ama a María”). Es, precisamente, gracias a que los constituyentes clásicos tienen esta propiedad, que es que siempre son accesibles a las operaciones que son definidas sobre los símbolos complejos que los contienen. Además, es justamente porque las representaciones mentales complejas clásicas tienen constituyentes, que ellos conforman un dominio para procesos mentales sensibles a la estructura (op.cit., 186-7).

La semántica de las representaciones clásicas es una extensión de lo que sucede con la sintaxis. F&M resumen estos supuestos sobre la sintaxis y la semántica en una condición C (op.cit., 187):

C. Si una proposición P puede ser expresada en un sistema de representaciones mentales M, entonces M contiene alguna representación mental compleja (una “oración mental”) S, tal que S expresa P y los constituyentes clásicos de S expresan (o refieren a) los elementos de P.

La explicación clásica de la sistematicidad asume que C se mantiene mediante necesidad nomológica, es decir, expresa una ley psicológica que subsume a todas las mentes sistemáticas. F&M consideran bastante claro cómo la sistematicidad es fácilmente explicable a partir de los supuestos, primero, que las representaciones mentales satisfacen C, y que, segundo, los procesos mentales tienen acceso a la estructura constitutiva de las representaciones mentales (op.cit., 188).

Luego de este recuento, la síntesis del argumento en contra de Smolensky sería la siguiente: por una parte la arquitectura cognitiva que él mantiene no provee representaciones mentales con constituyentes clásicos. Por otra parte, él no provee sugerencia de cómo los procesos mentales podrían ser sensibles a la estructura, ni menos cómo las representaciones mentales tendrían constituyentes clásicos. Es más, Smolensky no provee sugerencia alguna de cómo las mentes podrían ser sistemáticas si las operaciones mentales no son sensibles a la estructura (op.cit.). Esta síntesis es dada antes por F&M de entrar en detalle en contra de la primera de las propuestas de Smolensky: la “historia del café” que va en línea con lo que F&M denominarán “composicionalidad débil” La segunda propuesta a objetar será la “hipótesis de la representación mediante producto tensorial y una superposición de representaciones”, que será considerada como la propuesta “fuerte” de composicionalidad.

En el cuento del café de Smolensky se introducen cuatro puntos que F&M consideran interesantes señalar (op.cit., 190-1):

- i. “Café” y “taza con café” son vectores de activación que sirven como representaciones de los patrones de actividad de la red conexionista.
- ii. “Taza con café” contiene la representación de “café” como constituyente en el sentido que lo contiene como un vector componente; la relación entre los vectores y sus constituyentes no clásicos es que los primeros son derivables de los segundos por operaciones de análisis vectorial.

iii. Las representaciones, tanto de “café”, como de “taza con café”, se dan sobre unidades que pueden ser interpretadas como microrrasgos o microcaracterísticas.

iv. “Café”, y presumiblemente cualquier otro vector como representación, es contexto-dependiente.

F&M consideran que i y ii introducen las ideas que las representaciones mentales son vectores de activación y que tienen constituyentes no clásicos, no obstante estas ideas son neutrales con respecto a la distinción entre composicionalidad débil y fuerte. El punto iii es, desde la perspectiva de ellos, una cortina de humo, dado que la cuestión de los microrrasgos es ortogonal, tanto a la pregunta por la sistematicidad, como al tema de la composicionalidad. *El punto iv es más interesante, dado que es el que permite distinguir a una representación que tiene una estructura composicional débil, en tanto que, si solo si, contiene constituyentes dependientes del contexto (op.cit., 190).*

Esta composicionalidad débil tiene el problema de que si uno combina, siguiendo el cuento del café, una representación con el contenido de “taza sin café” con la representación con el contenido de “café”, uno obtiene no una representación con el contenido de “taza con café”, sino que una representación con contenido autocontradictorio de “taza sin café con café”. El procedimiento de sustracción del contenido representacional, presentado por Smolensky en su historia del café, tiende a confundir la representación de “taza sin café” con la representación de “taza” sin la representación de “café” (op.cit., 191).

Smolensky cuando habla de las conclusiones que se extraen a partir de su ejemplo sobre la taza de café, dice que la dependencia del contexto de las representaciones es tal que un vector que representa “café” en dos o más apariciones compuestas no tienen más en común que cierto “parecido de familia”. Luego, no hay un vector único que cuente como la representación de “café”, es decir, no hay un único vector que sea componente de todas las representaciones en las cuales, en un sistema clásico, estaría “café” como un constituyente clásico. ¿Cómo, entonces, Smolensky supone que el que tuviesen las representaciones mentales una estructura composicional débil, es decir, representaciones mentales dependientes del contexto, dé lugar a una explicación de la sistematicidad? Él simplemente no lo dice, ni siquiera dando una pista (op.cit., 193).

De hecho, independiente de si la estructura composicional débil explica la sistematicidad, esta explicación presenta dudosa coherencia. La primera incoherencia es que, si suponemos que la representación de café es tal en un contexto particular (el contexto que provee la taza), entonces ¿cómo es posible que una representación que no es café aisladamente, sino que meramente dentro de un contexto, pase a semánticamente dar sentido a la representación de “taza con café”? Los microrrasgos asociados con “café” no son, desde la perspectiva conexionista, *ni necesarios ni suficientes para hacer café*⁴⁴ (op.cit., 194).

Esta duda sobre la coherencia de la composicionalidad débil puede llevarse a casos más complejos donde representaciones altamente estructuradas requieran de, por ejemplo, representaciones de “taza

⁴⁴ En realidad F&M son más irónicos para decir esto, cito: "Take a liquid that has no properties specified by the microfeatures that comprise COFFEE in isolation, but that isn't coffee. Put it into a cup, et voila! it *becomes* coffee by semantical magic." (Fodor & McLaughlin, 1990, 194)

con café”, representación compleja que en sí misma es igualmente dependiente del contexto como “café”, el cual depende a su vez de microrrasgos, nuevamente ni suficientes ni necesarios (op.cit.).

Para F&M la segunda incoherencia que presenta esta noción débil de composicionalidad de Smolensky es que confunde la idea verdadera, pero poco interesante, de que cómo uno representa algún café depende del contexto, con *la idea tendenciosa de que la representación de “café” es dependiente del contexto*. Suponiendo que Smolensky es víctima de esta confusión, se puede dar sentido a muchas de las cuestiones intrigantes que él dice en su cuento del café (op.cit., 195).

Una vez discutidos los problemas de la propuesta de Smolensky sobre la composicionalidad débil, F&M presentan la propuesta de Smolensky sobre una estructura composicional fuerte para una red PDP. Esta propuesta de estructura composicional fuerte se diferencia de su versión débil en tanto que: primero, las unidades son explícitamente supuestas de modo de tener niveles de activación entre 0 y 1. Segundo, no se hace referencia a la idea de microrrasgos cuando se discute la composicionalidad fuerte. Tercero, introduce una nueva operación vectorial, la multiplicación a las operaciones ya mencionadas, de adición y sustracción. Cuarto, y en opinión de F&M, la diferencia más importante entre la propuesta “débil” y “fuerte”, *las nociones de representaciones mentales no son referidas como contexto-dependientes* (op.cit., 196).

La noción de estructura composicional fuerte apela a las ideas de representación mediante producto tensorial y superposición de representaciones⁴⁵. Sin embargo, el problema que ven F&M a tal propuesta es que el vector de activación representando el *filler* “J” será meramente imaginario, así como también el vector de activación representando el rol de la “primera posición de la letra”. Equivalentemente, para el vector de activación de producto tensorial representando “J” en la primera posición de letra. El único patrón de actividad que será, de hecho, implementado en la máquina, es el vector de superposición representando “Juan” (op.cit., 196-7).

F&M recuerdan lo que se ha dicho sobre C; *cuando un símbolo complejo clásico es implementado, entonces sus constituyentes lo son igualmente. Cuando un vector de producto tensorial o un vector de superposición es implementado, sus componentes no lo son, excepto por casualidad*. La implicancia de esta diferencia, desde el punto de vista de una teoría de los procesos mentales, es que mientras los constituyentes clásicos de una representación compleja están *ipso facto*, disponibles para contribuir a las consecuencias causales de sus implementaciones (en particular ellos están disponibles para proveer el dominio para las operaciones mentales), los componentes de un producto tensorial y vectores de superposición pueden no tener estatus causal como tal (op.cit., 197-8).

Además, vale la pena enfatizar que, en general, no existe una única descomposición del producto tensorial ni del vector de superposición en componentes. Perspicazmente, F&M observan que uno puede preguntarse legítimamente cuál es el sentido de hablar de *la* descomposición de una representación mental en constituyentes significativos, dada la noción de constitutividad que la teoría de Smolensky provee. En un modelo clásico la descomposición lingüística mediante *estructura arbórea*

⁴⁵ En el curso del artículo de F&M se da cuenta muy estrechamente de la propuesta más técnica de Smolensky. He preferido aquí no volver a abundar y me remito a lo expuesto en la parte anterior.

generará un árbol, que cuando está bien formado, individúa una descomposición única. Smolensky trata de dar una solución provisoria a este problema, que es que, en principio, dado que una red conexionista es un sistema dinámico, modos normales de un sistema cognitivo dinámico podrían generar descomposiciones únicas. No obstante, se observa, que no hay fundamentos fuertes para ser optimista respecto de la solución de Smolensky (op.cit., 198).

Ahora bien, ¿y cómo encaja el problema de la sistematicidad dentro de la propuesta fuerte? Un vector de producto tensorial y vectores de superposición “tienen constituyentes” en algún sentido extendido compatible: los vectores de producto tensorial tienen componentes semánticamente evaluables, y los vectores de superposición son susceptibles de ser descompuestos en vectores de producto tensorial semánticamente evaluables. Pero, la solución clásica al problema de la sistematicidad asume que *los constituyentes de las representaciones mentales tienen roles causales*; y que ellos proveen un dominio para las operaciones mentales. Los constituyentes clásicos de un símbolo complejo, así contribuirían a determinar las consecuencias causales de implementar tal símbolo y sería claro que los constituyentes “extendidos” de representaciones como las de Smolensky no pueden hacer esto (op.cit., 199).

A esto último, Smolensky respondería que el esquema representacional basado en superposición y producto tensorial no distorsiona la noción de constitutividad, dado que precisamente *este formalismo es usado en física para mostrar que el átomo como un todo está representado por un vector que es la suma de superposiciones de vectores*. F&M no dudan que el producto tensorial *puede representar* una estructura constitutiva, pero la pregunta relevante es si las representaciones de producto tensorial *tienen* estructura constitutiva, o si ellos tienen la clase de estructura constitutiva sensible a procesos causales. La diferencia entre psicología y física es que, mientras las leyes de la psicología son sobre las leyes causales que gobiernan las implementaciones de representaciones mentales, la física es sobre las leyes causales que gobiernan átomos, electrones y partículas de ese estilo. Puesto que, “ser una representación” no es una propiedad del dominio de la teoría física, la pregunta de si las representaciones mentales tienen estructura constitutiva no tiene análogo en física (op.cit., 200).

Por otra parte, si la eficiencia causal no es la meta de desarrollar la representación mediante producto tensorial, como Smolensky de algún modo dice, entonces se podría esperar que él tuviera alguna solución no clásica al problema de la sistematicidad, es decir, una solución que no dependa de asumir que las operaciones mentales sean causalmente sensibles a la estructura constitutiva. Recordemos que, la composicionalidad es la solución clásica que se da al problema de la sistematicidad, pero no necesariamente la única posible, aunque la argumentación de F&P junto con la de F&M dan a entender que es la más razonable. Lamentablemente, *Smolensky no da una salida al problema de la sistematicidad que no sea alguna variante de estructura composicional como la planteada en su propuesta débil y en su propuesta fuerte* (op.cit., 201).

El problema de la sistematicidad para los conexionistas no debe ser entendido como meramente la necesidad de mostrar que las capacidades cognitivas sistemáticas son *posibles* dadas las presuposiciones de la arquitectura conexionista, sino que explicar cómo la sistematicidad sería *necesaria* – cómo sería una *ley* que las capacidades cognitivas son sistemáticas – dadas estas presuposiciones. Sin duda, es posible para Smolensky cablear una red que soporte un vector que represente aRb, si solo si, soporta un

vector que representa bRa ; y quizás sea posible para él hacerlo sin explicitar sus “unidades imaginarias”⁴⁶. El gran problema es que, aunque la arquitectura permite esto, igualmente permite cablear una red que porte un vector que representa aRb , si solo si, soporta otro vector que representa zSq , o bien que represente “la guerra de las galaxias”. La arquitectura conexionista parece ser absolutamente indiferente a tales opciones (op.cit., 202).

En la arquitectura clásica, si uno conoce las condiciones para ser capaz de representarse aRb , uno no puede sino conocer las condiciones por las cuales es capaz de representarse bRa ; la arquitectura tiene este comportamiento debido a que: (i) la representación de a , R y b son constituyentes de la representación de aRb y (ii) se deben implementar los constituyentes de la representación compleja que es implementada, luego los constituyentes clásicos no pueden ser simplemente imaginarios. Por consiguiente, está *incorporado* en la visión clásica que no se puede pensar aRb , a menos que se sea capaz de pensar bRa , pero la visión conexionista es *neutral* sobre si se puede pensar aRb y pensar bRa . Es una *ley de la naturaleza* que uno no puede pensar aRb si uno no puede pensar bRa (op.cit., 203).

1.2. Una posible organización de los contenidos de la polémica a lo largo del tiempo.

Desde la publicación de los dos volúmenes PDP al año de la presente investigación que estoy llevando a cabo, ya han transcurrido tres décadas – y desde el momento en que se publica la última réplica de Fodor (1997) serán casi dos décadas – y en todo ese periodo de tiempo la producción de artículos sobre los contenidos detallados en la sección anterior no ha cesado ningún año. Esto es notable desde una perspectiva sociológica, no profundamente elaborada claro está, de la ciencia cognitiva, ya que significa que el impacto que ha tenido la polémica del conexionismo-simbolismo, no solamente ha dado quehacer a los involucrados más directamente, a saber, los adalides filosóficos e investigadores afines a cada bando, sino que *ha sido esencial para desarrollar el contenido teórico de la mayor parte de los enfoques más allá del simbólico*, tal como ha sido filosóficamente detallado por Fodor y Pylyshyn (1988).

Hilando más fino; el conexionismo no necesariamente debe interpretarse como dando una estocada fatal al simbolismo – menos al cognitivismo que tiene su propia publicidad gracias al trabajo de Steven Pinker (1997; 2003) –, sino que dando argumentos empíricos para fundamentar la necesidad de plantear alternativas explicativas al simbolismo, de modo de comenzar a ampliar el rango de posibilidades teóricas en ciencias cognitivas. Esta tendencia se fue acentuando en el tiempo, lo que se trasluce en que, si es que alguna vez hubo una hegemonía del simbolismo y el cognitivismo, el rango de posibilidades teóricas se fuera entremezclando, hasta el punto que ya no existe un enfoque dominante para tratar con la descripción y explicación de funciones cognitivas – y, por extensión, con la predicción de la conducta. Es por ello que no es exagerado afirmar que el debate simbolismo-conexionismo tuvo como consecuencia a largo plazo, también, el dar la forma contemporánea que poseen las ciencias cognitivas, forma que se trasluce en el contenido actual de muchos manuales introductorios al campo donde el

⁴⁶ Es decir, las representaciones de producto tensorial y de superposición, las cuales son “imaginarias” en tanto que carecen de un estatus causal según F&M.

“enfoque hegemónico clásico” solamente es una parte acotada del modo cómo tratar el fenómeno de la cognición⁴⁷.

Sin embargo, puesto que el impacto no puede ser dimensionado en su real magnitud en un trabajo acotado como el que estoy llevando a cabo, me limitaré a *organizar en parte las diferentes funciones que cumplen tanto la polémica simbolismo-conexionismo en su conjunto, como los elementos particulares de su contenido* – que han sido medianamente detallados en la sección precedente –, a partir de cierta cantidad de literatura de ciencias cognitivas que he seleccionado. La organización de la literatura es bastante sencilla; por una parte existen autores que tomaron el contenido de la polémica, o la polémica en su conjunto, y decidieron considerarla como un paso en falso respecto a la búsqueda de la verdadera naturaleza de la cognición. Por otra parte, existe literatura que continúa enfrascada en los detalles internos de la polémica y que con el tiempo se ha ido especializando en una cadena de intercambios que va siendo matizada, justamente, por la primera clase de contenido antes mencionada. De manera bastante simple la primera clase de artículos y textos será aludida como “considerando la polémica externamente”, mientras que la segunda como “considerando la polémica internamente”. Ergo, lo que será presentado a continuación no debe ser tomado como un estudio concienzudo de la evolución de los contenidos de la polémica simbolismo-conexionismo, sino más bien una manera práctica de introducirse a todas las aristas visibles que legó al campo de ciencias cognitivas tal intercambio desarrollado en (1.1.).

1.2.1. Considerando la polémica externamente.

(A) Considerando la polémica externamente: Es decir, esta literatura toma la polémica como parte de la secuencia evolutiva de un perfeccionamiento de la teorización sobre “el qué es realmente” la cognición, llevando a la aparición de otros “paradigmas” en las Ciencias Cognitivas, lo que enmarca al simbolismo y al conexionismo como fases que debían ser *superadas* para engendrar, tanto un modo de explicación novedoso, como una concepción ontológica “más precisa” de la cognición. Esta clase de literatura se puede subdividir en dos tipos dependiendo la generalidad con la que tratan la contienda simbolismo-conexionismo, así como también dependiendo del grado de énfasis con el cual se rechaza los lineamientos implícitos tanto en cognitivistas como conexionistas:

(A0) Literatura de objetivos generales: esta subclase de literatura coloca históricamente al simbolismo cognitivista o al conexionismo, o a ambos, como fases que han recorrido las ciencias cognitivas en la búsqueda de un “paradigma” mejor. Contiene a todos los que se han venido a denominar *enfoques alternativos*, que en parte se justifican mediante alguna clase de evaluación de los límites conceptuales

⁴⁷ Considérese algunos casos: La sexta edición de *Cognitive Psychology A Student's Handbook* (Eysenck & Keane, 2010) en su prefacio ya apunta a la ciencia cognitiva computacional como uno de varios enfoques para acercarse a las funciones cognitivas, las cuales se detallan con profundidad en el contenido mismo del manual entremezclándose en cada capítulo los diferentes enfoques (experimental, neurocientífico-cognitivo, neuropsicológico y computacional), poniéndose el acento está primordialmente en las funciones cognitivas y no en los enfoques. Una estrategia análoga seguirán Kolak, Hirstein, Mandik & Waskan (2006, primeros dos capítulos) y Britt Anderson (2014). Ligeramente más tradicional es la de José Luis Bermúdez (2014), donde el énfasis está puesto en la integración – y se introducen las ciencias cognitivas en la primera parte del libro sin desconocer ningún hito sin sesgarse por cognitvismo alguno –, desafío que es introducido con las herramientas ya bien conocidas desde el cognitvismo, tales como los niveles de análisis de Marr o la postulación de arquitecturas cognitivas.

en principio que poseerían simbolismo cognitivista y conexionismo. Normalmente, tales consideraciones son abstractas y no toman en cuenta modelos conexionistas o simbólicos específicos, sino que éstos pasan, en el mejor de los casos, a ser subsumidos por el “enfoque alternativo” en cuestión.

Dos ejemplos famosos, debido a que engendran por sí solo alternativas al simbolismo y conexionismo bien diferenciadas y con una identidad clara, son el *enactivismo* propugnado por Francisco Varela, Evan Thompson y Eleanor Rosch (1991) y el *dinacimismo* propuesto conjuntamente por autores como Ester Thelen y Linda Smith (1996) y Timothy Van Gelder (1995; 1998). Si se revisa sus trabajos puede notarse cómo la existencia del debate, no sólo los hizo en su momento tomar una postura (Van Gelder, 1990)⁴⁸, sino que paralelamente ir construyendo una alternativa que trascendiera teóricamente los “paradigmas” en oposición. No obstante, lo más interesante es que tales explicaciones alternativas aparecen cronológicamente justo posteriormente al debate simbolismo-conexionismo⁴⁹. A continuación profundizaré algo sobre tales alternativas y su relación con el simbolismo y el conexionismo:

Enactivismo:

Cuando se investiga lo mental, en el curso del entrenamiento y práctica de ciencia y filosofía occidental, el modo de proceder primero es con la reflexión teórica que se pregunta cosas como “¿qué es la mente?” o “¿qué es el cuerpo?”, para después proceder a la investigación científica experimental. Este procedimiento engendra una miríada de aseveraciones, experimentos y resultados con diversas facetas de las capacidades cognitivas, pero se olvida del rol de quién está haciendo las preguntas y la forma en que lo hace. Luego, al no incluirnos a nosotros mismos en la reflexión teórica, se estará persiguiendo una reflexión parcial, tornándose así nuestras preguntas en “descorporizadas”. (Varela et al., 1991,27)

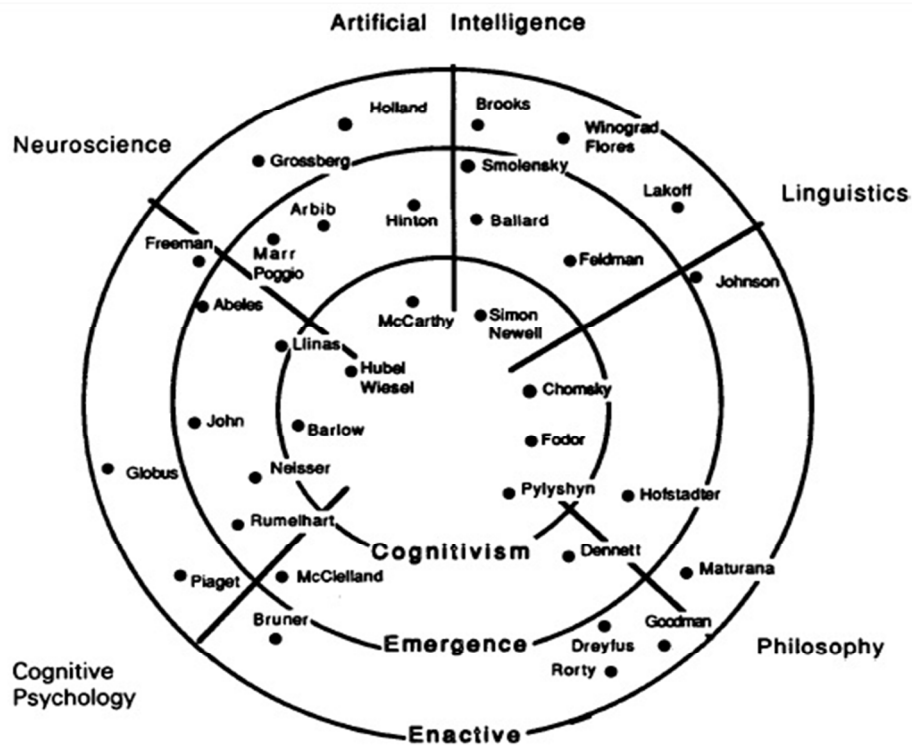
Así pues, la cuestión central y fundamental del yo o el sujeto cognoscente será relegado a una actividad puramente teórica, aun cuando, este problema obviamente toca nuestras vidas y la auto-comprensión directamente. La opinión de Varela et al. será entonces que, no solamente se han dejado de lado fenómenos cognitivos cruciales por explicar, sino que el estilo actual de la investigación es limitada y poco satisfactoria, tanto teórica como empíricamente (op. cit, xvii-xviii). El sesgo que tomarán los autores – y que reconocen explícitamente – para avanzar en un mejor estilo de investigación (tanto teórico como experimental) será pues hacer hincapié en la diversidad teórica de las ciencias cognitivas interpretadas como consistente en tres etapas sucesivas, las cuales son graficadas en una especie de mapa “polar” conformado con tres anillos concéntricos⁵⁰ (FIGURA 5 en la página siguiente).

⁴⁸ En particular este artículo propone la posibilidad de redefinir la composicionalidad no como *concatenación*, sino que *funcionalmente*, definición que es más amplia que la por concatenación (Van Gelder, 1990, 361-2). Dado que es un artículo crítico, entonces puede ser categorizado dentro de (B2). No obstante, puesto que el autor trabajaba paralelamente en esa época en la elaboración del marco de un enfoque alternativo, entonces debe considerarse como parte de lo que he clasificado “literatura externa de objetivos específicos” o (A1).

⁴⁹ Si bien es cierto que la disputa simbolismo-conexionismo permitió dar mayor visibilidad a los enfoques alternativos, ello no debe interpretarse que tales alternativas no preexistieran al supuesto periodo hegemónico del cognitivismo, dado que la historia de las ciencias cognitivas muestra, precisamente, que “lo alternativo” preexistía antes del advenimiento de simbolismo. Esto será un poco más claro con el esquema de la sección (2.5.).

⁵⁰ En realidad este modo de “cartografiar” las ciencias cognitivas es anterior a Varela et al. Considérese el mapa presentado por Dennett en *The logical geography of computational approaches: A view from the East Pole* (1986)

Las tres etapas se corresponderían con un movimiento sucesivo desde centro a periferia, donde el centro comprendería el cognitivismo, la circunferencia media subsumiría el conexionismo entendido como parte de los "enfoques emergentes" y la periferia que envuelve todo sería el enactivismo (op. cit., p. 6-7)



Los dos primeros círculos – donde se encontraría ubicado propiamente el debate tratado en (1.1.) y subsecciones – son presentados con cierta profundidad en su obra y, luego, sumariamente sintetizados por Varela et al. del siguiente modo (op. cit., 42-3/98-9):

	<i>Cognitivismo</i>	<i>Emergentismo</i> ⁵¹
<i>¿Qué es la cognición?</i>	Procesamiento de información entendida como la manipulación de símbolos llevada a cabo mediante computación simbólica basada en reglas.	La emergencia de estados globales en una red de componentes simples.
<i>¿Cómo funciona?</i>	Mediante todo dispositivo capaz de portar y manipular elementos discretos funcionales o símbolos. Tal sistema interactúa solamente con la forma de los símbolos (sus atributos físicos) y no con su significado.	A través de reglas locales para operaciones individuales y reglas para los cambios en la conectividad entre los elementos
<i>¿Cómo se sabe cuándo un sistema cognitivo funciona de manera adecuada?</i>	Cuando los símbolos representan apropiadamente algún aspecto del mundo real, y el procesamiento de la información permite llegar al sistema a una solución exitosa para algún problema.	Cuando las propiedades emergentes (y su estructura resultante) puede ser vista como correspondiendo a una capacidad cognitiva – una solución exitosa a una tarea requerida.

(reimpreso en Dennett, 1998). Lo novedoso de la presentación de Varela et al. frente a la del filósofo es el etiquetar los diferentes anillos y denominar el más externo y comprehensivo como "enactivo".

⁵¹ Varela et al. aclaran que esta orientación alternativa subsume al conexionismo, y que a esta orientación pertenecen visiones divergentes sobre qué es la ciencia cognitiva y su futuro.

La exterioridad desde el segundo círculo la habitarían autores como Jackendoff o Minsky y Papert, quienes tendrían propuestas que tratarían ir más allá del emergentismo. En específico, por ejemplo, Varela et. al. arguyen que el modelo de la mente como una sociedad de numerosos agentes de Minsky y Papert – modelo que intenta abarcar una multiplicidad de enfoques para el estudio de la cognición: distribuidos, redes auto-organizadas, módulos localizados, procesamiento en serie, etc. – sería una especie de camino intermedio en las ciencias cognitivas. Este camino intermedio va más allá de una concepción homogénea de la mente, al no adherir totalmente ni a redes conexionistas ni a procesadores simbólicos seriales. Esta búsqueda de equilibrio entre los extremos es evidente cuando Minsky y Papert argumentan que la virtud de un sistema cognitivo está tanto en la distribución como en el aislamiento, o sea, el que los mecanismos mantengan diversos procesos separados. Así pues, los agentes dentro de una agencia pueden estar conectados en forma de una red distribuida, pero debiesen existir, a la vez, mecanismos para mantener las diversas agencias aisladas entre sí, permitiéndose así que las agencias interactúen limitadamente, como el típico procesamiento secuencial simbólico prescribe (op. cit., 106-7).

Sin embargo, el enfoque de Minsky y Papert es incompleto, así como también la propuesta de Jackendoff, ya que ninguno de ellos estaría investigando, por ejemplo, una faceta fundamental del ser humano; cómo la práctica meditativa puede transformar la consciencia, lo cual también significará en el fondo las ciencias cognitivas – ni los caminos intermedios hasta entonces desarrollados – no son relevantes para explicar la experiencia humana como tal (op. cit., 126-7).

En este punto se vuelve más claro que parte de la inspiración decisiva – ésta es una presentación simplificada del enactivismo, al punto que he omitido la larga interiorización de Varela et al. a la fenomenología de Merleau-Ponty como inspiración de su propuesta (capítulos 1, 2 y 4) – para la propuesta conceptual de Varela et al., a saber, la *enacción*, es la noción de *sunyata* budista o *vacuidad*, es decir la noción de la no existencia de fenómenos por independientes por sí mismos, sino que siempre dándose de manera interrelacionada o codependiente (op. cit., 220-4). Luego la cognición, desde su perspectiva más fundamental, carecería de sustancia separable del ambiente, lo que es condensado precisamente en la noción de “enactuar” donde *nada es previamente dado*, sino que está siempre co-construyéndose en el bucle de interacción organismo-ambiente. En términos de Varela et al.: “In a nutshell, the enactive approach consists of two points: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided.” (op. cit., p. 173). Las dos características centrales de la enacción pueden ser ejemplificadas del siguiente modo:

Primero, por percepción definida como “acción perceptualmente guiada” se entenderá que la percepción no está suficientemente definida agregando la pura corporización a la cognición, sino que la acción es decisiva para definir la percepción. Considérese representativo para tal definición el experimento de Held y Hein, donde a un grupo de gatitos recién nacidos se le permitió una experiencia visual del mundo sin restricciones motoras, mientras a un segundo grupo de gatitos se les ató a un arnés y carrito que los unía a cada uno a un gatito del primer grupo, de modo que ambos grupos compartieran la misma experiencia visual, pero siendo el segundo grupo pasivo en su descubrimiento motor del mundo al contrario del primero. El resultado del experimento fue que el primer grupo de gatos se desarrolló normalmente y que el segundo, aun cuando podían ver, no fueron capaces de interacción

motora con el ambiente, lo que confirma que la percepción es acción es perceptualmente guiada (op. cit., 174-5).

Y, segundo, se ejemplifica la emergencia de estructuras cognitivas mediante la recurrencia de patrones sensoriomotores que permiten la acción perceptualmente guiada con la investigación de Rosch sobre la actividad cognitiva de la categorización. Rosch propondría un nivel básico de categorización en las taxonomías de objetos concretos en los cuales biología, cultura y necesidades cognitivas confluirían en su grado de economía e informatividad. Este nivel básico de categorización sería el nivel más inclusivo en el cual los miembros de una categoría (1) son usados, o se interactúa con ellos mediante acciones motoras parecidas, (2) tienen formas similares percibidas y puede ser imaginados, (3) tienen atributos identificables humanamente significativos, (4) son clasificados por niños pequeños, y (5) tienen primacía lingüística (en varios sentidos). Luego, el nivel básico de categorización sería el punto en el cual la cognición y el medio ambiente son “enactuadas” simultáneamente (op. cit., 177).

Así pues, el enactivismo como ha sido aquí sintetizado, se popularizó como la base de un proyecto alternativo en ciencias cognitivas que sigue vigente hasta el día de hoy y se ha erigido como común a muchas investigaciones que tienen en cuenta varias características derivadas de la obra original de Varela et al., tales como: rechazo de las representaciones internas desencarnadas – y por extensión a toda clase de representación, incluida la distribuida conexionista –, la necesidad de conciliar la experiencia de primera persona con la perspectiva la tercera persona, volver a dar un lugar relevante explicativamente a la cultura y la interacción en las explicaciones cognitivas, y, en general, con todas aquellas facetas de lo mental que indirecta o directamente sean susceptibles de ser tratadas a partir de la noción de enacción (Stewart et al., 2010).

Dinamismo – o “dinamicismo”⁵²

Thelen & Smith (1996) presentan su obra como llenando un espacio dentro de las teorías del desarrollo que no ha sido aún ocupado ni por el innatismo⁵³, ni por el ambientalismo ni por el interaccionismo, las cuales serían teorías imperfectas, porque esencialmente *prescriben* la forma adulta antes del propio desarrollo. Estos puntos de vista no tienen en cuenta el *proceso*, de cómo nuevas formas y funciones se materializan a lo largo del tiempo. El desarrollo no es la especificación de un resultado – del producto –,

⁵² Zednik (2009) distingue tres familias de investigación dinamicista: (i) *Kelsonian dynamicism* (ii) *Dynamical field theory* (iii) *Dynamical agent modeling*. El caso de Thelen y Smith cae en el segundo grupo. Estas variedades son importantes distinguirlas, ya que cuando se presente sucintamente el mecanicismo de Bechtel y Abrahamsen en la sección (2.1.) se apelará al dinamicismo antirepresentacionalista incompatible con el mecanicismo, es decir, la escuela derivada de Kelso con su obra – a estas alturas un referente para dinamicistas antirepresentacionalistas como Anthony Chemero y compañía – *Dynamic Patterns: The Self-Organization of Brain and Behavior* de 1995. He elegido la presentación de Thelen & Smith (1996) porque a estas alturas su visión se ha transformado en la dominante dentro de la subdisciplina de las explicaciones del desarrollo en ciencias cognitivas (Spencer et al., 2011). Esta nota de advertencia ha sido omitida en el caso del enactivismo, puesto que es más claro que el origen conceptual de éste está circunscrito a Francisco Varela. En el caso del dinamismo la cuestión de su origen no es tan obvia y requiere un trabajo aparte que puede emprender el lector interesado en las variedades dinamicistas.

⁵³ En el original se utiliza el término “maturationalism”. Por “maturationalism” se entiende aquella postura sobre el desarrollo que supone la existencia de etapas predeterminadas que debe recorrer el organismo hasta su madurez, siendo así básicamente una forma de innatismo y por eso he decidido dejarlo como tal.

sino que es la ruta a través de la cual el organismo se mueve desde un estado temprano a otro estado más maduro. Al asumirse la prescripción, o la teleología, simplemente resta afinar el proceso, ya que el resultado se encapsula en el plan (Thelen & Smith, 1996, xv). Luego, una teoría del desarrollo debiese cumplir con las siguientes metas centrales (op. cit., xvii-xviii):

- Permitir comprender los orígenes de la novedad
- Reconciliar las regularidades globales con la variabilidad local, la complejidad y la especificidad contextual.
- Integrar datos del desarrollo a través de muchos niveles de explicación.
- Proveer una explicación biológicamente plausible, pero no reduccionista, del comportamiento.
- Comprender cómo procesos locales permiten las salidas globales.
- Establecer las bases teóricas para generar e interpretar datos empíricos.

El capítulo 2 de Thelen & Smith contiene críticas tanto al “simbolismo” y “cognitivismo” como al conexionismo. Es necesario entrecomillar “simbolismo” y “cognitivismo”, porque las autoras hacen ciertas subdivisiones a precisar: la división racionalista-innatista y la de procesamiento de información (p38), subdivisiones que corresponderían *grosso modo*, respectivamente, al cognitivismo y simbolismo. Ambas subdivisiones teóricas son desestimadas con variados argumentos tales como:

- La falta de plausibilidad biológica y falta de restricciones empíricas de la distinción competencia/actuación, la cual deviene del racionalismo aplicado a la generación de modelos lingüísticos. (op. cit., 25-7).
- Segundo, el innatismo justificado como una respuesta que evita los peligros del argumento quineano de la indeterminación de la traducción, carece de fundamento cuando se toma en cuenta que el aprendizaje en el desarrollo es algo más allá de mera inducción y que la cognición puede ser también algo más allá de la formación de hipótesis y creencias (op. cit., 28-30).
- Tercero, la continuidad, que es un argumento para buscar estructuras de conocimiento que se mantengan inalterables a lo largo del desarrollo, parece débil ante evidencia que podría apoyar tanto la continuidad como la discontinuidad en niveles macro del desarrollo (op. cit., 30-32).
- Cuarto, la modularidad, al igual que la continuidad, pareciera ser un fenómeno justificable sólo localmente, pero no en la globalidad del desarrollo (op. cit., 34-7).
- Quinto, y finalmente, la metáfora de la “mente como máquina” que lleva a la descomposición en términos de mecanismos y a la descomposición de dominios de tareas separados se traduce en un fuerte énfasis en la pregunta sobre el *qué se desarrolla*, pero deja de lado la pregunta más general que conectaría todos los modelos de procesamiento de información del *cómo se desarrolla* (op. cit., 37-8). Nótese que esta última observación deviene trivialmente de la crítica que da pie y sustenta la obra dinamicista, a saber, que se asume tradicionalmente la prescripción del estado maduro en las teorías del desarrollo.

Por otra parte también hay un desarrollo en contra el conexionismo, de porqué en principio la propuesta conexionista tampoco es factible como base para una teoría del desarrollo (op. cit., 41-2):

- El primer problema del conexionismo es que no da cuenta de la heterogeneidad de estructuras cerebrales, pasando los nodos a ser homogéneos, siendo ello claramente ejemplo de que el conexionismo no es "estilo cerebro".
- La segunda falla, derivada de la primera, es que los sistemas conexionistas son "sistemas de juguete", solamente capaces de solucionar problemas aislados. Esto va contra la evidencia que inclusive niños son capaces de solucionar simultáneamente muchos problemas y realizar diferentes tareas. El dinamicismo tiene como meta dar cuenta del hilo del desarrollo en su complejidad absoluta sobre un sistema intacto.
- El tercer punto en contra es que mucho del modelamiento conexionista tiene como meta el análisis matemático del espacio teórico, meta que supone la existencia de un estado final o teleológico.
- El cuarto error deriva de una falla de inspiración. La teoría conexionista cae en responder viejas preguntas de una nueva manera, sin percatarse que las preguntas en sí mismas son las equivocadas. La psicología cognitiva tradicional busca entender qué es estable y constante, por ejemplo se explica qué es entender la palabra "gato" mediante la alusión a representaciones, y el conexionismo la única variación que presenta respecto al enfoque tradicional es formalizar las representaciones distribuidamente.

Desde la perspectiva de Thelen y Smith, el conexionismo es demasiado cercano a la teoría cognitiva tradicional, en cuanto que trata de explicar la estabilidad de la cognición y la mide con la misma vara del enfoque tradicional, es decir, busca comprender tareas cognitivas guiadas por la manipulación símbolos, pero sin "símbolos", o busca conducta guiada por reglas, pero sin apelar a "reglas". Dado que la teoría del desarrollo que buscan Thelen y Smith está fundamentada en el cambio de las preguntas esenciales, entonces el dinamicismo trata de ir más allá de la presuposición de un estado-final del desarrollo. Para ello utilizarán un acercamiento mediante el formalismo de los sistemas dinámicos, formalismo que ha sido aplicado a diversas ciencias empíricas y que permite representar los principios generales de formación de patrones que evolucionan en el tiempo (op. cit., p. 51).

La obra contiene posteriormente, no sólo un contenido puramente conceptual como lo sintetizado hasta aquí, sino que también hay tramos completos dentro del libro que tratan la operacionalización en contextos experimentales de los conceptos del paradigma dinamicista (op. cit., 99-125). Además de tal operacionalización, hay sugerencias sobre cómo el dinamismo puede ser complementado con explicaciones basadas en mecanismos neuronales, proponiéndose adherir – al menos heurísticamente – a la teoría de la selección de grupos neuronales de Edelman, la cual se basa en la existencia de una organización dinámica del cerebro (op. cit., 131-151). Sin embargo, y finalmente, más allá de estos brevísimos apuntes para quien quisiera abocarse al estudio del contenido teórico mismo del dinamicismo, me interesa rescatar el posicionamiento totalmente radical frente al simbolismo que es de manera clarísima declarado por las autoras reafirmado casi final de su obra:

We, like the symbolic computational theorist, view cognition as all one kind; but in our view; it is all embodied, all distributed, all activity, all a complex event in time. We believe that our theory of activity-driven, reentrant, high-dimensional cognition offers the best hope for understanding symbolic thought. We propose that symbolic thought, like walking up and down slopes, like the A-not-B error, like interpreting novel words, is emergent in activity and in the products of those

activities in the physical world and on us. As we act and explore our world, our actions will produce concrete physical changes in that world that we may perceive. By perceiving the products of activity and the reentrant mapping of that perceiving onto the ongoing activity from which it emerged, we create external symbols and discover their value. (op. cit. 337)

Otros casos

Posterior al debate simbólico-conexionista, además de Varela et al. (1991) y Thelen & Smith (1996) existen obras que contienen ideas de alternativas de la cognición, – por ejemplo, definida como acción/percepción, corporalización, emergencia, etc. – y ejemplo de esta clase de obras post-debate con objetivos generales, podrían nombrarse Menary (2007), Calvo & Gomila (2008), Chemero (2009). Entrar a detallar, aunque fuera superficialmente, el contenido de tales obras alargaría innecesariamente la presente parte de la tesis y considero que con los ejemplos del dinamicismo y enactivismo es suficiente para ejemplificar (A0). Queda al lector especialista más entusiasta la empresa de organizar y sistematizar la heterogeneidad de la miríada de enfoques alternativos contemporáneos.

(A1) Artículos o textos especializados de variadas clases, que contienen trozos conceptuales más particulares y mejor delimitados de la polémica, o directamente tocan exclusivamente un concepto del debate, desarrollándolos y criticándolos desde la óptica de los intitulados “paradigmas alternativos”. Esta categoría de textos es menos general que la anterior y, más que contener el desarrollo de un nuevo enfoque, proponen una crítica a las nociones antiguas del simbolismo y conexionismo.

Para ejemplificar (A1) considérese el caso de la **representación**⁵⁴, la cual dentro del contexto del debate tiene dos formas bien definidas – distribuida versus local –, pero que más allá del debate es puesto en duda. Quien articula tal duda de manera paradigmática es Van Gelder (1995), quien, para mostrar la no necesidad de usar representaciones como las usadas en el marco computacional, utiliza el siguiente argumento general: postula dos clases de solución al problema de traspasar el movimiento oscilatorio de un pistón que funcionaba por el vapor a una rueda para que esta girase de manera uniforme sin grandes variaciones en su velocidad. Por una parte está la solución real dada por James Watt que consistió en su famoso “regulador centrífugo”. Y, por otra parte, un hipotético regulador computacional cuya tarea principal puede ser dividida en un algoritmo como el siguiente (Van Gelder, 1995, p. 348):

- 1.- Medir la velocidad de la rueda
- 2.- Comparar la velocidad actual con la velocidad que se desea
- 3.- Si no hay discrepancia retornar al paso 1. De otra manera:
 - a.- Medir la presión actual del vapor
 - b.- Calcular la alteración deseada en la presión del vapor
 - c.- Calcular el ajuste necesario de la válvula
- 4.- Hacer el ajuste de la válvula y retornar al paso 1.

⁵⁴ Una “representación” tendría los rasgos de: ser capaz de mapear el ambiente, ser intencional – “aboutness” –, ser asimétrica de manera de distinguirse de las relaciones de semejanza comunes, “standing-in” o la característica de tener contenido en relación con otras representaciones y de tener poder causal para guiar la conducta. (Nagel, 2005) En mayor o menor grado conexionistas y simbólicos aseveran que sus representaciones poseen tales características, o al menos, que en principio pueden portarlas, dado que, por ejemplo, tanto para conexionismo como para el cognitivismo es difícil dar cuenta del “aboutness” si se toma en cuenta la crítica de Searle (1980)

Lo que se quiere subrayar en esta crítica a la representación no radica en la diferencia del cómo se soluciona la tarea, sino que en que el regulador de Watt no puede ser considerado como representacional de ningún modo y que, aun así, es capaz de “implementar un algoritmo” de modo de encontrar una solución óptima a un problema. Cuatro argumentos van en contra de considerar al regular de Watt como representacional (op. cit., 351-353):

i. El ángulo del brazo del regulador puede dar la impresión que “representa” la velocidad de la rueda mientras gira, dada la correlación entre estos dos elementos. Empero, esta idea es insuficiente para asignar un estado representacional al regulador, dado que no se puede caracterizar tal correlación como “standing in”, es decir portando información.

ii. Además, la correlación entre ángulo del brazo y la velocidad de la rueda es contingente exclusivamente en el estado de equilibrio, cuando no está variando la posición del regulador. Dado que casi siempre está variando la presión del vapor, entonces ni siquiera se puede hablar de una correlación completa.

iii. El sistema podría ser considerado representacional siempre y cuando las representaciones explican algo de modo que se gana con la intervención de éstas. Sin embargo, el sistema de Watt es descrito y explicado sin necesidad de representaciones.

iv. La explicación por representación es demasiado simplista para ser correcta para un sistema como el del regulador y la rueda, y aquí interviene propiamente el dinamicismo como el marco adecuado para hacer frente a esta inadecuación de la explicación computacional.

Con estos cuatro puntos quedaría demostrado que el dispositivo es, según Van Gelder, ni representacional ni computacional. Sin embargo, esta argumentación contraria a las representaciones desde el dinamicismo no está ni cerca de ser reconocida como decisiva. Para Clark y Toribio el problema básico que subyace a esta clase de ejemplos, es que pertenecen al dominio de tareas donde, sin excepción, *existen estímulos ambientales adecuados que pueden ser utilizados en lugar de representaciones internas*, luego no siendo suficientemente “hambrientos de representación”. Un problema dentro de este último tipo sería (Clark & Toribio, 1994, 418-419):

1. Un problema que involucre razonamiento sobre estados de hechos ausentes, no existentes o contrafácticos.

2. Que requiera que el agente sea selectivamente sensible a parámetros cuyas manifestaciones físicas sean complejas y sin reglas, por ejemplo en una disyunción sin un último término.

La primera clase de problemas es conocida, ya que la capacidad de seguir un estímulo distal o inexistente, *prima facie* requiere el uso de algún tipo de recurso interno que permite la conducta coordinada apropiada sin un estímulo constante del ambiente que nos guíe. Esta capacidad de razonar sobre lo ausente o lo plausible no puede - ni debería - ser descartado como "punta del iceberg" del fenómeno cognitivo. La segunda clase de casos es igualmente convincente, pero menos intuitiva. La idea detrás de esta clase es que existe un desarrollo de la sensibilidad a los estados del mundo cuyas manifestaciones como estímulo están al menos "atenuadas". Estos serían estados de hechos que serían altamente relacionales, funcionales o abstractos. Es difícil dar cuenta de la capacidad de seguimiento de un sistema sobre propiedades muy diferentes bajo una rúbrica común, a menos que existiese un patrón

o estado interno cuyo contenido corresponda a la propiedad abstracta en cuestión (Clark & Toribio, 1994, 419).⁵⁵

La representación es un concepto que, no sólo ha sido puesto en entredicho por el dinamicismo, sino que también por la denominada “nouvelle IA” de Rodney Brooks, quien ejemplificando a través de sus robots basados en una arquitectura de subsunción diseñada por él mismo, muestra que se puede lograr un comportamiento robusto en ambientes desconocidos sin recurrir a la noción clásica de representación perteneciente a la IA tradicional (Brooks, 1999). Quizás su más explícita tesis al respecto sea: “representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems” (Brooks, 1991, 140). Sin embargo, extenderme más sobre el estado de la cuestión me llevaría a dispersar demasiado el foco de la presente exposición.

Otros casos.

Más allá del tópico de la representación, la literatura crítica a la polémica simbólico-conexionista toca nociones como la de sistematicidad, composicionalidad o productividad, y subyace un cúmulo de otros conceptos más susceptibles de enlistar. Sin embargo, la categoría (A1), al ser definida por su carácter de crítica específica a conceptos simbólicos y/o conexionistas, se traslapa con la categoría (B2) que se expondrá en la subsección a continuación.

1.2.2. Considerando la polémica internamente.

(B) Considerando la polémica internamente: Esta clase de literatura considera que la polémica contiene contenido teórico relevante, tanto en el sentido de subsumir herramientas conceptuales necesarias para tratar el fenómeno en cuestión, como por describir fenómenos que requieren una explicación. Se puede subdividir en, al menos, tres tipos bien diferenciados de temáticas; primero, (B0) como profundizando los argumentos a favor y en contra ya esgrimidos por el grupo PDP y los conexionistas. Segundo, (B1) como propugnando una fusión de ambas clases de modelamiento. Y, tercero, aquella más crítica en sus contenidos, (B2) poniendo en duda parte del contenido del intercambio.

(B0) Literatura que extiende el contenido propio del debate; categoría que toma postura sin juzgar una salida intermedia que sea, o bien salirse de los bandos – como sucede en (A) –, o bien un intentar una construcción teórica a partir de ambos, como se da con aquel grupo de investigaciones enfocadas en hibridar ambos modelos. Pareciera haber más a favor del conexionismo que a favor del lado simbólico. A continuación me explayaré brevísimamente sobre una muestra de algunos artículos de cada bando:

A favor del conexionismo:

- *Connectionism, Classical Cognitive Science and Experimental Psychology* (Oaksford et al. 1990)

En tareas relacionadas a memoria e inferencia, mientras más rico es el contenido de los materiales usados el desempeño del sujeto mejor parece ser. Esto es misterioso desde la explicación clásica, pero

⁵⁵ Vale la pena notar que Andy Clark – al menos el de las obras revisadas para esta tesis – es un revisionista representacional y que, según Clark & Toribio (1994) Van Gelder no es anti-representacionista, sino que éste dejaría espacio a la evaluación caso por caso para la explicación representacional. Quizás la mejor imagen de la postura de Andy Clark está en *Being There* (1998), donde no desecha las representaciones, sino que es ecuménico.

no para el conexionismo. Codificar un estímulo más rico requiere más fórmulas a ser almacenadas en una base de datos declarativa y tanto el almacenamiento como la recuperación se vuelven más difíciles; en la visión clásica los límites de recursos computacionales se vuelven más aguzados mientras más rico es el estímulo. Sin embargo, la evidencia psicológica indica que la memoria humana no se ve mermada, sino que mejorada cuando el estímulo es rico. En el caso de la inferencia, según la explicación clásica de la cognición, el razonamiento de sentido común involucra la construcción de derivaciones lógicas muy elaboradas. Luego, el razonamiento de sentido común debiese ser mucho más difícil para los sujetos humanos que las deducciones lógicas triviales de prueba de teoremas. No obstante, el desempeño exactamente opuesto se observa en los datos psicológicos, por ejemplo con la tarea de selección de Wason (op. cit., 79). En síntesis, las limitaciones y patrones de errores e interferencia en tareas de memoria e inferencia parecen misteriosos desde la visión clásica, lo cual lleva a concluir que (op. cit., 87):

- En contra de las expectativas clásicas, el desempeño en memoria e inferencia mejora con la riqueza del estímulo y puede relacionarse al vasto arreglo interconectado de restricciones anulables las cuales codifican la organización predecible del mundo. Los sistemas conexionistas son capaces de explotar tal redundancia y, por ende, emular las características observables de la memoria humana, tales como capacidad de direccionamiento por el contenido en la memoria y degradación elegante.
- El desempeño humano en, a primera vista, tareas simbólicas está fuertemente influenciado por el contenido de los materiales. Luego, en la medida que los procesos simbólicos son implementados en el cerebro, como argumentan también F&P, éstos crucialmente debiesen interactuar con las propiedades de la implementación.
- Esto conduce naturalmente a la conclusión que la psicología cognitiva moderna ha estado siempre comprometida con la interacción entre los niveles cognitivo e implementacional. En consecuencia, la defensa de F&P por la autonomía es una restricción que, si los psicólogos adoptan, conduciría al desinterés por el conexionismo, privándose así a la psicología cognitiva de la mayor parte de su objeto y una de sus distinciones explicativas más potentes. Por lo tanto, las cuestiones de la implementación de la arquitectura cognitiva debiesen seguir siendo una fuente importante de restricción en la teorización psicológica. En síntesis, el conexionismo puede, naturalmente, ser visto como proporcionando una nueva e importante metáfora para el pensamiento, más que como una teoría de la implementación psicológicamente relevante.

- *Why Fodor and Pylyshyn Were Wrong: The Simplest Refutation* (Chalmers, 1993)

La “refutación más simple” de Chalmers sigue cuatro pasos:

- i. En el argumento de F&P no hay modelos conexionistas que tengan semántica composicional, tampoco hay cláusulas de escape excluyendo modelos como implementaciones clásicas de la conclusión.
- ii. Si el argumento de F&P es correcto como se mantiene, entonces se establece que ningún modelo conexionista puede tener semántica composicional.
- iii. Pero algunos modelos conexionistas obviamente tienen semántica composicional, a saber, implementaciones conexionistas de los modelos clásicos.
- iv. Luego, el argumento de F&P no es correcto si es correcto (de ii y de iii).

Chalmers establece una analogía entre el argumento de F&P y un científico loco que ha “probado” mediante un argumento *a priori* que la Tierra es el único planeta habitado en el universo. El argumento de este científico loco sería que es imposible que la clase de correcta de bioquímicos sea alguna vez ensamblada de la manera correcta, requisito básico para la complejidad organizacional de la vida. Entonces concluiría “la vida no ha surgido en ningún planeta dentro de este universo”. Por supuesto es obvio que la vida existe en la tierra, ante lo cual la respuesta del científico loco sería “ya sabemos eso, lo que he establecido es que la vida no puede aparecer en ningún lugar excepto en la tierra”. Esto nos suena bastante *ad hoc* y extremadamente pobre en su lógica. Análogamente, el argumento de F&P debe ser arreglado para que excluya las implementaciones clásicas de su propia conclusión, lo que significaría responder por qué ciertos modelos (implementaciones y otros) están fuera de los efectos del propio argumento de F&P.

Un defensor de F&P podría decir “claramente el argumento no aplica a las implementaciones de los modelos clásicos, ya que las implementaciones son diferentes a las representaciones de los símbolos clásicos, puesto que en tales no existiría representación a nivel del nodo, sino que en un nivel mucho más alto; estos símbolos serán los capaces de combinarse composicional y autónomamente”. Tal defensor habrá descubierto el poder de la representación distribuida.

Para Chalmers es posible codificar la información composicional en representaciones distribuidas, por tanto, a continuación se hará cargo del argumento de F&M que asevera que para que exista procesamiento sensible a la estructura, una representación composicional debe ser una concatenación de implementaciones explícitas de las partes constituyentes originales. Si tal argumento es correcto, entonces las representaciones conexionistas distribuidas e implícitas no tendrían los poderes causales para ser sensibles a la estructura.

Este argumento es respondido mediante un contraejemplo concreto, la red RAAM la cual: primero demuestra que no solo la estructura composicional puede ser codificada de forma distribuida, sino que la estructura implícitamente presente dentro de la forma distribuida puede ser usada directamente para el procesamiento más adelante. A pesar de la falta de estructura concatenada explícita en las representaciones de RAAM, estas permiten procesamientos sensibles a la estructura de todas maneras. Y, segundo, se demuestra la posibilidad de operaciones sensibles a la estructura conexionistas que no son implementaciones de algoritmos clásicos. Esto se muestra en que las operaciones sensibles a la estructura llevada a cabo clásicamente tienen como primer paso una descomposición explícita de la representación con implementaciones particulares siendo explícitamente extraídas. En un modelo conexionista la transformación toma lugar sin tener que extraer las partes constituyentes, siendo la operación *directa y holística*.

- *PDP networks can provide models that are not mere implementations of classical theories* (Dawson et al. 1997)

Usualmente, se visualiza a los modelos conexionistas como teorías no cognitivas, dado que se moverían en el nivel de la implementación, siendo esta una opinión bastante común en autores de ciencias cognitivas – de hecho es la opinión de F&P. El modo de que una red pase a ser una teoría cognitiva es

mostrando cómo es posible que tenga estados semánticos y demostrando que tales estados figuran en la explicación de su comportamiento (Dawson et al., 1997, 37-8).

El análisis de una red que se aplica a aprender inferencias lógicas provee un buen ejemplo de cómo se puede defender el carácter cognitivo, no implementacional, de la teoría conexionista. Un análisis de la estructura interna de tal red mostraría que estados particulares de ella – niveles de actividad específicos en las unidades ocultas – son asociadas a interpretaciones semánticas específicas. Luego, se podría mostrar que el conocimiento de tales interpretaciones semánticas sería utilizable para hacer predicciones precisas del comportamiento de la red. Inclusive, que tal conocimiento podría ser entendido como emergencia de reglas, las cuales, aun cuando no fueran recursivas, se extenderían más allá del conjunto de problemas sobre el cual se entrenó la red (op. cit., 38-9)

Así como la pregunta por "¿cuál es la naturaleza de una representación mental?" está claramente mal constreñida, para Dawson et al., la pregunta por "¿qué es una regla clásica?" y "¿puede una red conexionista ser de naturaleza clásica?" también lo están. Un análisis detenido de la estructura interna de una red entrenada para hacer inferencias lógicas permite proveer un marco específico y adecuado para un avance en responder estas preguntas o, al menos, para ir especificando en qué sentido más concreto las reglas clásicas diferirían de las reglas que emergen tras un análisis de la estructura interna de una red entrenada para la solución de tareas cognitivas (op. cit., 39).

A favor del simbolismo:

- *Connectionism, Generalization, and Propositional Attitudes: A Catalogue of Challenging Issues* (Barnden, 1992)

Aquí, más que presentar argumentos en línea de lo que sería imposible en principio – como sucede con lo que se observa en F&P o F&M – se especifican algunos problemas desafiantes para el conexionismo, tales como la generalización explícita o la generación de nuevas combinaciones de creencias. Para ejemplificar el primer problema se plantea una red hipotética que sea capaz de llegar a la conclusión de que "tal persona es atea" a partir de un entrenamiento donde se le enfrente a personas en un pueblo donde cierto grupo de éstas es comunista. Tal red sería capaz de llegar a actuar de manera correcta ante el estímulo de "alguien es comunista" de modo de reaccionar "tal persona es atea", pero esta conducta no implica que el sistema haya construido una generalización, sino que meramente ha logrado conectar de forma empírica estímulos y respuestas sin tener una generalización explícita (op. cit., p. 154).

La necesidad de tal generalización explícita se vincula a que el sistema conexionista necesita entender proposiciones generales como "todos los comunistas en el pueblo son ateos", y que sea capaz de tomar tal proposición, no meramente como exponiendo un hecho útil para futuras inferencias sobre "los comunistas de tal pueblo" – es decir, como permitiendo mejorar su desempeño en la tarea del "saber-cómo" generalizar –, sino que de tomar tal proposición exclusivamente como una creencia perteneciente a quien lo enuncia, sin que tal creencia afecte su desempeño. Esta capacidad de distinguir entre las propias generalizaciones y las ajenas requeriría una generalización explícita en la red (op. cit., 155-6).

Ejemplificando el segundo problema tómesese la proposición "la pera puede hablar". Una pera no puede ser sujeto de la acción "hablar", sin embargo es posible hallar contextos donde una pera "pueda" hablar,

como sucedería en bromas, historias de niños, metáforas, metonimia, contrafácticos, negaciones o contextos de actitud. Como resultado, no es rara en el discurso coloquial la existencia de combinaciones anómalas de ideas (op. cit., 160-1). Supóngase que un sistema conexionista puede entrenarse de modo de hacer inferencias plausibles cuyo contenido sea que, si alguien puede hablar, entonces también puede entender el lenguaje. Esto es, cada proposición “X puede hablar” para diferentes personas llenando X, ha sido emparejada con “X puede entender el lenguaje” en el set de entrenamiento. Existe el peligro que la representación interna de “la pera puede hablar” sea lo suficientemente diferente para que “la pera puede entender el lenguaje” no sea generada. Esta falencia sería causada porque las representaciones de las diferentes personas llenando X activarían el patrón o los microrrasgos de “persona”, los cuales claramente serían un conjunto diferente del que activa “pera”. (op. cit., 164)

Estas limitaciones empíricas del conexionismo no las sufriría un sistema simbólico, ya que en estos existe separación entre la representación de la estructura y la representación de los componentes; combinaciones arbitrarias pueden ser formadas, porque el procesamiento es (o puede ser fácilmente hecho) independiente de la naturaleza de los ítems combinados. Luego, por ejemplo, las técnicas de estructuras de datos convencionales como punteros, asignación secuencial de ubicaciones o de direccionamiento asociativo podrían ser usadas para unir la clase de información que se expone en estos problemas desafiantes. Esto contrasta fuertemente con las redes conexionistas, donde la codificación de asociaciones entre partes de una proposición están unidas inextricablemente con la codificación de lo que es asociado (op. cit., 165).

- *The Connectionism/Classicism battle to win souls* (McLaughlin, 1993a)

Este artículo puede verse en directa oposición con el de Chalmers (1993), ya que responde a la sugerencia hecha por este autor de que los modelos propuestos por Elman y Smolensky están entre “the most prominent counterexamples to F&P’s argument”, además del suyo propio. McLaughlin considera que ninguno de estos modelos – el de Elman, Smolensky o Chalmers – provee una pista sobre qué consiste la posesión de una capacidad para tener un estado intencional o cómo tales estados participan en los procesos cognitivos. Por esta sola razón, ninguno de estos modelos provee una pista de cómo la sistematicidad puede ser explicada por una arquitectura conexionista que no sea la implementación de una clásica (op. cit., 175). McLaughlin se dedica en detalle a revisar cada uno de estos modelos, para finalmente apuntar que lo que los conexionistas necesitan es articular una concepción de la cognición que sea diferente de la concepción clásica, es decir, necesitan decidirse respecto a qué clase de actividad dentro de la red constituye una actividad mental y por qué. Hasta que ellos no hagan tal cosa, entonces no serán capaces de lidiar con el desafío de F&P (op. cit., 179-180).

Lo que vale la pena notar es que McLaughlin considera que la analogía hecha por los conexionistas que dice que el conexionismo es al simbolismo cognitivista como la mecánica cuántica es a la mecánica newtoniana es incorrecta. La relación correcta entre ambos enfoques sería la que se establece entre la mecánica cuántica y la química, donde la mecánica cuántica implementa la química, de lo que se sigue que la primera no muestra que la segunda sea falsa (p.180-182). Sin embargo, quizás lo más conveniente sea que las técnicas del modelamiento conexionista y clásico co-evolucionen de modo de beneficiarse mutuamente. Puede ser el caso que la arquitectura cognitiva incluya la arquitectura clásica, y que la arquitectura clásica sea implementada por una arquitectura conexionista. Puede ser también que los

procesos intencionales sean implementados por procesos simbólicos clásicos, los cuales sean implementados por procesos conexionistas (op. cit., 183-184).

- *Exhibiting versus Explaining Systematicity: A Reply to Hadley and Hayward* (Aizawa, 1997)

Sin duda, no hay desacuerdo en que el encaje entre la teoría y los datos es un elemento esencial en la evaluación de los enfoques clásicos y conexionistas a la cognición. Empero, el ajuste entre teoría y datos no es el punto del desafío de la sistematicidad. Una manera de fundamentar a favor de una teoría científica sobre una rival es mostrando que los datos encajan mejor en ella que en su rival. Otra manera es mostrar que la teoría tiene cierto tipo de superioridad explicativa sobre su rival (Aizawa, 1997, 41).

Aizawa decanta por la segunda clase de argumentación y se basa en el contraste entre la astronomía copernicana y la ptolomaica. Ambas tienen explicaciones sobre el movimiento retrógrado de los planetas, ambos modelos exhiben el fenómeno deseado, pero solo la teoría copernicana tiene una explicación de por qué esto es el caso. En el sistema copernicano el fenómeno es necesitado por la explicación de la teoría sobre la naturaleza del movimiento retrógrado. En el ptolomaico, es producto de la fijación de parámetros, a saber, el supuesto que el deferente del sol es paralelo al epiciclo del planeta y que el período del epiciclo del planeta es igual a un año solar. El problema con la explicación del movimiento retrógrado en el sistema ptolomaico no es mostrar cómo esto es posible, sino mostrar cómo en el sistema copernicano este fenómeno sería necesario. No es suficiente que la astronomía ptolomaica exhiba el fenómeno; debe explicarlo (op. cit., 43).

El problema con los conexionistas es que no aprecian la distinción entre exhibir un fenómeno y genuinamente explicarlo. Cuando se pregunta por una explicación que se supone dan sus modelos, ellos simplemente exponen los mecanismos en el modelo dado. El problema, sin embargo, no es que nosotros necesitemos tener algún mecanismo que subyazca al pensamiento sistemático, lo que se necesita es un mecanismo, una hipótesis, a partir de la cual tal rasgo del pensamiento sea una consecuencia necesaria (op. cit., 49).

(B1) Literatura en torno a la hibridación del simbolismo-conexionismo (y más allá): Aquel grupo de artículos que, dando por sentada la relevancia de los fenómenos discutidos a lo largo del debate, toman las herramientas conceptuales de ambas clases de modelamiento y las combinan en lo que se denominan convencionalmente como modelos cognitivos híbridos⁵⁶. Esta clase de literatura puede subdividirse en (B1.1) *aquella que plantea la necesidad teórica/conceptual de hibridación desarrollándola más o menos* y (B1.2) *aquella otra que se pone “manos a la obra” y estipula posibles modelos que hagan convergen los dos tipos de formalismo*.

Dentro de **(B1.1)** puede ser subdividido entre propuestas filosóficas acotadas versus aquellas más amplias donde la hibridación trasciende los límites de lo impuesto por el debate conexionismo-simbolismo. Un caso acotado puede ser parte de lo expuesto en Bechtel (1988), donde el autor en cierto punto de su desarrollo considera deseable que una manera de concebir los sistemas cognitivos tome ventaja de las fortalezas, tanto del conexionismo, como de los sistemas basados en reglas. Una

⁵⁶ Por ahora restringiré la noción de “modelo híbrido” a aquellos presentados en esta sección (B1). Posteriormente la definición se refinará a lo que presentaré en la sección (2.2.).

posibilidad de integrarlos en una explicación de la arquitectura de la mente única sería considerar a la mente como un sistema modular fodoriano, permitiendo que algunos de los módulos empleen representaciones y reglas y otros implementen una arquitectura tipo PDP. Podemos imaginar que la selección natural haya elegido cuál de estas arquitecturas emplear para una tarea en particular dependiendo sobre cuál está mejor diseñada para ésta. Si esta clase de integración fuera apropiada, cree Bechtel, podríamos esperar evidencia que en algunas tareas exhibiésemos conductas más rígidas y baja tolerancia a los errores como ocurre en el procesamiento basado en reglas, mientras que en otras tareas exhibiéramos una conducta más fluida, aunque inhabilitados para recurrir a patrones recursivos de razonamiento (Bechtel, 1988, 9).

La manera de acomodar las teorías PDP al modelamiento cognitivo más tradicional acepta la imagen de una jerarquía de niveles, pero niega la afirmación que las representaciones son estructuras fijas, así como en los computadores Von Neumann. Además, estas pueden verse como emergentes, estructuras temporalmente estables producidas en un sistema PDP. Estas estructuras, las cuales codificarían las reglas para el procesamiento, podrían igualmente ser vistas como emergentes. Cuando tales estructuras emergen, un tipo de procesamiento basado en reglas debiese ocurrir en el sistema, así como en las tradicionales máquinas de von Neumann. Pero las reglas y representaciones no tendrían la rigidez de una estructura von Neumann (op. cit., 9).

Una forma desde la filosofía de las ciencias de defender esta integración sería apelar a que, históricamente, la investigación sobre la micro-estructura de muchos fenómenos redundaba en cambios para la concepción de la macro-estructura también. Esto es claro con el caso de la genética, donde el descubrimiento del sustrato molecular ha llevado al reconocimiento que la explicación mendeliana equivocadamente mezclaba diferentes nociones de gen (por ejemplo, como la unidad codificadora para un aminoácido, como la unidad de la mutación, la unidad del *crossing over*, etc). Nuevas explicaciones en el nivel mendeliano han distinguido entre las entidades involucradas en estas diferentes operaciones. Lo que las teorías PDP proponen es que sus investigaciones en la micro-estructura de la cognición pueden resultar en cambios en la concepción de la macro-estructura de la cognición también.

Esta *propuesta de hibridación acotada* que subyace en Bechtel⁵⁷, pero también en otros autores, puede verse como siendo subsumida por *propuestas mucho más amplias*, como la de Andy Clark (1998). Lo que inspira a Clark en su propuesta son los agentes autónomos de Brooks, caracterizados por no tener una planificación central, por guiarse en el mundo usándolo como su propio modelo, y navegando de forma fluida y robusta por el mundo. Su punto de partida es que si estos principios de la nueva robótica son

⁵⁷ En Bechtel & Abrahamsen (1990) se trata con un matiz más epistemológicamente amplio cómo la irrupción del conexionismo (entre otras propuestas) estaría posibilitado expandir la “visión proposicional” del conocimiento, no obstante su aparición no debe tomarse como eliminando la visión proposicional del conocimiento, sino que como permitiendo una comprensión más profunda; por consiguiente, aquí se observaría parte de lo que yo denominaría como “filosofía de la hibridación acotada”. Esta propuesta epistemológica, aunque estaría emparentada con la epistemología naturalizada quineana al tratar de hacer compatible epistemología y los descubrimientos en ciencias cognitivas, no dejaría de intentar de ser normativa (Bechtel & Abrahamsen, 1990, 224-5). Las investigaciones posteriores de Bechtel y Abramsen evolucionarán hasta llegar a la propuesta mecanicista, la cual presentará problemas al ser defendida normativamente, puesto que no daría cabida al dinamicismo (de tipo kelsoniano) de manera autónoma (con algo más de detalle se expondrá este tópico en la sección (2.1.))

considerados guía para describir la cognición, entonces hay dos problemas conceptuales inmediatos: el del descubrimiento, o cómo formular hipótesis sobre el posible funcionamiento de estas “mentes fragmentarias”. Y el problema de la coherencia, es decir, cómo los subsistemas independientes llegan a mostrar una conducta organizada, sin tener que apelar a una planificación central (op. cit., 31-3).

Inspirándose en trabajos en robótica de Rodney Brooks, Clark pasa a hablar de una serie de tópicos que tienen que ver con explicaciones alternativas a la cognición. Trata con los experimentos sobre el aprendizaje a caminar de los bebés tratados por Thelen y Smith, experimentos que muestran que el parámetro crucial para mostrar las diferencias conductuales y de aprendizaje entre los bebés tienen que ver con la masa de las piernas, y no con alguna clase de “plan global con factor único” (op. cit., 39-41). Otro enfoque alternativo es el propio conexionismo, comparando el desempeño de DECTalk contra el de NETtalk, ambos programas de producción fonética, el primero basado en “conocimiento estático” versus el segundo que “aprendía” a través de un proceso de entrenamiento a hablar (op. cit., 54-7). La moraleja de esto es que la mente no debe ser vista como un archivador estático o pasivo. Otro caso que toma Clark es el de las termitas y cómo a partir de instrucciones sencillas son capaces de transformar el ambiente y construir complejos termiteros. En síntesis, que la conducta “inteligente” puede aparecer colectivamente extendiéndose más allá de la caja craneana. Luego, su propuesta alternativa cognitiva parte con las características de ser: (i.) Centrada en el mundo real, en tiempo real (ii.) Reconocer de soluciones descentralizadas (iii.) Una perspectiva extendida de la cognición y la computación (op. cit., 81)

Ahora bien, el proceso de diseño natural supera el diseño teórico humano, prescindiendo de nítidas separaciones entre físico, computacional o informacional. La evolución utiliza recursos viejos para producir nuevas habilidades, rara vez existe el diseño ideal, lo que conlleva que el diseño sea opaco. El cerebro tiene un cableado opaco, con cantidades grandes de circuitos recurrentes, los cuales permitirían modificaciones continuas y repetidas, empero el rol del cerebro es nada más que el cuerpo llegue a ejecutar movimientos correctos. Luego, el éxito adaptativo no puede atribuirse solo al cerebro sino que a la unión cerebro-cuerpo situada en un entorno ecológico. Así pues, la explicación que va ganando terreno, según Clark, es la de los sistemas dinámicos que reemplaza la teoría computacional y el vocabulario representacional (Op.cit, 97-8).

La teoría de los sistemas dinámicos es un marco cuyas ideas fundamentales son las nociones de espacio de estados, trayectoria o conjunto de trayectorias posibles a través de tal espacio de estados y la descripción matemática de las funciones que determinan la forma de tales trayectorias. La comprensión basada en espacios de estados, aun cuando es valiosa e informativa, no es totalmente aceptada como un reemplazo de las explicaciones más tradicionales para Clark. Un primer desafío es que un sistema dinámico a medida que aumenta el número de parámetros y el tamaño del espacio de estados se vuelve cada vez más difícil de analizar. Un segundo reto es cuál clase de comprensión es posible obtener del marco de los sistemas dinámicos, ya que puede tender a convertirse más en una descripción abstracta que una explicación completa. Se sabe qué hace el sistema y cuándo lo hace, además la norma temporal de evolución de su conducta, pero parece ser esto una imagen empobrecida. Luego, la postura que se tomará será la ecuménica, donde los instrumentos teóricos de los sistemas dinámicos se verán como complementarios a las interpretaciones computacionales y representacionales (op. cit. 99-102).

Existirían tres clases de explicaciones científicas cognitivas: la basada en componentes – v. gr. cuando explicamos cómo funciona una lavadora apelando a la función de cada una de sus partes –, la explicación tipo “atrapar y lanzar” – donde el mundo lanza entradas al cerebro, que las atrapa y responde lanzando acciones, lo cual implica una explicación interactiva – y, por último, la basada en la emergencia. Esta última apela a la existencia de *variables colectivas* que generan cierta auto-organización. Estas variables deben ser distinguida entre aquellas controladas (que reflejan conductas del sistema manipulables de manera simple y directa) y variables incontroladas (que reflejan conductas del sistema surgidas de la interacción de múltiples parámetros de modo que se resisten a la manipulación directa) (op. cit., 103-8).

La explicación dinámica difiere de la tradicional basada en componentes, siendo su diferencia más evidente que los sistemas dinámicos explican conductas describiendo otras. Ergo, intuitivamente, esto distaría mucho de ser una explicación, ya que ésta comúnmente se basa en explicitar los mecanismos ocultos que provocan alguna conducta. Además, está la creencia que el sistema físico del cerebro tiene principios organizativos especiales, esencialmente distintos de los que se usan para explicar, por ejemplo, la sincronización de dos péndulos. No obstante, para un teórico de los sistemas dinámicos, tanto dinámica neuronal como corporal derivarían de los mismos principios que subyacen en la auto-organización de sistemas complejos. Tales principios no serían meramente descriptivos, sino que son explicativos en tanto que permiten dilucidar contrafácticos del sistema, aunque careciendo detalles sobre los dispositivos que describen y explican. En esto contrastan, por ejemplo, con los modelos computacionales donde se descomponen tareas complejas en más simples, correspondiendo a cada función cierto componente material posible. Aun así, las explicaciones dinámicas, gracias a las variables colectivas y la capacidad de dar cuenta de acoplamiento de conductas, permite naturalmente abarcar múltiples elementos en interacción, tal como un sistema tipo agente-ambiente (op. cit. 115-8).

Para hacer que la postura ecuménica cuaje deben conjugarse las nociones de mentes corporeizadas y situadas como, además, la de cerebros que computan y representan; ¿es posible que computación y representación tengan definiciones no triviales en este nuevo marco ecuménico? La representación interna, como noción, ha sido terreno común tanto para IA clásica como conexionista, sin embargo la capacidad de desacoplamiento que requiere la versión más tradicional de representación interna es demasiado exagerada para Clark. Considérese el caso de las neuronas que contienen información sobre la dirección hacia la que se orienta la cabeza de un animal, por ejemplo, una rata. Hasta ahora no existe descripción que muestre que tales neuronas sean capaces de desempeñar su función sin la presencia de una continua corriente de señales propioceptivas cuyo origen es el cuerpo de la rata. El desacoplamiento no es posible, sin embargo tal complejo de neuronas que contiene información es ciertamente un esquema de codificación sistemática (op.cit, 143-5). Estos sucesos internos pueden ser interpretados en una ciencia cognitiva madura de diversas formas posibles (op. cit., 148):

1. Pueden interpretarse como carentes de importancia explicativa, aunque con cierto papel heurístico.
2. Pueden interpretarse como induciendo a error, es decir, el vincular procesos internos concretos con funciones de portar contenido acarrearía conclusiones teóricas erróneas.
3. Las interpretaciones forman parte de la explicación, reflejando verdades importantes sobre las funciones de diferentes estados y procesos.

El ecumenismo de Clark funciona argumentando contra el radicalismo (postura 2), que tiende a aumentar la competición cuando el progreso se daría con la cooperación; la mayoría de explicaciones alternativas que ponen énfasis en el cuerpo y el mundo podrían ser vistas como complementarias a aquellas computacionales y representacionales. Un primer argumento es que la representación no necesariamente debe ser pensada clásicamente, es decir, por ejemplo, que la representación de “taza” tenga que siempre invocar contenidos con propiedades profundas e independientes del agente, como forma y capacidad. La representación puede ser local y orientada hacia la acción, como por ejemplo, cuando buscamos cierta taza, y acudimos a propiedades superficiales como color o brillo, propiedades que para los clásicos son insuficientes para individuar el concepto “taza” (op. cit., 149-150).

Un segundo argumento, es que si uno analiza el desarrollo de enfoques de postura radical con respecto a la representación y la computación, lo que se encuentra es que la evidencia de estos enfoques alternativos apunta a que las explicaciones computacionales y representacionales son incorrectas por su insuficiencia respecto del cuerpo y el ambiente, pero no necesariamente siendo la representación y la computación en general los “villanos de la película”. Clark, tomando el caso de Thelen y Smith (1996), considera que son compatibles las afirmaciones:

- i. Que el desarrollo no está dirigido por un plan totalmente desarrollado de antemano. Luego, el desarrollo (y la acción) presentan un orden meramente ejecutor, siendo las soluciones el resultado de un “montaje blando” de múltiples componentes heterogéneos.
- ii. Que la cognición adulta no implica operaciones lógicas internas sobre estructuras de datos proposicionales. Luego, aun cuando la cognición del adulto parezca lógica y proposicional, ésta se basa en recursos que se desenvuelven activamente en tiempo real cuya base es la experiencia corporal.

Con la existencia de soluciones *parcialmente programadas*. Estas soluciones se traducen en que habría un programa inicial en el ser humano configurado por la evolución, de modo de permitir que la dinámica corporal y las contingencias del entorno próximo permitan determinar, tanto el curso, como el resultado del proceso de desarrollo (Op.cit, pp. 153-6).

La postura de Clark se puede resumir en que la resolución del debate sobre la necesidad de representaciones internas que subyazcan al éxito conductual depende de futuros estudios empíricos. Es obvio que existen procesos cognitivos que serían inútilmente descritos con un excesivo énfasis en las representaciones internas, sin embargo es discutible aun cuántas representaciones internas máximas requiere la explicación de una función cognitiva (op. cit., 169). *Para Clark una explicación concisa, general y clara de buena parte de nuestro éxito adaptativo dependerá claramente de, al menos, reinventar en algún sentido las ideas de procesamiento complejo y estados internos*. Él considera que a medida que se desarrollen las investigaciones se observará una co-evolución cooperativa entre las múltiples ideas y análisis de los distintos enfoques (op. cit., 174-5).

Dentro del segundo subgrupo de literatura **(B1.2)** es posible hacer más subdivisiones de acuerdo si se sigue una filosofía de hibridación “acotada” o “expandida”. Ejemplificaré con algunos casos para que sea más fácil visualizar el grado de mezcla entre diferentes enfoques:

- Hibridación “intra-conexionismo”; SAARCS

Existen tres arquitecturas que pueden ser consideradas “conexionistas” en un muy amplio sentido: redes distribuidas, redes localistas y redes *marker-passing*, esta última la más conexionista de las arquitecturas simbólicas. Cada una de estas redes conexionistas en diferente grado contienen ventajas y desventajas, tales como (Lange, 1992, 237-8):

- Las distribuidas tienen las ventajas de ser resistentes al ruido, robustas y capaces de generalizar. Sus desventajas son dificultad con el enlazado dinámico de variables y la representación de la estructura necesaria para manejar relaciones conceptuales complejas. Luego, no están bien adaptadas para modelar tareas cognitivas de alto nivel como planeamiento o comprensión lingüística.
- Las ventajas de una red localista se ejemplifican en las redes estructurales entre conceptos, las cuales solucionan fácilmente tareas cognitivas como desambiguación del sentido de las palabras. Lamentablemente, estas redes carecen de un poderoso aprendizaje y capacidades de generalización como ocurre con las redes distribuidas. Por otro lado, también tienen problemas con el enlazado dinámico de variables – esta última siendo especialmente bien lograda con los modelos simbólicos.
- Las redes *marker-passing* tienen cierto parecido con las redes localistas, en tanto que representan el conocimiento en redes semánticas y retienen un grado de paralelismo al nivel del conocimiento. No obstante, en vez de propagarse valores de activación numéricos, esta clase de redes propagan marcadores simbólicos, así permitiendo el enlazado de variables necesario para la aplicación de reglas. Su desventaja reside en que las unidades de esta clase simbólica de red son más complejas que las distribuidas y localistas, luego no posee la capacidad de aprendizaje de una red distribuida y no exhiben las capacidades de satisfacción de restricciones de una red localista (op. cit., 237-8).

Esto permite justificar a Lange que a veces se vuelve necesario construir modelos híbridos usando elementos de más de un nivel de modelamiento conexionista, entendiéndose por “conexionista” cualquiera de las tres clases de redes antes mencionadas. A menudo no es posible construir un modelo exitoso para una tarea cognitiva dada con los elementos de un solo nivel conexionista, enfrentando a uno esto a una disyuntiva: o bien abandonar o disminuir la escala de la tarea, de manera de que las habilidades del nivel conexionista puedan lidiar con el modelamiento de la tarea, o bien utilizar elementos de otro nivel, generándose así un modelo híbrido. Es obvio que abandonar o disminuir la escala de la tarea no es una salida muy deseable, por consiguiente la hibridación parece la mejor solución (op. cit., 250).

SAARCS es un modelo híbrido entre una red localista y otra de *marker-passing* que integra comprensión lingüística y recuperación analógica. Dada una representación sintáctica de una oración de entrada como pista, la red de SAARCS primero desambigua e infiere una interpretación de tal pista, luego recupera y retorna una oración o un episodio de la memoria de largo plazo que sea la más análogamente cercana a tal interpretación. SAARCS consiste en una red localista que codifica una base de conocimiento de conceptos (por ejemplo objetos, acciones, planes y metas) y reglas generales de conocimiento para inferir entre conceptos. También indexado en esta red semántica están las unidades representando episodios de la memoria de largo plazo que son objetivos potenciales de recuperación. Usando esta red, la comprensión y el proceso de recuperación analógico consiste en cuatro etapas (op. cit., 270):

1. La activación se propaga a través de la red semántica para desambiguar e inferir una interpretación de la pista que se le da al sistema.
2. Los marcadores simbólicos se propagan desde las unidades del camino de inferencia ganador para encontrar los objetivos semánticamente similares al contexto actual de la interpretación de la pista.
3. Una red de unidades es construida dinámicamente para representar los mapeos posibles en competencia entre la interpretación de la pista y los objetivos semánticamente similares encontrados por la propagación de los marcadores. Las conexiones de excitación e inhibición entre unidades de esta nueva red de mapeos refuerza la consistencia semántica y estructural con la pista.
4. Una red nueva de mapeos se asienta mediante un proceso de satisfacción de restricciones; el mapeo de unidades activas después del asentamiento constituye el resultado más coherente para la pista.

Mediante la integración de una red localista conexionista con *marker-passing* que permiten la creación dinámica de redes de mapeo, SAARCS es capaz de combinar partes de comprensión lingüística y procesamiento de recuperación analógica en una red sin una supervisión externa. Así, SAARCS sería un primer paso para un modelo que explique la influencia de la comprensión lingüística en la recuperación de memoria y viceversa. Contrario a los modelos que se concentran primariamente en la recuperación de memoria – como los modelos de razonamiento basados en casos – y los modelos conexionistas de recuperación de memoria, SAARCS puede, potencialmente, dar cuenta de muchos fenómenos psicológicos que involucran *priming* (recuperación aumentada de cierto estímulo que no ha sido almacenado en la memoria explícita, sino que quedaría en la memoria implícita) y los efectos del lenguaje en la recuperación de la memoria humana, como, por ejemplo, la recuperación incrementada debido a la repetición, todos los cuales pueden ser modelados por variaciones en los niveles de activación de evidencia antes de la presentación de la pista (op. cit., 277).

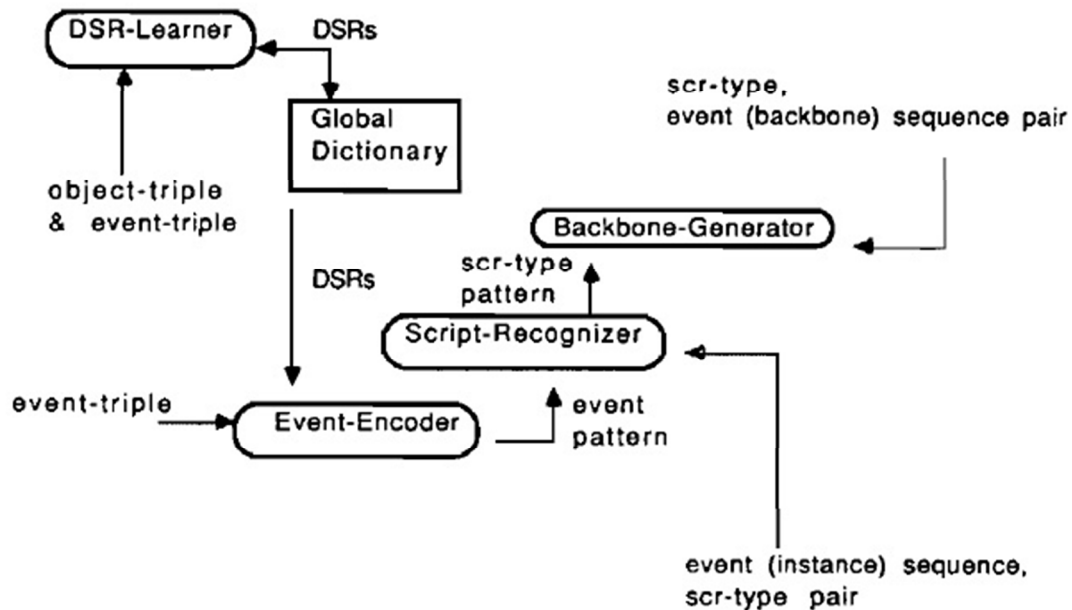
- Hibridación “simbolismo-conexionismo”; DYNASTY

La mayor parte de los modelos híbridos son modelos que combinan redes conexionistas con procesamiento simbólico tradicional. Esta clase mayoritaria de sistemas híbridos son especialmente valiosos para los enfoques funcionales del diseño de modelos, donde un sistema es diseñado como un conjunto de módulos funcionales interconectados que son primero implementados simbólicamente, pero que son gradualmente reemplazados con módulos conexionistas. Un ejemplo de este enfoque es DYNASTY, un sistema que comprende historias mediante guiones y que usa múltiples módulos de redes distribuidas recurrentes que acceden a un diccionario simbólico.

En este modelo cada red distribuida es entrenada para servir como módulo desempeñando una subtarea de procesamiento separada, tales como análisis sintáctico secuencial del texto de entrada en representaciones de eventos individuales, reconocimiento que una secuencia de eventos encaja en un guion particular y parafraseando el guion reconocido. Muchos de estos módulos distribuidos eran implementados al principio simbólicamente para permitir el testeo sobre todo el modelo del concepto, y fueron reemplazados uno a uno cuando el tiempo y la oportunidad lo presentaban. El componente simbólico restante de DYNASTY es una tabla *hash* simbólica, usada como “diccionario global” para almacenar las representaciones de los conceptos y los eventos que ha aprendido junto a los símbolos

que los representan. Diferentes módulos de redes distribuidas acceden y almacenan valores en este diccionario simbólico global cuando es requerido (op. cit., 253-4).

FIGURA 6 (Lee et al., 1990, 325); Esta figura muestra la arquitectura del sistema en la fase de entrenamiento. Los óvalos representan módulos PDP, mientras que la caja del diccionario global es una memoria simbólica. Las líneas designan flujos de datos uni-/bi-direccionales. Cada módulo es entrenado separadamente con su propio conjunto de datos de entrenamiento.



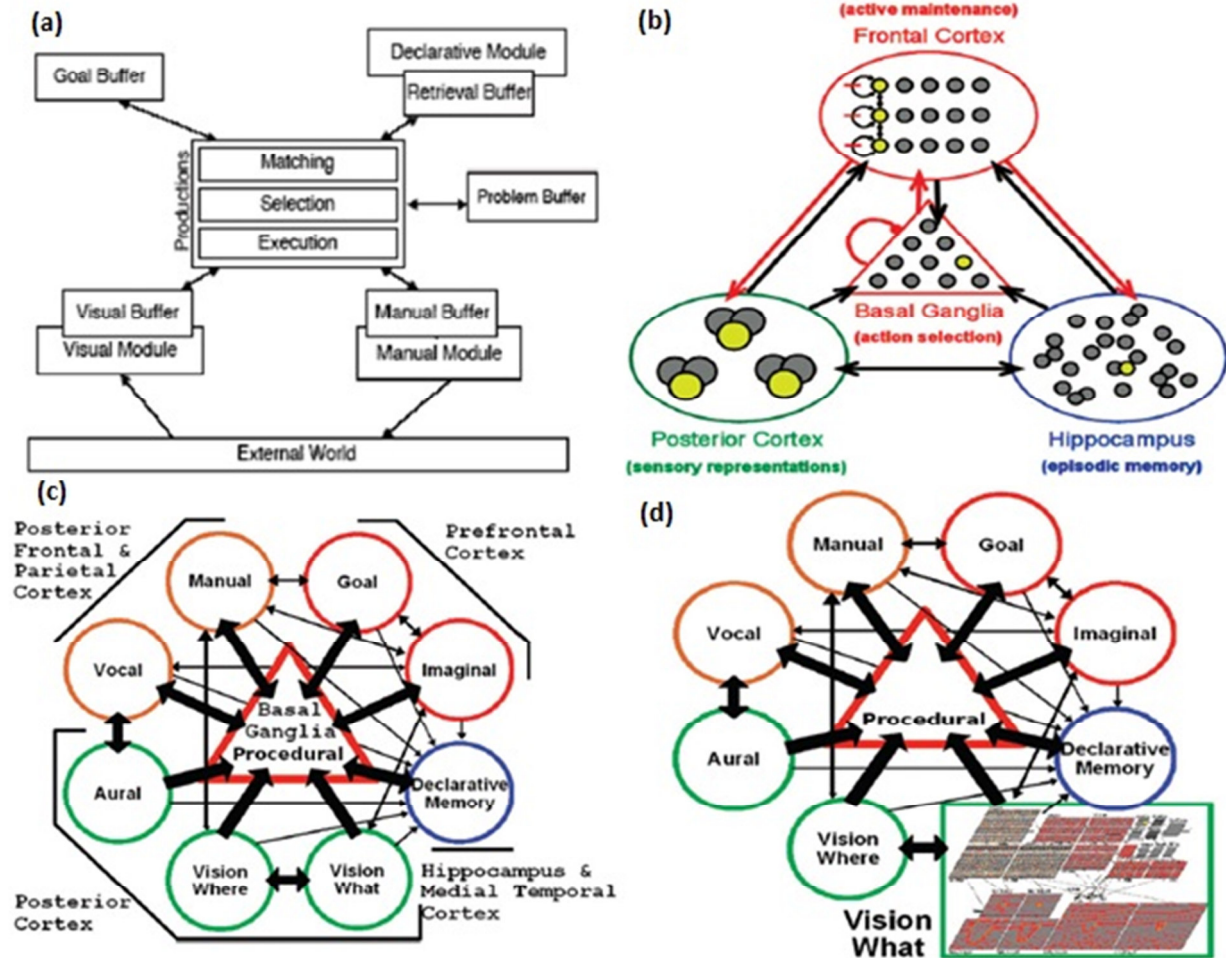
- Hibridación “meta simbolismo-conexionismo”; SAL

SAL es una arquitectura pluralista que nace del trabajo colaborativo del grupo que ha desarrollado ACT-R y el grupo que ha desarrollado Leabra. ACT-R es una continuación de HAM (Human Associative Memory), el cual representaba el conocimiento declarativo en forma de red proposicional. Por otra parte, Leabra nace del deseo de reproducir resultados de laboratorio de mecanismos neuronales, por tanto está compuesto de simulaciones de potenciales de membrana con umbrales no lineales a otras neuronas. Cuando los dos grupos empezaron a trabajar en conjunto se percataron de algo importantísimo: las dos arquitecturas, a pesar de sus orígenes virtualmente opuestos y ampliamente diferentes en niveles de abstracción, tenían una visión de la arquitectura total del cerebro muy similar (Jilk et al., 2008a, 205).

Aún más, los investigadores descubrieron que los mecanismos subsimbólicos que subyacían en ACT-R tenían similitud conceptual e inclusive matemática al comportamiento de las representaciones emergentes en Leabra. Se reconoció así que cada arquitectura reflejaba un compromiso explícito con el *pluralismo*, ambos verticalmente (grueso versus granular) y horizontalmente (v.gr. módulos reemplazables en ACT-R, múltiples estrategias y mecanismos de aprendizaje en Leabra) (Op.cit).

Al nivel de sistemas a larga escala, el acuerdo teórico es evidente cuando se traslapan las organizaciones de las arquitecturas (véase la letra (c) en la **FIGURA 7** en la siguiente página). Ambas dan un rol central al ganglio basal en recibir entradas convergentes desde un amplio rango de áreas de procesamiento cortical, las cuales conducen el desempeño de acciones motoras específicas o cognitivas.

Anatómicamente el ganglio basal envía primariamente su output a la corteza frontal, la cual es asociada con el mantenimiento activo de la información de la tarea relevante en Leabra, y con los *buffers* homólogos de ACT-R. Ambas arquitecturas subrayan la importancia de los sistemas de memorias declarativas/episódicas, cuyo sustrato sería el hipocampo y estructuras anatómicas relacionadas. Finalmente, ambas adoptan vías especializadas para el procesamiento sensorial y motor, caracterizadas por el córtex posterior (Op.cit, 206).



A nivel cognitivo, una división en componentes procedimentales y declarativos es compartida por ambos marcos. Esta distinción y disociación tiene clara validez cognitiva; las personas pueden tener conocimiento declarativo abstracto de cómo hacer algo, y aun así ser incapaz procedimentalmente de hacerlo, y viceversa. El aprendizaje por refuerzo es un elemento conductual central en la cognición procedimental de animales y humanos. El sistema de ganglio basal en Leabra está fuertemente modulado por un modelo de la dopamina, cuyas señales envían información de recompensa y castigo. Las señales de recompensa son asociadas a acciones procedimentales, mientras que señales de retroalimentación negativas reducen la tendencia de producir acciones asociadas. Un tipo de aprendizaje por refuerzo más abstracto existe en el sistema procedimental de ACT-R, donde la historia integrada de los éxitos del pasado y el tiempo consumido son las determinantes más fuertes para seleccionar qué regla productiva se gatillará (Op.cit, 206-7).

Como todas las teorías, ACT-R y Leabra están incompletas. Mientras ACT-R utiliza mecanismos subsimbólicos y puede interactuar con módulos que no son simbólicos, en última instancia entradas y representaciones se traducen a símbolos puros. Leabra tiene una arquitectura tripartita a larga escala, pero aún no había sido implementada en una simulación integrada de larga escala. *Ninguna de estas arquitecturas provee mucho detalle sobre la cuestión central de cómo representaciones simbólicas, el rasgo humano que nos hace cognitivamente únicos entre los animales, aparece orgánicamente en mente y cerebro.* SAL es una síntesis de la teoría de Leabra sobre el funcionamiento neuronal, la conducta de las redes y la arquitectura tripartita, con la teoría de ACT-R sobre la toma de decisiones simbólica y subsimbólica, la organización y activación representacional, y una arquitectura con organización modular. Además, SAL está basado en datos de psicología humana y su biología (Op.cit, 208).

En un esfuerzo inicial, el equipo de SAL construyó una demostración representando una síntesis preliminar de las dos arquitecturas. El modelo se desempeñó correctamente en navegación y tarea de búsqueda (letra (d) en la anterior **FIGURA 7** página anterior). La integración aquí es muy simple, donde el módulo visual de un modelo existente de ACT-R de navegación es reemplazado con un modelo de reconocimiento visual de objetos de Leabra, el cual es capaz de procesar imágenes tipo mapas de bits sin procesamiento previo en una manera que el módulo visual de ACT-R no es capaz. Igualmente, los modelos existentes de Leabra no son capaces de organizar una conducta de resolución de problema dentro de un período de pocos minutos, así como el modelo de ACT-R es capaz en la búsqueda de un objeto en un ambiente complejo. Luego, este modelo híbrido de SAL representa un nuevo nivel de funcionalidad que va más allá de las capacidades de sus arquitecturas constituyentes (Op.cit, 208-9).

Esta sencilla integración modular ilustra los significativos beneficios conductuales que se presentan a partir de una síntesis. *El valor obvio es que el modelo SAL puede completar una tarea que ninguna de las arquitecturas podría desempeñar aisladamente:* ACT-R no podría hacerlo porque su módulo visual no puede reconocer objetos a partir de mapas de bits, y Leabra porque no tendría propiedades demostradas de control que le permitiesen navegar en ambientes espaciales complejos. *Más interesante todavía, es que este modelo arroja luces sobre un enigma teórico básico, a saber, el problema del enraizamiento de símbolo – symbol grounding.* El modelo SAL demuestra cómo la arquitectura podría entrenar la red visual de Leabra para asociar imágenes de un objeto a su representación simbólica, de modo de usar esta capacidad para, robustamente, reconocer objetos en el ambiente, extraer su identidad simbólica, y usar tal información para controlar conductas complejas (Op.cit, 209-2010).

(B2) Literatura crítica; aquella que juzga la validez de los conceptos usados dentro de la polémica simbolismo-conexionismo y que a partir de, o bien el análisis conceptual, o bien de la recopilación de experimentación empírica, o bien de una combinación de ambas, llega a concluir el grado de verosimilitud del o los fenómenos sobre los cuales se está discutiendo y, por tanto, es útil para llegar a juzgar la validez del debate en su conjunto.

Dentro de este subtipo que se mantiene aún en los márgenes del debate – aun cuando muchas veces terminan traspasando tales límites – resulta especialmente interesante el caso de la “sistematicidad del lenguaje y el pensamiento”, que es tomada por autores posteriores como fenómeno fundamental que requiere, básicamente, una buena justificación de la pertinencia de una explicación.

Por ejemplo, en Johnson (2004) se considera la sistematicidad del lenguaje como algo problemático, donde es lícito preguntarse si es siquiera importante como para teorizar sobre ésta. La sintaxis recursiva (y la semántica) y un lexicón altamente estructurado son componentes bien conocidos y bien estudiados del lenguaje humano, mientras que la sistematicidad es en el mejor de los casos un subproducto del primero, y en el peor poco más que un “fuego fatuo”. (Johnson, 2004, 129-130)

Por otra parte, el mismo autor sobre la sistematicidad del pensamiento también será lapidario, diciendo que hay, al menos, dos razones generales para dudar que un sistema empírico dado sea sistemático. Primero, sea lo que sea precisamente la sistematicidad, parece ser más bien rara en sistemas naturales. Por ejemplo, el hecho que una especie de un género pueda interactuar de una manera específica con un rasgo del ambiente, no implica que todas las especies de tal género puedan hacerlo también – siendo esta una alusión directa a la suposición de F&P que apuesta por la sistematicidad en organismos infrahumanos. Cummins asevera que el componente sistemático parece estar restringido a la psicología humana, desafortunadamente no da un ejemplo claro de esto, recurriendo solamente a considerar la percepción de dos objetos en el espacio: cualquiera que pueda ver (o imaginar) una escena que involucra un objeto O1 y otro O2, puede ver (o imaginar) una escena donde sus localizaciones están intercambiadas. Pero, tal ejemplo es endeble, ya que sin especificar las clases de objetos psicológicos relevantes y sus posibles localizaciones, tal ejemplo general es evidentemente falso. Parece natural interpretar “localización” como relativa a las características del otro objeto y del contexto, luego uno puede imaginar una esfera pequeña transparente dentro de una caja negra más grande, pero no se puede sencillamente hacer el intercambio de localizaciones. De hecho el único sistema sistemático que es capaz de cumplir con lo que el ejemplo de Cummins trata de mostrar son los lenguajes de la lógica y las matemáticas, pero estos sistemas no aparecen naturalmente, al contrario, son específicamente diseñados para tener propiedades como la sistematicidad, justamente de modo de proveer un medio más exacto para razonar que el modo proveído por el lenguaje natural. Que estos lenguajes creados por diseño “posean sistematicidad” no es evidencia de la sistematicidad del pensamiento (op. cit., 131-2).

La segunda razón para demandar evidencia sobre la afirmación de que un sistema dado es sistemático, es que tal propiedad es una propiedad matemática inusualmente fuerte. Las nociones estándar de “sistematicidad” pueden ser capturadas del siguiente modo:

- Donde C sea alguna colección de conjuntos de objetos $\{C_1, \dots, C_n, \dots\}$ y F sea un conjunto de funciones, una colección L es sistemática respecto a C y F si para cualquier $f \in F$, cualquier k -tupla c_1, \dots, c_k de elementos desde $C_{i_1} \times \dots \times C_{i_k}$ (con cada $C_{ij} \in C$), y cualquier $x \in C_{ij}$ ($1 \leq j \leq k$), $f(c_1, \dots, c_k) \in L$ si y sólo si $f(c_1, \dots, c_{i-1}, x, c_{i+1}, \dots, c_k) \in L$

El lenguaje se pensaba que era sistemático, porque se pensaba que habían algunas categorías C_1, \dots, C_n y algunas operaciones sintácticas F que, si el resultado de aplicar una de estas funciones a la k -tupla de expresiones pertenecientes a las categorías genera una oración gramatical (esto es, un elemento de L), también lo hará el resultado de aplicar la misma función a cualquier k -tupla que difiera del original en la mayoría en que sustituye un elemento por otro de la misma categoría. Pero, nótese cuán fácil es para los sistemas de varios tipos fallar.

Otros casos.

No solamente se ha juzgado sobre las características de la cognición que tratan de ser capturadas explicativamente por los rasgos arquitecturales clásicos, sino que también sobre otros tópicos asociados tales como la posibilidad de diferenciar entre asociaciones y reglas (Sloman, 1996; Gigerenzer & Resier, 1996). Existe literatura que sin duda se encontrarían en un límite de la categoría (B2), ya casi en la categoría (A1), como podría ser el caso de obras extensas como Gomila (2012), donde se dice que existiría dualidad en la mente humana, por una parte existiendo una cognición independiente del lenguaje y, por otro lado, una cognición de alto nivel que sería constituida desde el lenguaje. De haber algo así como “sistematicidad”, entonces ésta derivaría de la utilización del propio lenguaje natural y no desde dentro del pensamiento. En general, a pesar que algunas cuestiones desde el cognitivismo quedan, la tesis dual de Gomila pasaría a estar mejor adecuada a la línea de pensamiento de las corrientes alternativas de las ciencias cognitivas.

1.3. Desde las reflexiones en torno a la polémica a la síntesis de un punto de partida teórico

Creo que la actitud más natural ante la apabullante cantidad de literatura sería simplemente suspender la investigación centrada en la relectura y re-organización de los contenidos, dado que, lo que más pareciera importar hoy, son los detalles más que la búsqueda de algún gran relato que re-unifique los tópicos, relato cuya utilidad *a priori* sería bastante dudosa. Sin embargo, he llegado a la convicción – quizás aún no lo suficientemente bien justificada, quedando ello al juicio del especialista – que la manera de reorganizar la polémica, no solo es posible construirla conceptualmente, sino que podría ser de ayuda para el propio campo de las ciencias cognitivas, en tanto vuelve a encauzar las preguntas donde corresponden, y no permitiendo a la larga malgasto de energía en investigaciones cuya forma más abstracta ya estaría bastante bien definida. Empero, esta reorganización no vendría dada por un nuevo relato cognitivo o un enfoque alternativo “mejorado” como los expuestos en (1.2) (A0), sino que se llevaría a cabo mediante las herramientas conceptuales de la filosofía de las ciencias.

¿Cómo es que he llegado a tal convicción? A través de la lectura de los mismos artículos que componen la polémica, de algunas de sus consecuencias y de observar cómo los protagonistas esenciales tienen una visión restringida del campo de las ciencias cognitivas así como de su historia⁵⁸. Esta visión restringida tiene que ver con una postura tanto de las ciencias en general, como de los supuestos ontológicos de la cognición que terminan por restringir la gama de formalismos posibles con los cuales tratar lo mental. Sin embargo, por mor de la transparencia, antes de saltar directamente al punto de partida en la sección (1.3.4.) que explicitará cómo desde, exclusivamente, un único concepto proveniente la filosofía de las ciencias es posible reorganizar y re-comprender la relación interteórica simbolismo y conexionismo, explicitaré parte del conjunto de reflexiones que me sirvieron, tanto de motivación, como de guía teórica para desarrollar el resto de la tesis.

⁵⁸ La historia de las ciencias cognitivas y de la psicología es muchísimo más amplia que la oposición entre el paso de un conductismo hegemónico a un cognitivismo transgresor, ya que, por ejemplo, paralelamente al conductismo existían otras escuelas de psicología que hacían lo suyo (Bechtel et al., 1998, 17-20) Por otro lado, la propia historia del simbolismo y el conexionismo es muchísimo más compleja que la que declaran, usualmente, sus defensores.

1.3.1. Reflexiones y cuestionamientos en torno a la polémica simbolismo-conexionismo y su evolución.

No son pocas las dudas filosófico-conceptuales que pueden surgir a lo largo del estudio de la polémica y de sus consecuencias. Por mor de la claridad teórica detallaré buena parte de aquellas preguntas y cuestionamientos a lo largo de todos los momentos de la polémica expuesta y posterior evolución, siendo éstas las cuales me guiarán a reorganizar, en parte, la polémica (sección (2.5.vi.)):

(.01) Dentro de la primera parte del volumen 1 del grupo PDP, en su exposición general, el conexionismo es presentado como una “metateoría”, desde la cual se generarían modelos concretos sobre fenómenos más específicos. Sin embargo, cuando se traslapa tal noción con la de idea general de F&P, a saber, que conexionismo y modelos clásicos serían competidores en la postulación “arquitecturas cognitivas” (Fodor & Pylyshyn, 1988, 4), sería legítimo preguntar, ¿es posible establecer una relación de *competencia* entre el conexionismo, en términos de “metateoría”, y la noción de “modelos clásicos” que presenta F&P? Si es así, entonces ¿de qué modo están compitiendo? De lo contrario, ¿por qué no puede decirse que están compitiendo?

(.02) Resulta natural suponer que existen diferentes grados de generalidad en las comparaciones entre el simbolismo y el conexionismo⁵⁹, es decir, que cuando se está comparando la metateoría conexionista se tiene en mente una versión abstracta de los modelos simbólicos, utilizándose la misma estrategia en el caso de la crítica de los cognitivistas. Ahora bien, más allá de las comparaciones generales, en el caso de las comparaciones específicas se parece presuponer la pertinencia de tener que dar cuenta de conceptos – v. gr. los esquemas – cuyo origen tiene relación con el uso de la estrategia *top-down*. No obstante, tal hecho ¿no parece ser una *validación implícita de que la metateoría conexionista es implementacionista* en su esencia? Si no fuera el caso que es así, entonces ¿no sería una muestra que, como mínimo, el grupo PDP es implementacionista es su versión del conexionismo?

(.03) En línea con (.02), tratando de ir aún más profundo respecto a los niveles de análisis y las estrategias opuestas bottom-up versus top-down, cabe preguntarse, en caso que pueda darse por legítima la separabilidad del conexionismo de sus posibles relaciones con el simbolismo, ¿acaso el conexionismo debe ser competencia de otra estrategia top-down⁶⁰ o es posible plantear la existencia de *otras metateorías bottom-up* que puedan entrar a disputar el espacio que ocupa el conexionismo frente al simbolismo?

(.04) Ya entrando en algunos detalles sobre los elementos del marco general conexionista, uno puede cuestionarse, ¿son acaso las “neuronas artificiales” el único modo de conectar el nivel de organización fisiológico neuronal con los procesos cognitivos que darían lugar a la conducta inteligente? Dado que, obviamente, no es el único modo de abstraer los detalles neuronales de las neuronas reales (ya que los propios investigadores del grupo PDP reconocen que no toman en cuenta el detalle preciso de los *spikes* ni de la neuromodulación) entonces, la primera cuestión que surge es, *¿cómo de estas abstracciones*

⁵⁹ Para corroborar véase la parte de contrastes entre simbolismo y conexionismo en la sección (1.1.).

⁶⁰ Para reafirmar este punto puede complementarse la postura PDP con los dichos de Smolensky quien asevera que “subsymbolic models should not be viewed as neural models. If the subsymbolic proves valid, the best subsymbolic models of a cognitive process should one day be shown to be some higher-level approximation to the neural system supporting that process.” (Smolensky, 1987b, 9).

neuronales podrían surgir teorías o “metateorías” cognitivas alternativas al conexionismo? y una segunda que cabría responder es ¿en qué son fundamentalmente diferentes con las ANN?

(.05) Existen diferentes arquitecturas de ANN, así como también diferentes reglas de aprendizaje y, tomado en cuenta que el conexionismo es susceptible de variadas aplicaciones ¿cómo asegurarse de que son *unívocamente aplicables a la explicación-predicción cognitiva*? En caso de que se pueda asegurar tal aplicabilidad unívoca ¿qué argumento existe para ello? En caso contrario ¿cómo diferenciar las diferentes maneras de usar las ANN?

(.06) La relación entre el ambiente y una red PDP es considerada esencial para distinguir el enfoque “*tabula rasa*” versus el innatismo, estando el propio grupo PDP en un punto medio de los extremos de la discusión. Respecto a esto, caben contestar un par de cuestiones: primero, ¿de qué modo estas discusiones son atingentes como argumentación a favor o en contra del conexionismo o los modelos clásicos? Y, segundo, en el caso que no sean atingente, ¿pueden ser separadas estas discusiones de la investigación científica justificadamente *sin pérdida substantiva de legitimidad* de la clase de *argumentación apriorística* usada normalmente en los círculos de filosofía?

(.07) Tanto los conexionistas – el grupo PDP – como los defensores del simbolismo, particularmente Fodor y Pylyshyn, se basan en intuiciones psicológicas o cognitivas, que conllevan posteriormente generalizaciones, más o menos fundamentadas en observaciones de la conducta humana. Sin embargo, a partir de las mismas observaciones, no solamente llegan a teorías contrapuestas, sino que, aunque se compartan conceptualmente los mismos constructos teóricos constituyentes de sus teorías, estos ocupan diferentes espacios dentro de sus arquitecturas. Por ejemplo, el contenido representacional, mientras en las arquitecturas clásicas está estructurado de manera de constituir un sistema simbólico, por tanto pasa a ser parte del *explanans* de la teoría computacional de la mente, en el caso del conexionismo el mismo contenido representacional es parte del *explanandum*. El primer fenómeno, la formulación de teorías contrapuestas entre sí sobre un mismo corpus de datos tiene ya bastante de conocido en la filosofía de las ciencias⁶¹, no obstante, ¿cómo explicar que, por ejemplo, compartiendo el realismo intencional, conexionismo y simbolismo pongan en diferentes lugares teóricos el “mismo” concepto?

(.08) Precisamente, en continuidad con la última pregunta (.07), la relación entre conexionismo y simbolismo no parece ser sencillamente la de “teorías contrapuestas”, ya que habría una clara parentela entre ellas al compartir la tradición del realismo intencional. Otro ejemplo de parentesco ideológico existiría, según Chemero y Silberstein, en que simbolismo y conexionismo tienen al “individualismo metodológico” como elemento filosófico compartido (Chemero & Silberstein, 2008, 2-4). Dado que esos componentes no pueden ser ignorados meramente por la facilidad que conllevaría quedarse con una imagen simplificada de la relación simbolismo-conexionismo, entonces, ¿cómo explicar su relación completa sin caer en la imagen simplificada?

(.09) La necesidad de utilizar unidades subcognitivas, cuya interacción permitiría la emergencia de los procesos cognitivos en sus modelos, es la *directriz de generación crucial* de los modelos PDP. En

⁶¹Este fenómeno se denomina *subdeterminación de las teorías científicas* y es descrito por Quine (Gibson, 2004, 15)

contraste, los rasgos arquitecturales [1] y [2] de F&P serían también directrices de generación de aquellos modelos de tradición representacional para ser considerados como “sistemas de símbolos”. Ahora bien, estos elementos pertenecientes, respectivamente, al conexionismo y al simbolismo ¿de qué modo participan del concepto metateórico del conexionismo y el simbolismo? Uno estaría tentado a decir que no funcionan simplemente como directrices de generación y construcción, sino que *definirían* a cada clase de modelamiento, empero en opinión de F&P el paralelismo y la interacción de procesos cognitivos no es incompatible con la clase más general de “arquitecturas clásicas”⁶². Luego, la duda más inmediata sería, ¿existe modo de diferenciar de manera unívoca conexionismo y simbolismo?

(.10) Tanto (.08) como (.09) conducen de modo paralelo a una cuestión más profunda: la hibridación de los modelos, tema que ha sido tratado dentro de la literatura apuntada en (1.2.)(B1.1) y (1.2.)(B1.2), pero que *no ha sido aún agotado*. Por una parte, desde lo que se ha insinuado en (.08), sobre superar lo que se denominó “imagen simplificada”, podemos suponer que el conexionismo debe su trasfondo conceptual al simbolismo cognitivista⁶³, de lo que se sigue que puede ser pensado en ciertas de sus partes como siendo una variación desde el simbolismo. Por otra parte, de (.09) existen dos posibles respuestas: en caso de existir rasgos unívocos, entonces la hibridación podría ser caracterizada como teniendo un trasfondo – conexionista o simbólico – dominante. En caso de no existir rasgos unívocos, entonces se podría hablar de un continuo desde el simbolismo al conexionismo con los rasgos característicos de cada clase de modelos como los extremos ideales – pero nunca alcanzados – por la gama completa de implementaciones de modelos específicos, ¿cuál de estas salidas sería la adecuada?

(.11) Abundan los ejemplos donde el grupo PDP ciñe sus modelos a la explicación de tareas motoras. Estos ejemplos se mueven en el terreno de la simulación puramente y no involucran una aplicación directa en la robótica. No obstante, existen a partir de los 90’ implementaciones de la clase de ANN que son usadas paradigmáticamente por el grupo PDP⁶⁴. La implementación material y su éxito, o su fracaso, ya ha sido usado en literatura en ciencias cognitivas como argumento a favor o en contra de la fuerza explicativa, predictiva y descriptiva de enfoques completos. No obstante, me parece *a priori* apresurado dar el salto de modo de suponer que, abstractamente, es válida toda argumentación que parta desde cierto éxito técnico de alguna prototeoría a su potencial éxito o fracaso como perspectiva teórica desarrollada. Sin ir más lejos, los relativos fracasos de la investigación alrededor del perceptrón y los

⁶² Para ejemplificar un caso recuérdese el caso de SOAR apuntado en la nota 30.

⁶³ Trásfondo intelectual “compartido”, que más allá del uso común de los niveles de análisis de Marr, el realismo sobre el contenido representacional y el individualismo metodológico se agregan cuestiones como la *modularidad* (Rumelhart & McClelland, 1986, 84-6), empero este último concepto no será tratado para no sobrecargar el análisis

⁶⁴ Aun cuando los robots de Rodney Brooks constituirían ejemplos del fracaso del simbolismo cognitivista aplicado a la conducta robótica – aunque Brooks, no habla de simbolismo, sino que de “traditional IA” en contra su versión que sería la “nouveau IA” –, tal fracaso no encaja como una demostración de la utilidad de las redes PDP para llevar a cabo tareas motoras implementadas en un robot, dado que la arquitectura de subsunción no es ni conexionismo ni ANN, ni mucho menos simbolismo (Brooks, 1999, 96-7). Sin embargo, derivado del trabajo sobre ANNs se pueden encontrar aplicaciones desde principios de los 90’ donde una conducta motora muy simplificada de un insecto de 6 patas es reproducido por un robot (Beer et al., 1992). No obstante, la clase de ANN *heterogénea* usada dista bastante del enfoque del grupo PDP (Beer et al., 1989). Tardíamente se puede encontrar una implementación de una red *feedforward*, red neuronal paradigmática del grupo PDP, en el modelo de Walknet (Cruse et al., 1998) el cual ha evolucionado hasta el punto de, no sólo servir como modelo robótico para una tarea motora, sino que como una solución muy elegante de cómo emergería una consciencia primitiva (Cruse & Schilling, 2015).

relativos éxitos de la naciente escuela simbólica, según casi todas las variantes de la historia estándar del conexionismo, fueron el detonante de que la segunda se desarrollase hasta el punto de volverse, como se dice normalmente, “hegemónica”. Por tanto, la cuestión es bastante peliaguda y considero que requiere un tratamiento que es imposible de ser abarcado en el presente trabajo. Luego, la presente pregunta (.11), aun cuando es una reflexión más que válida que surge como pregunta por el desarrollo que tomó el conexionismo aplicado a las tareas motoras, no puede ser desarrollada en su alcance real, ya que involucraría el terreno de intersección entre filosofía de las ciencias y filosofía de la tecnología⁶⁵.

(.12) La conexión entre los modelos PDP y el *reduccionismo* es claramente falsa si se toma en cuenta que los mismos conexionistas se declaran “*emergentistas*”. Pero en sí mismas las etiquetas del reduccionismo y el emergentismo parecen estar construidas a partir de una *sobresimplificación del funcionamiento del modelamiento científico de los fenómenos de la naturaleza*. Si se da por sentado la legitimidad de esta premisa, entonces parece claro que muchas discusiones, tanto en filosofía de la mente, como en el campo de filosofía de las ciencias, debiesen ser desandadas. Empero, aún persistirá la pregunta que entrelaza las nociones del reduccionismo, el emergentismo y el fenómeno de lo mental, a saber, ¿cuál es la relación entre la mente y el cuerpo? o el denominado “problema mente-cuerpo”.

(.13) El *razonamiento explícito* es el fenómeno básico que da lugar a la escuela simbólica (Newell & Simon, 1963 [1961]). Es concedido por los conexionistas que – contextualizándose en la época que corrían los trabajos del grupo PDP – dar cuenta del razonamiento explícito y la consciencia serían las dificultades más grandes a afrontar para sus modelos, y que tal obstáculo quedaba a cargo de la investigación del futuro. Dejando de lado los intentos post-conexionistas ulteriores sobre esas facultades cognitivas⁶⁶, considero pertinente la pregunta; si dejamos de lado el cognitivismo y nos centramos en las primeras versiones del simbolismo pre-cognitivista, ¿no sería lo más razonable visualizar al *conexionismo* y al *simbolismo* como *complementarios* y no como *competidores*? Esta pregunta involucra directamente (a1.) y (.03), pero su respuesta requiere antes las respuestas tanto de (.01) como de (.03)

(.14) F&P consideran que de lo no representacional sin intencionalidad nada representacional se sigue, esto debido a que, por definición, una arquitectura cognitiva tomaría ciertos primitivos cognitivos de modo de dar cuenta de los procesos y el espacio de representaciones complejas que conforman las funciones cognitivas. Sin embargo, la noción de arquitectura cognitiva que está presupuesta en el marco conexionista dista de la noción simbólica, siendo esta la causa más probable de que, a pesar de la expresa intención común de clásicos y conexionistas de estar lidiando con el mismo fenómeno, terminen

⁶⁵ Esta finta filosófica puede parecer extraña a ojos no entrenados, ya que la salida fácil es tomar por equivalentes la tarea de dar sentido a datos conductuales/empíricos mediante un modelo susceptible de simulación, con la tarea de simular (o bien reproducir) mediante una implementación material los datos conductuales u observables de un fenómeno. Pero, obviamente, no son equivalentes; considérese fenómenos rudimentarios como la combustión, donde, mientras es un hecho histórico el dominio técnico de su implementación material – ¡el dominio del fuego! – su explicación devino a ser parte del corpus científico recién milenios después, con la teoría química que describió la relación entre la combustión y el oxígeno. *Replicar el fenómeno, aun cuando por casualidad tenga como base los mismos poderes causales, no es equivalente a la explicitación científica de tales poderes causales.*

⁶⁶ Un artículo contemporáneo que propone desde el conexionismo cómo se podrían solucionar tales problemas es Cleeremans (2014). Quizás el caso más destacable, por su simpleza y elegancia conceptual sea el de Cruse & Shilling (2015).

ubicando las representaciones en lugares totalmente opuestos dentro de sus construcciones teóricas – en el *explanans* versus el *explanandum* respectivamente. Más allá de la causa específica más probable de este fenómeno – que en cierta medida contestaría parcialmente (.07) –, lo interesante es *situar* el “contenido representacional”, así como el “realismo representacional”, en *una estructura que dé sentido a estos movimientos relativos de los conceptos dentro de teorías distintas*.

(.15) El argumento de que la relación entre los nodos es puramente causal, versus causal y estructural – o, específicamente, de “constitutividad”, en el caso de los símbolos – es sostenida por F&P fuertemente y es repetida por Fodor a lo largo de todas sus declaraciones contra Smolensky. Sin embargo, parece haber algo profundamente equivocado en suponer que los símbolos están “causalmente” relacionados entre sí, aunque si pueda existir entre ellos la relación “lógica” o “estructural” que Fodor les achaca, así como también parece ser erróneo suponer lo mismo con la relación “causal” entre los nodos. El error es simple, pero requiere de un argumento anterior que quite el status privilegiado que tienen las simulaciones computacionales, tanto en el enfoque cognitivista⁶⁷, como el conexionista tipo PDP (el cual, pese a quien le pese, carga con los mismos problemas conceptuales del cognitismo por entrar en una relación demasiado cercana con éste). Suponiendo este argumento anterior como ya demostrado, entonces el error es éste: *los instrumentos formales carecen de causalidad, dado que la causalidad es un hecho metafísicamente perteneciente al reino de lo empírico, de lo que se sigue que parasitan de esta propiedad en cuanto son interpretables como dando sentido a los datos del mundo*. No obstante, la pregunta que se sigue es aquella que da lugar a toda la literatura indicada la sección (1.2.2.)(B2) a saber, ¿existen “en realidad” relaciones de constitutividad como las descritas por Fodor, así como “sistematicidad”, “composicionalidad” o “productividad” de la cognición? ¿O son ellas más bien impuestas desde la teoría cognitivista sin tener en cuenta la realidad? ¿O existe un punto medio entre ambos extremos, es decir, una *existencia relativa* de tales nociones?

(.16) Nuevamente, en continuidad con otra reflexión anterior, a saber, (.10), la hibridación parece un problema muchísimo más profundo que lo que en principio puede ser visualizado. F&P plantean que un modelo clásico puede ser paralelo sin perder su condición de tal, y Lange (1992) considera que los modelos *marker-passing* se mueven aun en el mismo nivel que las redes conexionistas. No obstante, si uno recurre a una historia más detallada de las ciencias cognitivas, entonces es posible observar que la *hibridación* pasaría de ser un hecho aislado que se sitúa como una reacción específica a una situación anómala dentro de una disciplina – en este caso pensada como una reacción a la polémica simbolismo-conexionismo –, a transformarse en *un fenómeno que lo permea todo indirectamente en el campo del modelamiento cognitivo*. Luego, lo pregunta que surge es ¿qué consecuencias tiene esto para la propia comprensión del debate entre el simbolismo y el conexionismo?

(.17) Ahora ya moviéndose un poco más lejos de los márgenes de lo expuesto aquí hay varias cuestiones relaciones más indirectamente con el tema. Por ejemplo, ¿cuál es la diferencia “esencial”⁶⁸ entre la crítica de Minsky y Papert (1969) y la crítica de F&P (1988)? Obviamente abstrayéndose del hecho de que

⁶⁷ El status especial que poseerían las simulaciones computacionales cognitivas en el cognitismo es uno de los rasgos que derivan del principio de “teoría” = “programa” (Newell & Simon, 1961).

⁶⁸ Dado que no encuentro una palabra mejor me remitiré a esa, aunque la diferencia, como se verá más adelante, remite a factores que serán individualizados según mi propia tesis, la de los “Instrumentos Conceptuales”.

el primero tenía como foco crítico los perceptrones de Rosenblatt y los segundos el conexionismo general expuesto por el programa PDP. Por otro lado, ¿en qué sentido vale la pena revisar la historia estándar del conexionismo y las ANN que sea atingente al resto de las reflexiones aquí hechas?

(.18) Finalmente, y quizás la pregunta que da comienzo a todas las reflexiones, pero que la colocaré como cierre, es *¿cómo desde la filosofía de las ciencias en general es posible comenzar una empresa que, no solamente dé sentido tanto al debate central como a sus consecuencias, sino que sea útil para dirimir entre diferentes “teorías en competencia” en ciencias cognitivas?*

Pues bien, antes de llegar a la síntesis – sección (1.3.4.) – que indicará cual es el “nudo gordiano” de esta serie de reflexiones y de preguntas de modo de permitir una investigación filosófica mejor delimitada, me interiorizaré en un par de cuestiones básicas que están justo en la intersección entre lo metafísico y la filosofía de las ciencias en general. Estas dos digresiones tienen un objetivo común; asentar la irrupción en la próxima parte de la tesis del *pluralismo explicativo*, sin tener que dar el larguísimo rodeo que implicaría presentarlo desde el progreso de la historia de la filosofía de las ciencias.

1.3.2. Digresión 1: Causalidad, instrumentos formales versus empíricos y el cognitivismo.

Intuitivamente diferenciadas, sean por un lado las teorías de ciencias naturales (física, química, biología, psicología, etc.) denominadas “instrumentos empíricos”, mientras que por el otro, las teorías en ciencias formales (matemática, lógica, ciencias de la computación, etc.) denominadas “instrumentos formales”. Mi primer supuesto a defender es:

Los instrumentos formales carecen de causalidad, dado que la causalidad es un hecho metafísicamente perteneciente al reino de lo empírico, de lo que se sigue que parasitan de esta propiedad en cuanto son interpretables como dando sentido a los datos del mundo.

Este supuesto, lamentablemente, soy incapaz de presentarlo de modo universal para todo instrumento formal, ya que hacerlo implicaría un trabajo más amplio desde las filosofías de las matemáticas y lógica (entre otras). No obstante, puede aplicarse a dos instrumentos formales de uso común en la teorización dentro de las ciencias cognitivas, a saber, sobre los sistemas dinámicos⁶⁹ y las máquinas de Turing.

Un sistema dinámico es presentado con diferentes definiciones dependiendo el contexto (Bezruchko & Smirnov, 2010, 26-7):

- Como sinónimo de la expresión “un conjunto de ecuaciones diferenciales ordinarias $dx/dt = f(x, t)$ ”;
- Como sinónimo de la expresión “un conjunto de ecuaciones diferenciales ordinarias autónomas $dx/dt = g(x)$ ”;
- Un modelo matemático de algún sistema mecánico⁷⁰.

⁶⁹ La hipótesis dinámica de Van Gelder supone una separación estricta entre los sistemas dinámicos entendidos matemáticamente y su propia noción de sistema dinámico cognitivo (Van Gelder, 1998). Ergo, no se tome esta digresión ni a favor ni en contra del dinamismo, dado que ella es ortogonal a la validez del mentado enfoque.

⁷⁰ Nótese que, en general, un sistema dinámico siempre es tomado como un modelamiento basado en un conjunto de ED cuya variable independiente por antonomasia es el tiempo.

Para simplificar quedémonos con la primera que dice que es un conjunto de ecuaciones diferenciales ordinarias (EDO). Considérese una EDO lineal de segundo orden cualquiera como⁷¹:

$$F(x, t) = \alpha \frac{d^2x}{dt^2} + \beta \frac{dx}{dt} + \gamma x$$

En esta ecuación α , β y γ representan coeficientes cualesquiera, siendo t el tiempo como la variable independiente y x alguna variable dependiente de éste. Puesto que la primera definición de “sistema dinámico” implica más de una ecuación, entonces me apoyaré en que es siempre posible que una ecuación diferencial lineal de segundo orden sea transformada en un sistema de dos ecuaciones diferenciales lineales de primer orden para cumplir tal primera definición. Supongamos $y = dx/dt$, entonces nuestro simplísimo sistema constará del siguiente conjunto de ecuaciones diferenciales:

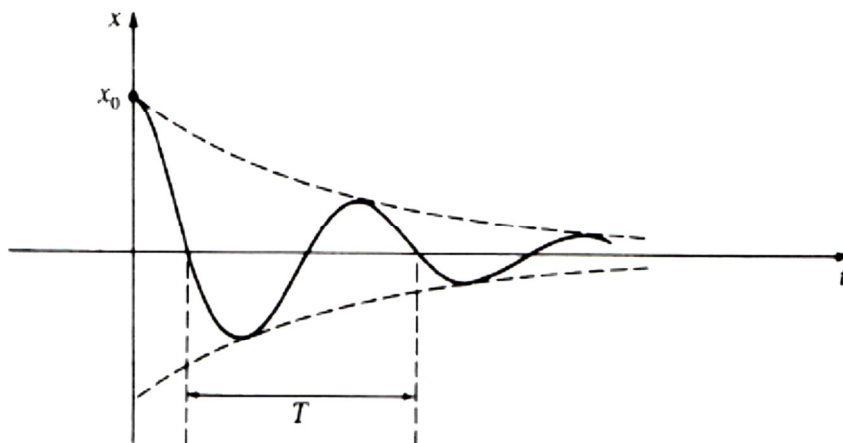
$$F(x, t) = \alpha \frac{dx}{dt} + \beta y + \gamma x$$

$$0 = \frac{dx}{dt} - y$$

La anterior argucia es hecha exclusivamente para adaptarse a la noción de “sistema dinámico” antes elegida, ya que por sí sola no permite avanzar, precisamente, en cómo tales ecuaciones en sí mismas permiten decirnos algo respecto al mundo. De hecho, *inclusive si se especificasen los coeficientes α , β y γ mediante magnitudes físicas, aun así no tendríamos claro cómo es que tales magnitudes físicas interactúan causalmente entre sí sin apelar al instrumento empírico o teoría física específica de origen.* Considérense los siguientes coeficientes: $\alpha = L$, $\beta = R$ y $\gamma = 1/C$ versus $\alpha = 1$, $\beta = c/M$ y $\gamma = k/M$, cada conjunto de ellos permite derivar las siguientes EDO lineales de segundo orden:

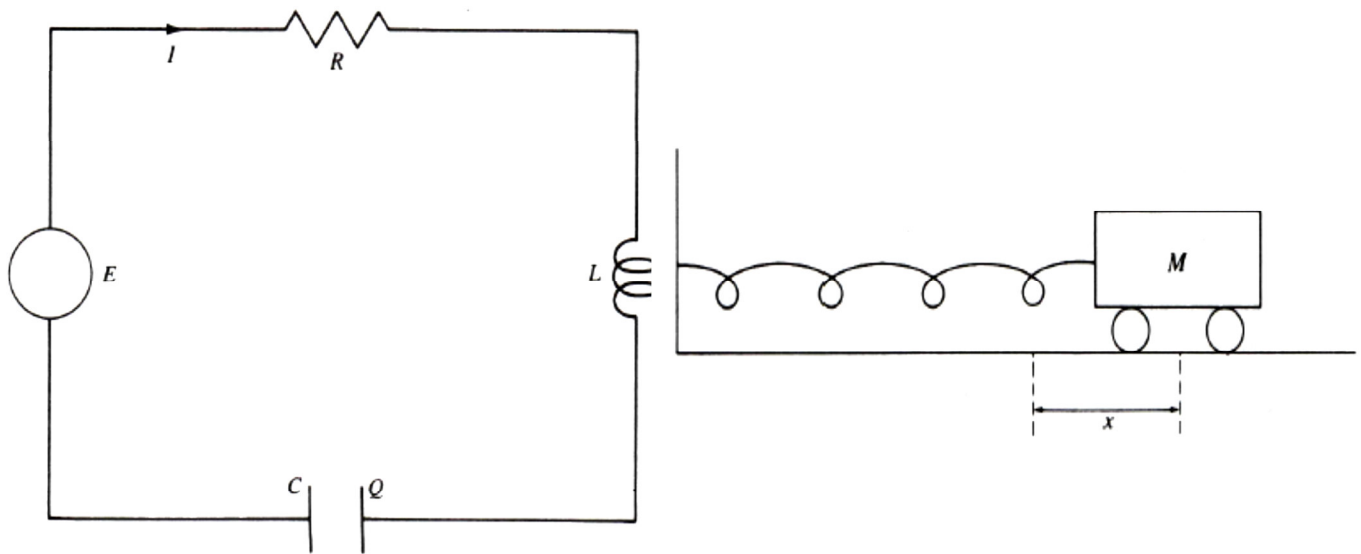
$$(1) E = L \frac{d^2Q}{dt^2} + R \frac{dQ}{dt} + \frac{1}{C} Q \quad (2) 0 = \frac{d^2x}{dt^2} + \frac{c}{M} \frac{dx}{dt} + \frac{k}{M} x$$

Llevaremos la cuestión más lejos ahora; tanto (1) como (2), si son graficadas, desplegarán una curva sinusoidal amortiguada como la siguiente (caso ecuación (2) **FIGURA 8**):



⁷¹ Las ecuaciones y las figuras a continuación son extraídas de los primeros dos capítulos de Simmons & Robertson (2002).

Y aun con las ecuaciones y los gráficos, *si no recurrimos a un instrumento empírico para determinar Q en (1) y x en (2), entonces sencillamente no tenemos manera de saber aún a cuál fenómeno nos estamos refiriendo, menos aún la causalidad de los fenómenos dados.* Las ecuaciones (1) y (2) representan, respectivamente, un circuito cerrado compuesto de una fuente de fuerza electromotriz (o voltaje) E , una resistencia R , un inductor con inductancia L y un condensador de capacitancia C , mientras que el otro sistema se trata de una masa M sujeta por un muelle a un muro, donde k es una constante positiva que mide la rigidez del resorte y c es otra constante positiva que mide la resistencia del medio (la **FIGURA 9** a continuación muestra el circuito versus el sistema mecánico).



Luego, respectivamente, Q es una carga en cierto instante t y x es posición en cierto instante t , es decir, la ecuación (1) es la oscilación de una carga hasta llegar a 0 y la ecuación (2) la oscilación de una masa sacada de su punto de equilibrio hasta volver al origen. Ergo, *la causa de la oscilación en cada sistema no va ínsita en el formalismo que se utiliza, sino que depende de nuestro conocimiento de relaciones entre objetos en el mundo que son abstraídos en sus rasgos relevantes, es decir, su comportamiento eléctrico o su cinemática.* El punto a destacar es bastante sencillo; un mismo conjunto de EDOs puede ser utilizado para una misma situación⁷², lo que significa que *el formalismo por sí solo es insuficiente para implicar una semántica unívoca, lo que se extiende con mayor razón para el caso de la causalidad.*

Por otra parte, una máquina de Turing (MT), análogamente a las múltiples acepciones de “sistema dinámico”, tiene diferentes presentaciones y utilidades dependiendo el contexto. Si uno va a la presentación original, la MT es presentada por Alan Turing (1936) en el curso de la demostración formal de que el problema de decisión⁷³ planteado por Hilbert y Ackermann no tiene solución. Partiendo de la definición de que un número es computable si sus decimales pueden ser escritos por una máquina en un tiempo *finito* – restricción justificada en “the fact that the human memory is necessarily limited” (Turing, 1936, 231) –, la MT aparece como definida abstractamente del siguiente modo (op. cit):

⁷² De hecho Edelman (2008, 182-3) utiliza una situación análoga para terminar diciendo que un sistema dinámico cualquiera es una *forma de computación*. Creo que esta conclusión tiene fuertes compromisos ontológicos que aquí estoy tratando de deshacer, o al menos mantener al mínimo.

⁷³ Mejor conocido como el *Entscheidungsproblem*.

- Como un dispositivo con un número finito de condiciones o m-configuraciones – las cuales serían los constituyentes de una “tabla de máquina”
- Tal dispositivo tendría una cinta (un análogo al papel) corriendo a través de éste y que estaría dividida en secciones cuadradas o celdas, cada una capaz de portar un “símbolo”.
- Tal símbolo podría ser escaneado desde una celda o impreso en una. El primero, el escaneado, sería aquel que es directamente “consciente” para la máquina. Sin embargo, mediante alteraciones de las m-configuraciones sería posible que la máquina “recordase” símbolos en celdas anteriores.

Por consiguiente, si me restringiese a esta descripción de la MT no sería necesario entrar ni a especular ni a tener que mostrar cómo la MT no contiene ninguna directriz que la conecte con la causalidad del mundo, ya que es por sí misma una idealización que tiene una utilidad que se restringe a cierta prueba formal. Sin embargo, esta clase de argumentación sería mero hombre de paja, ya que el cognitivista cuando apela a la parentela entre los modelos clásicos de la cognición y la MT va por otro lado; la mente es tratada “como-si-fuera-MT” y, en consecuencia, *la causalidad de los símbolos contenidos de tal mente-como-si-fuera-MT derivaría de ser una extensión de la cadena causal entre las entradas, su transducción y las salidas*. Sin embargo, ello es simplemente remitirse a una versión muy metafórica de la MT, que no se condice ni con su origen, ni con su contexto contemporáneo, la teoría de autómatas, donde la MT se considera aquel autómata más general que modela el poder computacional de los computadores reales, que permite estudiar la decidibilidad – aquello que puede hacer y no hacer un computador – y que posibilita estudiar la dicotomía entre problemas solucionables en tiempo polinomial de aquellos que no (Hopcroft et al., 2006), es decir, *no hay ninguna cuestión en el formalismo en sí que lo conecte al mundo, si no es mediante una teoría extra que debe antes justificar cómo el fenómeno se entrelaza causalmente entre sus partes*. Nuevamente el formalismo, inclusive usado muy metafóricamente, *parasita de supuestos que tienen que ver con lo empírico*.

En este punto el especialista podría decir “pero esto es una perogrullada que no requiere una ejemplificación tan detallada”, a lo cual solo podría dar la razón. El supuesto sobre la carencia de causalidad en los instrumentos formales puede ser demostrado de manera aún más simple que adentrándose en entender EDO y sus aplicaciones o las nociones originales y contemporáneas de MT. Tal demostración se basaría, solamente, en remitirse a la noción de “función” en matemáticas, noción que puede ser definida de manera muy intuitiva y sencilla como “la *relación* que se establece desde los elementos de un conjunto A, o dominio, a los elementos de un conjunto B, o recorrido.”. Luego, en tanto que es una relación no definida más que como aquella que permite “ir” desde cierto a_i a cierto b_j , no requiere que sea conceptualizada como portando causalidad en el mapeo entre los elementos. De hecho, una función lineal, elementalísima, puede graficar la *relación analítica* – en sentido kantiano – entre el paso de cierta cantidad de años y su correspondiente cantidad de meses, pero tal relación no es de causalidad, ya que el paso de los años no es causa del paso de los meses, ni viceversa, sino que su transcurrir tiene como causa algo común a ambos, la traslación de la Tierra alrededor del Sol y no puede ser capturada por el formalismo, a menos que éste esté constreñido por alguna semántica de un instrumento empírico.

Por consiguiente, dada cualquier función, *a priori* es imposible saber qué clase de relación está graficando, lo mismo se seguiría para cualquier algoritmo, en tanto que un algoritmo puede ser comprendido como una función, puesto que por definición un algoritmo “hace” exactamente lo mismo que una función, es decir, un mapeo entre conjuntos de elementos. Ahora bien, si damos por suficientemente autoevidente que *los instrumentos formales carecen de causalidad*, entonces ¿cómo este supuesto es relevante para la discusión entre conexionismo y cognitivismo? Pues bien, para avanzar en esta digresión vale la pena recapitular sobre la argumentación de F&P.

Para F&P las relaciones tipo “parte/todo” y “conjunto/subconjunto” – achacadas al conexionismo – son diferentes entre sí, dado que las primeras serían las que permiten la operación sobre representaciones complejas, operaciones que se llevarían a cabo gracias a *cierta descripción estructural o lógica común* que trascendería a representaciones tales como “ $P \wedge Q$ ” y “ $P \wedge (A \vee B \vee C)$ ” y que explicaría naturalmente cuestiones cognitivas como la sistematicidad de la inferencia. Luego, dado que *las representaciones mentales tienen una estructura interna sensible a operaciones sobre ellas* – que sería una síntesis en pocas palabras de los rasgos arquitecturales [1] y [2] (sección (1.1.2.)) –, en el mismo grado que se da en las expresiones lingüísticas, entonces existiría *composicionalidad* en las representaciones mentales, lo cual a su vez explicaría que los pensamientos estén sistemáticamente relacionados. Fodor (1997) dirá que la esencia de la explicación de la composicionalidad en los modelos clásicos es: si los efectos de instrumentar “vaca café” se traslapan parcialmente con los efectos de instanciar “tronco café”, eso es porque, en perfecto sentido literal, *las causas de estos efectos se traslapan también* (Fodor, 1997, 113-4), siendo tales causas las relaciones de “parte/todo” o constitutividad entre representaciones de diferente complejidad.

Ahora bien, es dudoso que la relación “parte/todo” sea de causalidad, en tanto un modelo clásico es un formalismo derivado de nociones en tales como “máquina de Turing” e instrumentos formales aledaños los cuales, por definición – o por lo que he desarrollado antes, como se guste –, no portan causalidad. Sin embargo, por mor de no crear un hombre de paja, vale la pena considerar un instante la argumentación fodoriana. Supongamos que F&P están en lo correcto, y *concedamos la existencia empírica de representaciones mentales complejas*, tales como “tronco café” y “vaca café”, de manera de transformar a la teoría clásica cognitivista en un instrumento empírico que tenga, por definición, acceso a la propiedad metafísica de portar causalidad en sus conceptos ¿se sigue de esto, por ejemplo, la existencia de composicionalidad – ergo, de la capacidad explicativa del cognitivismo de la sistematicidad? Por supuesto que no, dado que aseverarlo sería caer en la afirmación del consecuente, y el consecuente en este caso es una proposición de alcance universal. Este hecho ya está bien caracterizado en filosofía de las ciencias y es lo que llevó a Karl Popper a la postulación del falsacionismo.

Para F&P es obviamente cierto que la composicionalidad no es una característica de las representaciones conexionistas, simplemente porque sus representaciones mentales no tienen estructura combinatoria – representaciones mentales atómicas debido a las restricciones conexionistas, según la perspectiva cognitivista –, pero en su contra pesa que *no necesariamente existe univocidad para llegar desde la presencia de representaciones mentales complejas empíricas a la composicionalidad*, lo cual deja en el mismo nivel de incertidumbre al conexionismo como al simbolismo cognitivista en cuanto a su capacidad de dar cuenta de la sistematicidad. En F&M se hace explícito este hecho, al aceptarse que la

composicionalidad no necesariamente es requerida para explicar la sistematicidad (Fodor & McLaughlin, 1990, 185) y que el modo dar cuenta de la sistematicidad (como ley de la naturaleza) es mediante los rasgos arquitecturales [1] y [2] (op. cit., 188), siendo esto último lo que define a los modelos clásicos.

Pues bien, la conclusión hasta aquí es que, aun cuando se diese por autoevidente que *existe una causalidad ínsita en el “flujo mental de símbolos complejos”*, esto no significa que la explicación cognitivista sea la única posible, más bien que cualquier otra teoría cognitiva concebible puede colocar una causalidad ínsita en el nivel de análisis que se le plazca, dado que *no pareciera existir razones empíricas para suponer que la causalidad esté en uno u otro nivel de análisis*. Aquí Smolensky, como representante del conexionismo, cae en el error de suponer que aceptar nociones desde el cognitivismo⁷⁴, lo deja libre para suponer que no es el nivel simbólico donde la causalidad se presenta, sino que sería en el nivel subsimbólico, lo cual sería incompatible con partir con nociones cognitivistas que consideran, por ejemplo, que *la sistematicidad es una ley de la naturaleza*. La importancia de que las representaciones mentales complejas dentro de los modelos clásicos porten causalidad en F&M es aceptada en su cabalidad cuando dicen que:

No doubt, we do get greater precision when we go from generalizations about operations on constituents to generalizations about operations on constituents to generalizations about operations on units. But if that shows that symbol-level processes aren't really causal then it must be that unit-level processes aren't really causal either. After all, we *get still more* precision when we go down from unit-sensitive operations to molecule-sensitive operations, and more precision yet when we go do from molecule-sensitive operations to quark-sensitive operations. The moral is not, however, that the causal laws of psychology should be stated in terms of the behavior of quarks. Rather, the moral is that whether you have a level of causal explanation is a question, not just of how much precision you are able to achieve, but also of what generalizations you are able to express. The price you pay for doing psychology at the level units is that you lose causal generalizations that symbol-level theories are able to state. Smolensky's problems with capturing the generalizations about systematicity provide a graphic illustration of these truths. (Fodor & McLaughlin, 1990, 203-4).

Luego, un cognitivista hasta aquí podría decir, “ok, algo de duda has plantado al poner en entredicho que la causalidad esté en el nivel de análisis simbólico, en tanto es posible ponerla en el nivel subsimbólico si se aceptase desde cero la construcción teórica de Smolensky, sin embargo, ¿cuáles pruebas puedes aportar tú para decir que la causalidad no está en el nivel donde se desenvuelven las representaciones mentales complejas?”. Ante esta pregunta simplemente se replica mediante la carga de la prueba, es decir, corresponde al cognitivista demostrar que las representaciones mentales complejas, primero, existen y, segundo, poseen poderes causales que medien entre ellas, lo cual el cognitivismo, al ser mitad psicología especulativa y mitad filosofía, carece. El cognitivista, entonces, podría responder con una pregunta retórica, “si no se acepta la existencia de causalidad en el nivel simbólico, entonces ¿cuál es el sentido de postular una teoría cognitiva que sea dependiente de la causalidad de un “nivel más abajo”?” Pero, la respuesta irá también por el lado del paladar filosófico donde ya nada se puede hacer, es decir,

⁷⁴ Tales como las contenidas en los principios (i), (ii), (iii) y (iv) en F&M, sección (1.1.3.) segunda parte del debate Fodor-Smolensky.

se puede contra-preguntar retóricamente ¿y cuál sería el sentido de hacer una teoría cognitiva que sea autónoma⁷⁵ al punto que no sea lícito preguntarse de dónde deriva la causalidad de sus constructos?

Es posible construir un argumento más elaborado sobre *porqué los “símbolos” pertenecen al “reino” de los instrumentos formales y no a los empíricos, de lo que se seguiría que no heredan la causalidad, a menos que existan motivos desde los instrumentos empíricos para lo contrario* – aunque, en lo personal, es suficientemente satisfactorio que F&P declarasen parentela entre las máquinas de Turing y los modelos clásicos. Sin embargo, tal argumento más elaborado puede ser erigido a partir de ciertos puntos que van en línea con una construcción conceptual adecuada desde la filosofía de las ciencias, la cual dicha en una línea sería: *el problema del cognitivismo está en la etapa de asentamiento-madurez debido a tratar de naturalizar el vocabulario de instrumentos formales, obviando que, a lo más, pueden ser usados representando, pero no portando causalidad*. Este argumento más elaborado deriva de una abstracción basada en una historia de las ciencias muy general – pero no por ello menos plausible – que mostraría cómo cualquier teoría adquiere nuevos conceptos y de la cual se seguirá la razón de por qué no podría adscribirse causalidad a los símbolos fodorianos⁷⁶.

Quizás, entonces, la última embestida del cognitivista sería “me parece bien que exista algo así como un argumento elaborado, pero si tal argumento depende de una construcción que yo no necesariamente compartiré sobre filosofía de las ciencias, entonces puedo seguir afirmando que la causalidad puede seguir siendo adscrita a las representaciones mentales complejas.”, a lo cual cabría responder que es legítima tal aguda observación final. Sin embargo, con lo dicho hasta aquí en la presente digresión, me parece que he planteado suficientes objeciones como para demostrar lo que postulaba al principio de la subsección, a saber, que los instrumentos formales o formalismos en general, *parasitan la causalidad de la semántica que se les dé, esta última propiamente derivada del instrumento empírico donde están siendo usados*. Luego, lo que faltaría no sería una demostración de que el simbolismo cuenta como un formalismo – ello puede estipularse o no, pero en el fondo no es una discusión que le interese al cognitivista –, sino que su semántica que lo naturaliza depende del enfoque cognitivista, el cual, por razones del mismo desarrollo histórico de las ciencias cognitivas, *jamás ha llegado a pasar desde la fase de asentamiento a la de madurez*⁷⁷.

1.3.3. Digresión 2: Sobre la interpretación de la Paradoja de Smolensky y sus consecuencias.

Sucintamente, poniéndolo en términos filosóficos, mi segunda suposición para sustentar el trabajo a continuación es que:

La resolución de la paradoja de Smolensky no deviene de tomar partido ontológico, sino que deviene de la suspensión del juicio con respecto a un enfoque sobre el qué de la cognición, lo que posibilitaría una interpretación puramente epistemológica de ésta.

⁷⁵ La autonomía es defendida por Fodor en su artículo clásico *Special Sciences (or the disunity of science as a working hypothesis)* (1974).

⁷⁶ Este “argumento elaborado” ya tiene a la vista lo que se desarrollará en la sección (2.3.2.2.). Remito al lector a su lectura directa para comprender más exactamente a qué me refiero con nociones como “asentamiento” o “naturalizar” las cuales son claves en estos dos párrafos finales de la digresión 1.

⁷⁷ Específicamente este punto será retomado en (2.5.iv)

Para mostrar los problemas de tomar ontológicamente la paradoja, al igual que en la digresión anterior, comenzaré con un caso relativamente concreto (que funcionará cual “bomba de intuición”) y de ahí directamente extraeré la conclusión. El caso relativamente más concreto es este: imaginemos una situación donde ponemos cuatro individuos uno en cada esquina de una plaza a buscar un fajo de billetes *escondido* en algún lugar de ésta, de modo de contestar una pregunta como la siguiente, ¿cuál de ellos lo encontrará primero?

Dada que es una situación simple de constreñir experimentalmente y replicar en un contexto controlado, entonces podemos, sencillamente, tomar cuádruplas de sujetos de prueba voluntarios y extraer de varias repeticiones alguna forma de predecir cuál de ellos será aquel que más fácilmente encontrará el fajo, en relación con, supongamos, la distancia entre su ubicación en una esquina y la ubicación de la recompensa, el fajo de billetes, o la velocidad con la que se mueven en nuestra plaza preparada para la prueba, u otras variables externas que refieran a la *conducta observable* de nuestros sujetos.

Sin embargo cuando preguntamos “¿cuál de ellos lo encontrará primero?”, *desde nuestra perspectiva como científicos cognitivos no nos interesaba meramente extraer un patrón conductista*, como aquel que se describió justo arriba, más bien nos interesaría saber, por ejemplo, ¿cuál mecanismo es el que llevó a cabo la tarea?, ¿cuál fue la relación entre tal mecanismo interno y el ambiente?, ¿cuántas estrategias cognitivas de búsqueda existen?, etcétera.

Luego, nuestro experimento deberá variar de modo de dar cabida a responder la mayor cantidad posible de nuestro indeterminado etcétera de preguntas. Siguiendo en este caso, puesto que nuestra cantidad de fajos de recompensa crece aritméticamente a medida que requerimos nuevas cuádruplas de sujetos, y requerimos responder variaciones sutiles de la misma pregunta original traducida a un *vocabulario científico cognitivo* – más no cognitivista –, entonces se vuelve necesaria *alguna metodología* que facilite el trabajo que, por razones puramente *prácticas*, se ha vuelto progresivamente más complicado.

El camino a seguir por excelencia que se toma frente a esto es la simulación computacional, la cual presenta diferentes ventajas prácticas frente a la experimentación empírica, por ejemplo, que es más barata y con mayor grado de generalidad. En este ejemplo, una simulación podría traducirse en el uso amplio de autómatas celulares (CA), más específicamente, de alguna simulación que implemente un modelo basado en agentes (ABM) que represente nuestra búsqueda del fajo perdido en la plaza.

Informalmente un ABM puede ser pensado como un CA con un territorio y actores más sofisticados que se vuelven más cercanos para emular a los seres humanos en diferentes aspectos tales como el razonamiento, la toma de decisiones y la conducta. El comportamiento complejo que emerge en un ABM es similar al de un CA, pero sus agentes pasan a tener características como (a) autonomía, mientras que en un CA cada célula depende fuertemente de su vecindario, (b) libertad de movimiento, contra el confinamiento de las células a localizaciones fijas, (c) comportamiento basado en razonamiento, entre otros rasgos. El formalismo de los ABM es especialmente útil para las ciencias sociales y conductuales, por tanto nos servirá para ejemplificar cómo sería un caso de simulación computacional que intente crear un modelo de cómo se lleva a cabo una tarea donde intervienen agentes cognitivos en competencia, como es nuestro caso hipotético. (Cioffi-Revilla, 2014, 287-290)

Supongamos que llevamos a cabo dos simulaciones partiendo de conjeturas sobre la racionalidad totalmente distintas de los agentes. Por un lado, una simulación que parta de una racionalidad cuya regla explícita general portada por cada sujeto sea algo así como “siga en línea recta hasta encontrar la recompensa o chocar contra un obstáculo; en caso de encontrar la recompensa detenerse, en caso de chocar contra un obstáculo cambiar de dirección.” y, por otro lado, una simulación donde la racionalidad sea guiada por “micro-decisiones” totalmente aleatorias. El primer caso sería una especie de extremo de tipo “cognición rígida”, siguiendo la terminología de Smolensky, mientras que el segundo caso sería el otro extremo de “cognición flexible”.

Cada una de las búsquedas se vería más o menos del siguiente modo simple:

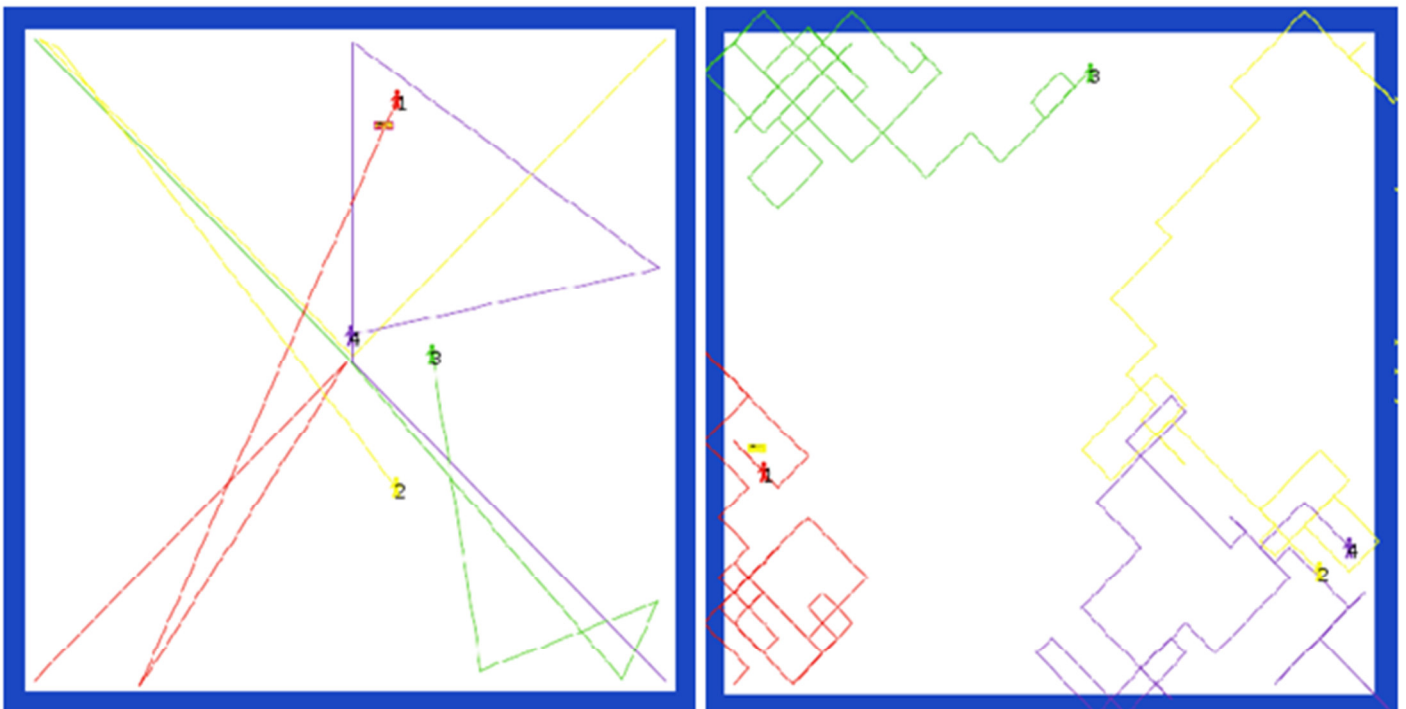


FIGURA 10; Aquí se observa el funcionamiento de dos sencillísimas simulaciones en Netlogo (Wilensky & Rand, 2015), “búsqueda en la plaza versión cognición rígida” versus “búsqueda en la plaza versión cognición flexible”. En ambos casos el sujeto 1 (recorrido marcado en rojo) es el primero en encontrar la recompensa colocada aleatoriamente, representada por el pequeño rectángulo.

Ahora bien, dejando de lado la notoria artificiosidad “rígida” que ambas poseen – debido a que ambas simulaciones las he construido a partir de la misma especie reactiva de agentes⁷⁸, por tanto, por ejemplo, carecen de memoria para recordar si recorrieron o no la misma trayectoria –, pareciera ser bastante obvio que ninguno de los dos extremos representa la “verdadera” situación, es decir, ninguno captura el comportamiento real de los sujetos en la búsqueda en la plaza, ya que, la primera simulación basada en un tipo de *cognición rígida* no logra cubrir extensivamente el terreno, mientras que la *flexible* (aleatoria) no muestra un mínimo coherencia que sin dudar adscribiríamos a la conducta de búsqueda.

⁷⁸ Los agentes pueden ser de tres tipos: reactivos (“reflexive”), basados en utilidad y basados en metas. Estos dos últimos serán adaptativos (Wilensky & Rand, 2015, 224-232) Para el punto que deseo mostrar no es necesario el agente de mayor sofisticación cognitiva, el basado en metas, aun cuando también puede hacerse.

Por consiguiente, si lleváramos a cabo el experimento de búsqueda en la plaza, recolectásemos los datos y quisiésemos representar mediante una simulación que cumpliera el primer requisito de adaptarse mejor a tal recolección empírica, ya en ese punto tendríamos que abdicar de continuar tozudamente en nuestra trinchera ontológica de modo purista y conceder el considerar una mezcla de ambos enfoques cognitivos. Por tanto, al hacer esta concesión ya se estaría dando por hecho, implícitamente, algo que es básico en los instrumentos empíricos, a saber, que son los hechos los que determinan nuestras explicaciones, y que si debemos forzar la evidencia para que calce en nuestros marcos, entonces claramente algo incorrecto estamos haciendo. *El privilegio a ultranza de ciertos supuestos ontológicos sobre otros es una mala consejera para desarrollar descripciones, explicaciones y predicciones.*

Llevando esta intuición al plano del debate entre los extremos del racionalismo y el eliminativismo, entonces – asumiendo que estuviese suficientemente probada la necesidad de postular representaciones y un plano cognitivo estrictamente en el nivel computacional de Marr – al tomar cualquiera de estas opciones la investigación sobre la cognición se paralizaría, porque la cognición no podría continuar siendo modelada con ninguno de los dos extremos en solitario. En conclusión, la paradoja contiene en sí el peligro de, si es tomada ontológicamente, poder conducir a un punto en que estancaría el progreso científico cognitivo, lo que implica que, incluso las posturas más moderadas como emergentismo desde lo flexible o emergentismo desde lo rígido, *deban ser interpretadas en su estricto valor epistemológico, sin hacer mayores adendas metafísicas.*

1.3.4. Síntesis del punto de partida teórico.

Todo lo que se desarrolló con cierto detalle en las secciones (1.1.), (1.2.) y sus respectivas subsecciones puede ser capturado, sin grandes pérdidas de contenido, en un mapa conceptual como el siguiente.

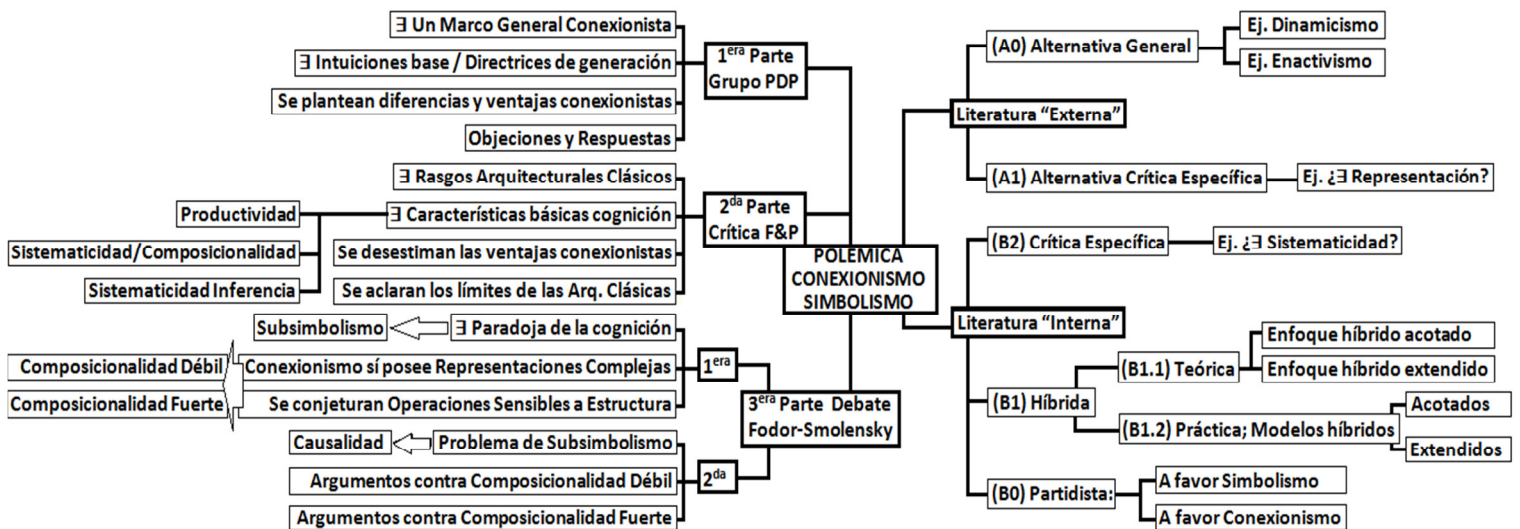


FIGURA 11. Mapa conceptual síntesis de la primera parte de la tesis. Casi en la mitad está situada la polémica, a la izquierda las cuestiones centrales de la sección (1.1.) y a la derecha las cuestiones referentes a la sección (1.2.).

Pues bien, los conceptos claves de cada una de las reflexiones desde (.01) a (.17) puede ser sintetizada por la pregunta (.18), a saber, “¿cómo desde las filosofías de las ciencias en general es posible comenzar una empresa que, no solamente dé sentido tanto al debate central como a sus consecuencias, sino que sea útil para dirimir entre diferentes “teorías en competencia” en ciencias cognitivas?” Esta pregunta tiene como trasfondo uno y solamente un concepto clave: la *competencia* entre teorías.

La polémica simbolismo-conexionismo, parece bastante claro, no puede ser solucionada “desde dentro”, ya que existen desacuerdos bastante fundamentales entre ambos, como por ejemplo respecto a la causalidad (sección (1.3.2.)) y respecto a tomar dogmáticamente una visualización estrictamente “ontológica” del problema de la cognición (sección (1.3.3.)). Literatura posterior, en general toda aquella que no cae dentro de la categoría de (B0), trata de salir en menor o mayor grado de tal “interioridad” de la polémica que le permita ir en la búsqueda de una salida de ella.

Mi estrategia a seguir irá en línea con esta literatura que mira la polémica “exteriormente”, sin embargo mi lugar más exacto será definido en el transcurso de la segunda parte (específicamente, al final de la sección (2.1.) y toda la sección (2.2.)). Puesto que creo que más allá de las desavenencias postuladas en (1.3.2.) y (1.3.3.), las cuales respectivamente se resumen en las *versiones incompatibles de causalidad* y del “*privilegio ontológico*” respectivamente, existen otros muchísimos vacíos en la misma discusión, los cuales dan lugar a las reflexiones desde (.01) a (.17), como por ejemplo ¿qué se entiende, un poco más detalladamente, por “metateoría” en el conexionismo? o ¿qué se entiende por “modelos clásicos” en F&P? Por consiguiente, el nudo gordiano será cortado a partir de redefinir lo más descriptivamente posible el fenómeno de la competencia, *núcleo de toda relación interteórica*.

Ergo, toda la siguiente parte de la tesis debe verse como un esfuerzo por llegar a una noción lo menos normativa posible que sea capaz de describir las diferentes maneras en que es posible la competencia – siendo tal punto de llegada lo que se desarrollará en la sección (2.3.) y sus subsecciones, específicamente lo que denominaré la “tesis de los IC” (“instrumentos conceptuales”) – para, finalmente, dentro de la sección de cierre de la tesis (2.5.) retomar más específicamente las reflexiones desarrolladas en (1.3.1) y alguno que otro tópico descrito en el transcurso de (1.1.) y (1.2.).

II. El Pluralismo Explicativo, la abstracción de los Instrumentos Conceptuales y su aplicación

Dentro de la primera parte, la sección (1.1.) y sus subsecciones, se trató en el mayor detalle posible la polémica entre el simbolismo cognitivista y el conexionismo, dejando de lado aspectos más específicos de la disputa, tales como los que corresponden al contraste entre modelos sobre procesamiento del lenguaje. Algunos de los tópicos tratados en la polémica y su tratamiento en conjunto fueron considerados para la organización de cierta muestra de textos en ciencias cognitivas dentro de la sección (1.2) y sus subsecciones. Una serie de reflexiones con cierta profundidad sobre (1.1.) y (1.2.) fue presentado en (1.3.1.), donde se conjeturó sobre la posibilidad de dar cuenta de las nociones de filosofía de las ciencias que trascienden a la polémica del simbolismo y el conexionismo. A partir de ahí las digresiones (1.3.2) y (1.3.3) sientan las premisas básicas que permiten en (1.3.4) sintetizar el punto de partida teórico del resto de la tesis que se está exponiendo.

Lo que da lugar a la presente segunda parte es que, si se aceptan las dos digresiones presentadas en (1.3.2.) y (1.3.3.), a saber, primero, que existe una separación entre instrumentos formales y empíricos, heredando la causalidad de los segundos los primeros y, segundo, que el proceso de modelamiento se restringe fuertemente si se da preeminencia a interpretaciones ontológicas sobre el fenómeno, entonces lo que se sigue es presentar un nuevo marco desde la filosofía de las ciencias generales – pero que tenga en consideración las limitaciones que existieron, por ejemplo, con la concepción heredada y el primer historicismo kuhniano – que sea capaz de empezar a responder al grupo de reflexiones apuntadas en (1.3.1.) y reordenadas (1.3.4.).

Esta segunda parte se dividirá en 5 secciones. La sección (2.1.) desechará la posibilidad, conjeturada y definida en la introducción, de encontrar un terreno común para ciencias cognitivas. Sin embargo, lo anterior no será impedimento para conjeturar un trasfondo que permita explicar la diferencia entre una disciplina madura y una inmadura, siendo este punto retomado en las conclusiones. La siguiente sección (2.2.) tratará de presentar la noción de “pluralismo explicativo”, la cual tiene su origen en variadas fuentes, pero que más fuertemente está originada por la discusión dada en una edición especial de la *Journal of Experimental & Theoretical Artificial Intelligence*, cuyo tópico a tratar era *Pluralism and the future of cognitive science* Pluralism. Esta noción de pluralismo explicativo, más las digresiones en la primera parte, serán la base para que en (2.3.) exponga lo que es el corazón de la tesis, a saber, la tesis de los IC, la cual es fabricada como un medio conceptual fundamentalmente descriptivo que permita dar una cuenta más detallada de las teorías, modelos y las relaciones interteóricas en general, siendo ejemplificada con un caso dentro de las ciencias cognitivas en la sección (2.4.). La sección (2.5.), concretamente el apartado (2.5.iv.), será justamente la aplicación de la tesis de los IC a la polémica simbolismo-conexionismo, para así, finalmente, lograr responder lo que está planteado en la introducción como la tesis de fondo del presente trabajo investigativo, a saber, hasta qué punto existe una competencia entre ambas “teorías”.

2.1. La búsqueda de un terreno común en filosofía de las Ciencias Cognitivas.

En la introducción, de manera provisional postulé que un *terreno común* (TC) se define por tres puntos:

- (i) Un cierto grupo de consensos conceptuales fundamentales sobre hechos básicos, los cuales permiten definir (ii) y (iii).
- (ii) La superación de posturas conceptuales antagónicas a partir de los anteriores consensos.
- (iii) La existencia de cierta metodología común para tratar sobre los hechos básicos del campo dado.

De (i) se encarga en determinarlo, principalmente, la misma ciencia en cuestión. De (ii) se encarga fundamentalmente la filosofía de las ciencias especiales de cualquier disciplina de la cual se esté tratando de dar cuenta su TC. Y de (iii) se encarga, tanto la filosofía de la ciencia particular, como la misma ciencia en cuestión en determinarlo.

Sin embargo, *tal TC ha sido y hasta el día de hoy sigue siendo imposible de establecer en ciencias cognitivas*. No sólo eso, *tal TC pareciera ser en principio una ilusión debido a la complejidad del fenómeno en estudio*. Estos dos puntos de partida no son tan evidentes cuando, por ejemplo, se estudia las funciones cognitivas desde un enfoque específico – por ejemplo si se es parte de una escuela simbólica o una enactivista cerrada – o cuando uno es introducido a la historia de las ciencias cognitivas también desde una perspectiva interpretativa cerrada. Por ello, analizaré en dos partes cada uno de los puntos propuestos para definir un TC en la búsqueda de tratar de encontrar tal TC; primero, trataré el punto (i) y, luego, los puntos combinadamente de (ii) y (iii). Finalmente, esta sección servirá de trampolín para llegar a conjeturar una salida pluralista la cual, en principio, podría hacerse cargo de construir un TC, el cual claramente sería redefinido, más allá que como un consenso sobre generalizaciones, como *una actitud pluralista en la práctica del modelamiento cognitivo* (se delinearé en conclusiones).

(i) ¿Hechos básicos?

Existe la imagen ingenua de que las ciencias tienen la finalidad de describir, explicar y predecir el mundo que directamente se desenvuelve ante nuestros ojos, pero ni siquiera lo que está más cerca de nosotros, los objetos cotidianos como los vemos, son tratados por las ciencias naturales como fenómenos de estudio. Esto se debe a que las ciencias no tratan de lo particular, sino que de lo universal (noción de las ciencias tan antigua como Aristóteles) y la conexión de tales *generalizaciones* contenidas en las teorías con el mundo dista de ser intuitivamente inteligible. Prácticamente, la mayoría de la filosofía de las ciencias (tanto en general como “especial”, i. e. de cada una de las disciplinas por separado) toma parte de la tarea de volver inteligible e investigar esta cuestión no muy obvia para el lego, a saber, que los *hechos básicos son generalizaciones establecidas desde la interioridad de una ciencia*, es decir que los hechos básicos no son aquellas cosas que uno intuitivamente pensaría lo son.

Pero, antes de adentrarse en la salida desde la filosofía de las ciencias, es menester ir un poco más despacio. Antes de hablar de las generalizaciones que pueden establecerse desde la filosofía, es preciso abocarse un instante a las generalizaciones que se han establecido desde las mismas ciencias cognitivas. La primera generalización (o “hecho básico”) que puede servir para cumplir el punto (i) de lo que, conjeturalmente, sería un TC podría ser la generalización de “representación”. No obstante, como se ha

apuntado en (1.2.1.)(A0) y (1.2.1.)(A1) existen dudas en los enfoques alternativos, desde la irrupción a la escena cognitiva del conexionismo, respecto a si realmente son legítimos como “hechos básicos” las representaciones (Thelen & Smith, 1996, 337) y, aún más complicado todavía, la misma definición de “representación” parece no estar bien clara, al punto de que algunos autores han afirmado que “without a properly constrained notion of representation, the debate between representationalists and anti-representationalists is bound to remain a debate for the sake of appearance.” (Haselager et al., 2003, 21) Se ha llegado así a un estado de cosas donde la noción de “representación”, no sólo es discutible como hecho válido, sino que además es discutible si es suficientemente transparente como para ser debatible.

Un simbolista cognitivo puede en este punto apuntar, “pero ello se dio meramente después de la aparición del conexionismo, además de que los propios conexionistas eran reformadores de la noción de representación, pero nunca la negaron.” Sin embargo, aun cuando es cierto que el esfuerzo del grupo PDP estaba dentro del mismo “marco amplio” del cognitivismo, ello no significa que la noción de “representación” haya sido antes del conexionismo un hecho básico claro para todos los investigadores. Esto puede visualizarse cuando se expande la historia de las ciencias cognitivas de modo de incluir la corriente de modelamiento antecesora del simbolismo, la cibernética⁷⁹, donde el conjunto de los hechos básicos aceptados es diferente⁸⁰. Por ejemplo, el modelo TOTE (Miller et al., 1960) es visto en Waterman & Mesarovic (1968) cómo “novedoso” para modelar mecanismos de control biológicos, dado que los propios Miller et al. postulaban que TOTE podría servir para reemplazar el arco reflejo conductista (Waterman & Mesarovic, 1968, 19-21). TOTE, identificado como “modelo tipo programa computacional” – siendo también uno de los más antiguos, aunque no el más original, es paradigmático respecto de los modelos simbólicos cognitivistas (Pylyshyn, 1980, 113) –, aparecía así en el ruedo como un modelo que trataba de utilizar la noción de “información” de un modo alternativo como lo hacía la cibernética, es decir, mediante la postulación de un “plan”, el cual es cualquier proceso jerárquico en el organismo que puede controlar el orden en el cual una secuencia de operaciones se va a realizar y que para el organismo es esencialmente lo mismo que es un programa para una computadora (Miller et al., 1960, 15-16)⁸¹. Luego, *la diferencia entre un modelo generalizado cibernético aplicado a alguna función cognitiva versus el tratamiento que recibía mediante TOTE no era meramente una diferencia en grados de abstracción, sino que radica en la existencia de generalizaciones que no poseería un modelo generalizado cibernético aplicado.*

Vale la pena una pequeña digresión, pues puede parecer poco intuitiva la existencia de competencia entre un “modelo generalizado cibernético aplicado” y TOTE, sin embargo esto es fácilmente visible si se individualizan algunos componentes conceptuales básicos de este último modelo. La cibernética es la

⁷⁹ No es necesario citar a Varela (2002) como voz autorizada sobre la historia de las ciencias cognitivas (hacerlo sería caer en otra interpretación de la historia), sino que se puede acudir a la versión mucho más completa y neutral respecto de la historia de las ciencias cognitivas de Boden (2006), o con el propio, relativamente menos neutral, Gardner (1985, 23-7).

⁸⁰ Esta “expansión de la historia de las ciencias cognitivas” es brevemente detallada en su significado en la sección (2.4.), donde será en el contexto de la aseveración de la pluralidad de las ciencias cognitivas que se mostrará cómo la cibernética ya contiene el germen de la modelización cognitiva simbolista.

⁸¹ La influencia es directa de Newell et al. (1958) al punto que Miller et al. dirán “thus, we are reasonably confident that “program” could be substituted everywhere for “Plan” in the following pages.” (Miller et al., 1960, 16)

confluencia de la teoría de la información y la teoría de control (Novikov, 2016; Mindell, 2006, 4) y, al menos la cibernética más "clásica" que es desarrollada por Norbert Wiener en 1948, constituye una síntesis de esfuerzos que ya existían en el periodo entreguerras, por ejemplo, dentro de la ingeniería de las comunicaciones, el control de artillería antiaérea y el control de vuelo y navegación (Mindell, 2006, 6-7). Por lo tanto, debe ser visualizada no como una teoría, sino que como una *metateoría* que tiene sus orígenes en la ingeniería y que es traspasada al ámbito de la fisiología, neurología, para así llegar a la conducta y la psicología (op. cit., 280-3). En cambio, TOTE es una teoría más acotada que tiene como disciplina objetivo la psicología. Miller et al. consideran que el arco reflejo (definido como una cadena estímulo → receptor → nervio aferente → fibra conectiva → nervio eferente → efector → respuesta (Miller et al., 1960, 22)) ha sido parcialmente exitoso para poner a la psicología por el camino científico, pero que es demasiado elemental y que no se adapta a todas las necesidades del estudioso de la conducta, quien debe readaptar continuamente el arco reflejo a cada conducta que trata de modelar. Los mecanismos neuronales que están involucrados en la acción no pueden ser diagramados simplemente como arcos o cadenas de conexión entre estímulo-respuesta (22-25). Citando a Miller et al:

"A much more complex kind of monitoring, or testing, is involved in reflex action than the classical reflex arc makes any provision for. The only conditions imposed upon the stimulus by the classical chain of elements are the criteria implicit in the thresholds of each element; if the distal stimulus is strong enough to surmount the thresholds all along the arc, then the response must occur (op. cit., 25).

Este "umbral" no sería la única manera en que la entrada puede ser testeada. La acción comienza cuando cierta "incongruencia" entre el estado del organismo y el estado que está siendo testeado existe, y la acción persiste hasta que la "incongruencia" (por ejemplo, un estímulo proximal) cese. El patrón general de la acción refleja será así testear las energías del *input* contra algún criterio interno del organismo hasta que la incongruencia desaparezca. Luego, hay un "feedback" o retroalimentación desde el resultado de la acción a la fase de testeo y se estará frente a un bucle recursivo. El diagrama para esta concepción de la acción, alternativa al arco reflejo, será como la que está a continuación (op. cit., 26):

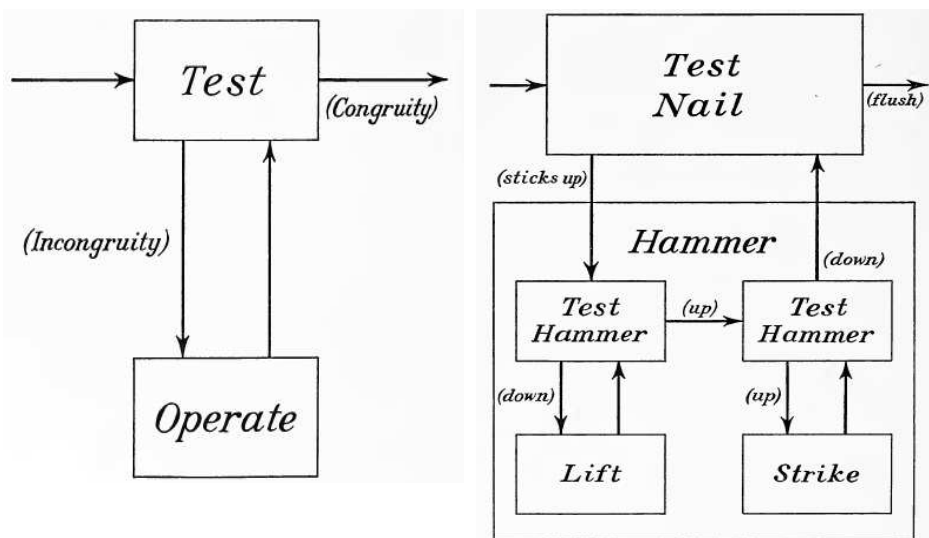


FIGURA 12; A la izquierda una unidad TOTE sin contenido. A la derecha e muestra un caso de descomposición de una tarea que puede ser subdividida en unidades TOTE anidadas cada vez más sencillas. Nótese el paralelo entre esta suposición de Miller et al. y el análisis funcional de Robert Cummins que será expuesto más adelante.

El reflejo no es la unidad que se deba usar como elemento básico de análisis de la conducta, sino que debe ser el lazo de retroalimentación en sí mismo, el cual pasará las fases de testeo-operación-testeo y salida (*exit*), lo cual llevará a Miller et al. a denominarlo "TOTE" (op. cit., 27). Las flechas en este diagrama (**FIGURA 12**) representan tres posibles flujos: *energía, información y "control"*. Si fuera el caso que el flujo corresponde a energía, entonces las flechas estarían representando estructuras nerviosas para mover tal energía, luego sería tal unidad TOTE un simple reflejo. Si se supone que es información (Shannon y Wiener son nombrados en este punto), entonces se estaría correlacionando meramente los eventos a los finales de las flechas, lo que es exactamente que se hace cuando se habla de los modelos S-R (estímulo-respuesta). Luego, la salida de Miller et al. es apelar a algo intangible que denominan "control". Este concepto aparece frecuentemente en las discusiones de las máquinas computadoras, donde el control de las operaciones de máquina pasan de una instrucción a otra, sucesivamente, lo que la lleva a ejecutar la lista de instrucciones que comprende el programa que se le ha dado. Evidentemente esta idea no está limitada a los computadores, como se puede ver en el caso familiar de buscar un tópico particular en el índice analítico de un libro, donde puede decirse que la conducta de búsqueda está bajo el control de la lista de números que contiene el índice⁸² (op. cit., 28).

Por consiguiente, la manera en que TOTE comparte un trasfondo común con la cibernética, *el cual permite decir que ambas teorías entran en competencia, dado que es posible mapear elementos desde la cibernética a TOTE, pero no viceversa*, es mediante la noción de "retroalimentación" y "control". La primera es básica dentro de la literatura cibernética, y se define de modo amplio como *cualquier cadena de transmisión y retorno de la información* (Wiener, 1961, 95-8)⁸³ – de hecho, la unidad TOTE está *explícitamente* basándose en la noción de retroalimentación cibernética (Miller et al., 1960, 26-7). Por otra parte, la noción de "control", en tanto Miller et al. se han inspirado de máquinas computadoras y sus "instrucciones", implica la existencia de cierto lenguaje y, dado que el lenguaje mínimo que debe poseer cualquier computadora es un lenguaje de máquina, entonces debe forzosamente implicar la existencia de alguna arquitectura de computador para tal máquina. Dado que, la arquitectura más sencilla para un sistema digital involucra la existencia de compuertas lógicas de entrada, un reloj y *al menos una memoria binaria implementada sobre interruptores para el almacenamiento de estados* (East, 1990, cap. 1), entonces cualquier arquitectura de una máquina computadora implica el uso de la teoría

⁸² Lamentablemente extenderme en el modelo alargaría demasiado la exposición, pero nótese que el significado de "plan" se matizará en este punto a "una organización jerárquica de TOTEs". Ahora bien, ¿cómo es posible la evolución en el tiempo de estos planes? Miller et al. apuntan que sería no meramente la relación directa con el ambiente, como es el caso con los mamíferos inferiores donde el ambiente mismo es lo que desplegaría la próxima etapa de actividad en éstos, sino que al nosotros tener la capacidad de manipular símbolos, entonces esta capacidad permitiría asignar nombres a nuevas unidades TOTE, lo que facultaría un re-arreglo de símbolos para formar nuevos planes. Cito "We have every reason to believe that man's verbal abilities are very intimately related to his planning abilities." (Miller et al., 1960, 36-8) No deja de ser curioso este apoyo en una "rewiring hypothesis" (Bermudez, 2005, 287-8) susceptible de ser correlacionada con cierto anti-innatismo totalmente contrario a la tesis como LOT, con la cual TOTE es normalmente emparentado.

⁸³ Puede presentarse retroalimentación sin la necesidad de acudir a la noción sofisticada de información. Un sistema determinista como un interruptor para prender la luz, en un contexto donde la pieza está oscura y, además, sin conocimiento de donde está el interruptor, se transforma su búsqueda en un proceso con bucle retroalimentado.

de la información como la de Shannon⁸⁴ (Shannon & Weaver, 1949, 31-5). Siendo este el caso, entonces la noción de “control” también estará siendo fabricada *sobre la de información*. En palabras de Boden sobre TOTE “the core ideas—feedback, internal model, programs, hierarchy, and the flow of control—were drawn from both sides of the emerging cybernetic/GOFAI divide” (Boden, 2006, 338-9).

Ahora bien, siendo inconsciente de la pequeña digresión aquí hecha, en este punto puede nuestro adversario simbólico-cognitivista replicar “es imposible comparar TOTE con un modelo generalizado cibernético aplicado, ya que realmente TOTE está explicando la cognición, mientras que el otro meramente lo describe usando una generalización *incorrecta*”. Esta réplica me lleva a la esencia del problema de fondo que existe con el propio punto (i) para definir TC, a saber, contiene la premisa de la existencia de un *hecho básico privilegiado* que, se presupone, está presente en la base de las ciencias cognitivas – de hecho, en la base de cualquier ciencia. Decir que se “usa la generalización incorrecta”, supone que, en este caso, antes del simbolismo cognitivista, simplemente no existía aun ciencia cognitiva entendida como “la apropiada ciencia de lo mental”, ya que se habría estado usando un hecho básico demasiado vago o *incorrecto* que no debiese ser considerado como un hecho básico.

En el fondo, se hace necesaria una respuesta a la pregunta más trascendental de *¿existe siempre una generalización privilegiada dentro de una ciencia en general?*, ya que una respuesta positiva implica un apoyo directo a la imagen de que un día nacerá una ciencia cognitiva “buena y definitiva” después de estar mal encaminada, mientras que una negativa, no es meramente derrumbar el supuesto de que un TC debe contener hechos básicos – por ende que ni la “representación” ni cualquier otra generalización servirá como contenido de (i) en un TC para ciencias cognitivas –, sino que significa afirmar que *ninguna ciencia contiene alguna generalización ni necesaria ni suficiente que funcione como definitiva de ésta*.

Para responder esta pregunta volveré al principio de esta partecita dentro de la sección (2.1.) sobre *hechos básicos*. Asumamos una postura metafísica realista sin muchos adornos (simplemente asumiendo que el mundo ha estado, está y seguirá estando ahí, aunque nosotros ya no existamos) y partamos nuevamente reflexionando con la cuestión de las generalizaciones y su relación con los fenómenos que tratan, tomando un caso como guía: la gravedad. Me parece intuitivamente falso imaginar que el día justo antes de que Newton postulara la ley de gravitación universal no existía la gravedad para que, precisamente, al día siguiente existiera. Parece bastante claro que existía la gravedad antes y después que Newton la “descubriera” ¿o acaso no la descubrió y más bien la inventó? Esto en realidad es una falsa disyuntiva, porque me parece intuitivamente obvio que las cosas caían antes y después del día que Newton postuló su ley, pero también es bastante obvio que el término “gravedad” se *naturalizó* a medida que la teoría newtoniana fue aceptada posteriormente⁸⁵. Luego, la generalización hecha por

⁸⁴ Teorías que utilizaban la noción de información existían antes de Claude Shannon y Norbert Wiener (también preexistían otras que utilizaban la de “control”), no obstante Shannon crea una teoría que revolucionará el campo, al presentar una teoría de la optimización de la comunicación con aplicabilidad universal a todos los medios de transmisión de información, específicamente mostrando como la información puede ser transmitida sin pérdida, aun en presencia de ruido, mediante la individuación del código correcto y la cantidad adecuada de redundancia.

⁸⁵ Esto es una versión en extremo simplificada de la evolución interna de la teoría y posterior popularización de la ley de gravedad newtoniana. Debe notarse que la teoría de la gravedad dentro de la vida de Newton fue una idea que demoró al menos dos décadas en madurar, y la física newtoniana fue vuelta a exponer comprensivamente recién más de un siglo después con la *Mecánica Celeste* de Laplace (Cohen & Smith, 2004, 6-7).

Newton no tuvo un lugar privilegiado dentro de la física *a priori*, sino que bastante *a posteriori* y, aún más interesante notar es que, aunque es decisiva hoy la gravedad como hecho básico dentro de la física, la misma física como disciplina es capaz de albergar investigaciones que traten de sacarla del grupo de las fuerzas fundamentales de la naturaleza (Verlinde, 2011). Y es posible afirmar sin dudas que, aun cuando sucediese que la gravedad dejara de tener el estatus de fuerza fundamental, y por extensión que dejara de tener la preeminencia como hecho básico que tenía, la identidad de la física como disciplina no se perdería, sino que acaecería sencillamente cierto “cambio de paradigma” o “revolución conceptual”.

En consecuencia, aun cuando en la introducción asumí preliminarmente lo contrario, un TC no puede ser construido ni en ciencias cognitivas ni en cualquier ciencia a partir de la preposición de algún conjunto estable de generalizaciones. Esta presuposición puede hacerse, claro está, como un modo *pedagógico* de ir interiorizándose en una disciplina en un instante dado, pero *tal proceso debe ser diferenciado del que es propiamente el del progreso teórico* de cualquier ciencia. Por consiguiente, esta conclusión no debe tomarse como una declaración contra la posibilidad de determinación de hechos básicos interiormente en alguna ciencia, sino que como una aseveración del carácter *esencialmente dinámico de la determinación de tales hechos básicos de naturaleza inestable*.

(ii) (iii) ¿Superación de posturas antagónicas?, ¿Metodología común?

Un TC sin hechos comunes no puede dar lugar ni a una superación de posturas antagónicas ni al desarrollo de una metodología común, sin embargo, a pesar que esta afirmación es válida a partir de lo que se ha presupuesto es un TC, tomaré una postura caritativa con la definición que di en mi introducción y supondré que un TC puede existir sin la necesidad de haber hechos comunes. Teniendo esto en cuenta, entonces procederé a evaluar los puntos (ii) y (iii) combinadamente.

Las posturas antagónicas que tomé, introductoriamente, eran las del localizacionismo y el holismo en neurociencias, la primera interpretando los datos que se tenían por entonces por conocidos sobre la relación entre conducta y cerebro como apoyando una tesis donde las funciones estaban localizadas en áreas específicas del cerebro, y la segunda negando tal tesis y apoyando una versión opuesta. No obstante, posturas antagónicas como éstas pueden ser encontradas a lo largo y ancho de todas las ciencias, siendo su lugar de debate en un punto intermedio entre el límite y el “afuera” de las investigaciones empíricas, más exactamente en la filosofía de las ciencias especiales.

En el caso de las ciencias cognitivas desde sus comienzos existió preocupación por la búsqueda de un vocabulario o alguna “metateoría” común. Esto es evidente si se expande la historia de las ciencias cognitivas a las conferencias Macy, las cuales, en su última versión de 1953, llegaron a un acuerdo muy general respecto a, por ejemplo, la búsqueda de los mecanismo básicos de la conducta teleológica y el flujo de información que atraviesa desde máquinas, organismos y llega a sistemas completos sociales. En esta búsqueda los conceptos de *retroalimentación*, *bucles* e *información* lo permean todo (Von Foerster, 1955, 69-73). Sin embargo, tanto en el “detalle” (dado que cada disciplina aplicará de manera distinta los conceptos de la metateoría cibernética, al manejar cada una métodos diversos de extracción de datos, además de la naturaleza divergente de cada una), como en otras grandes cuestiones, como por ejemplo el debate de si el sistema nervioso debe ser considerado un sistema digital como se hace con los

computadores, *simplemente no se llegó a consenso*. En una brevísima síntesis de lo que es la crítica de Ralph Gerard a lo que es la visión original de McCulloch-Pitts⁸⁶ sobre el funcionamiento del sistema nervioso, y que se da en el contexto de las discusiones de la séptima conferencia Macy (1950), citaré la crítica de Gerard a la “digitalización” del sistema nervioso central:

My point, in emphasizing the actual functioning of the nervous system in the forbidden continuous region, is that much thinking about the nervous system and much of the theoretical interpretation of memory, learning, and many other things has been based upon its not functioning in this region, upon its working digitally, upon the all-or-none behavior of firing or not firing an impulse. What I am suggesting is that, although it is certainly true that impulses either do or do not fire through large parts of the nervous system, this may not be the critical mechanism in its effective functioning. Just as has been said, in order to operate with a continuum one breaks it down into units with which one can work; but that is just an incidental procedure, which could perhaps be avoided, and might have little relation to the ways the whole functions. So, first, we do have many continuous mechanisms operating in the nervous system and my feeling is, although I confess at once this is not established except by collateral evidence, that some of those continuous mechanisms have coding value and are critical to the functioning of the nervous system. I am further suggesting that, even though we find digital operation in the nervous system, this may not be the essential mechanism accounting for its behavior but may be incidental. (Gerard, 1951, 194)

Si se toma este fenómeno de desacuerdos en todos los niveles – desde lo más específico, a las grandes cuestiones como la antinomia “digital-analógica”⁸⁷ – como la norma, entonces se sigue necesariamente un reconocimiento de la dificultad de conciliar las diferentes posturas sobre cómo modelar lo mental desde el comienzo: *un periodo donde es patente que la propia ciencia cognitiva es indistinguible de su propia filosofía*. Esta “indistinguibilidad” de tareas autoimpuestas por la naciente comunidad de los investigadores, y que puede observarse si se va a las fuentes históricas como he tratado de hacer, es el porqué las interpretaciones sobre lo mental, no solamente guiaron debates entre posturas antagónicas

⁸⁶ En realidad la afirmación de la originalidad del trabajo de MP no está libre de polémica. No es seguro afirmar taxativamente si es Wiener (1961 [1948]), Turing (1950) o McCulloch & Pitts (1943) quienes abogan por primera vez por la “computarización” de lo mental. Sin duda, cada uno puso un énfasis diferente, pero considero que la versión de M&P es la más evidente al *conectar una neurofisiología muy abstracta con funciones lógicas*, las cuales son susceptibles de implementación. En cambio, tanto Turing como Wiener lo tratan como una posibilidad más abstracta sin entrar en detalles sobre el cómo. Algo del trabajo de M&P se desarrollará en la sección (2.4.1.)

⁸⁷ Esta oposición es tratada por Allen Newell (1982, 9) al historiar los tópicos intelectuales que han marcado la evolución de la inteligencia artificial, y la explica a partir de la dicotomía entre las generaciones de computadoras analógicas versus las digitales – las cuales diferían en que, las primeras, representaban cantidades mediante variables continuas, mientras que las segundas representaban cantidades mediante estados discretos – que estaba implícita en la discusión sobre cuál es la clase relevante de procesamiento de información esencial al funcionamiento del sistema nervioso. Sin embargo, si se acude a las discusiones dentro de la presentación de Ralph Gerard, entonces se encontrará con que es muchísimo menos clara la noción de “analógico” o “digital”, al punto que dentro de la misma discusión, primero, existen muchísimas definiciones y salidas conceptuales que se van tanteando y, segundo, no se llega a ningún consenso respecto a la salida adecuada. El largo párrafo citado de Gerard puede ser trasladado a críticas contemporáneas contra la digitalización – “computarización” – del fenómeno cognitivo, de lo que se sigue que el tópico “analógico vs. digital” sencillamente ha mudado sus características, pero tiene cierta continuidad desde las conferencias Macy hasta el día de hoy.

en la “exterioridad” de las ciencias cognitivas, por ejemplo en filosofía, sino que también generaron una *guía diversificada* sobre cuáles eran las metodologías “comunes” aceptables para estudiar el fenómeno dependiendo de las posturas de cada investigador – v. gr. en la dicotomía ya nombrada – y de cuál era la disciplina especialidad de cada uno.

La cantidad de análisis mínimo que sería requerida para tratar esta diversidad de posturas que se develarían de un análisis directo de las diferentes aplicaciones de la metateoría de la cibernética y el naciente simbolismo sería bastante alta – una pizca de tal análisis mínimo fue hecho al exponer, brevemente, algo del modelo TOTE para derivar el cómo entra en competencia con un “modelo generalizado cibernético” – y, dado que *es posible tratar esta diversidad de posturas indirectamente mediante la filosofía de las ciencias cognitivas*, la cual iba tomando más forma a partir de las primeras conferencias Macy y el naciente simbolismo, entonces haré una variación del foco hacia la filosofía de las ciencias cognitivas a partir de la especificación de las *estrategias explicativas comunes*.

Luego, dando un salto de tres décadas respecto a la primera conferencia Macy en 1946, las ciencias cognitivas adoptarían el “análisis funcional” como estrategia explicativa común – al menos éste pareciera ser el acuerdo –, encontrando en Robert Cummins parte de las ideas que darían contenido a la que se perfilaría como la estrategia central cognitivista⁸⁸. La idea de Cummins (1975, 759-760) sobre qué constituye un análisis funcional es capturada en la imagen de la fabricación de cualquier producto industrial mediante una línea de ensamblaje. El proceso de producción se divide en diferentes tareas, siendo cada punto en la línea responsable de llevar a cabo cierta tarea, y es la función de los trabajadores o máquinas llevar a cabo tales tareas. Si la línea como un todo tiene la capacidad de producir el producto, esto será en virtud de que los trabajadores/máquinas son capaces de llevar a cabo las funciones que les corresponden en una cierta manera organizada. Luego, *se explicaría cómo la línea es capaz de producir el producto mediante un análisis de las funciones de sus trabajadores o máquinas*.

Para Cummins los diagramas esquemáticos en electrónica dan otra ilustración de su noción de “función”, ya que cada símbolo representa cualquier objeto que es definido por alguna capacidad, así el diagrama esquemático de un dispositivo complejo constituye un análisis de las capacidades electrónicas del dispositivo como un todo basado en las capacidades de sus componentes. El análisis funcional en biología es, esencialmente, similar al electrónico, es decir, se analiza al organismo a partir de sistemas – circulatorio, digestivo, nervioso, etc. – cada uno teniendo sus capacidades distintivas. A su vez, estas capacidades se descomponen en las capacidades de tejidos y órganos. Esta estrategia se continúa hasta que fisiología pura toma lugar, es decir, hasta que una estrategia no funcional es suficiente, específicamente, la estrategia de subsunción basada en generalizaciones nómicas (op. cit., 758-761).

In nuce el cognitivismo tomará esta noción de análisis funcional, que por aquella época estaba siendo relativamente más elaborada desde la filosofía de las ciencias que se preguntaba por la noción más

⁸⁸ El especialista podrá objetar que no estoy haciendo uso de autores como Putnam (1975 [1967], 435-8), desde donde es posible encontrar más rápidamente la relación que se va estableciendo entre el funcionalismo, en este caso como una postura filosófica de la mente, y las jóvenes ciencias cognitivas. Este recorrido alternativo desde la filosofía de las ciencias lo hago, básicamente, porque mi enfoque desde el comienzo no va en línea con una discusión con la filosofía de la mente, a pesar que desde ella es más fácil derivar las bases cognitivistas de las estrategias comunes aceptables y su relación con la metafísica de lo mental.

exacta de “función” (Godfrey-Smith, 1993), y la complementará con los hechos básicos sobre la cognición que pueden resumirse en la *conservación de la psicología popular* y la *existencia de representaciones mentales ligadas en su contenido conceptual al vocabulario computacional*, además de especificar una estrategia explicativa ligada a la *equivalencia fuerte entre la simulación computacional de un proceso cognitivo y el mismo proceso cognitivo*, que se diferencia de la mera simulación de otros procesos físicos tales como el movimiento de los planetas, el tráfico vehicular o un ciclo económico (Pylyshyn, 1980, 119-120). Tal estrategia explicativa, originalmente cognitivista, tendrá una variación con la aparición del análisis que diferencia entre tres niveles hecho por David Marr (2010 [1982]). Dado que su trabajo es de importancia en ciencias cognitivas, vale la pena detenerse sucintamente en él para evaluar desde su contenido hasta qué punto puede considerarse una estrategia común.

David Marr comienza su obra clásica sobre la visión con una perspectiva general de algunos de los éxitos tempranos en neurociencias, los cuales fueron capaces de identificar algunos mecanismos concretos involucrados en la visión. Por ejemplo, el trabajo de Horace Barlow, *Summation and inhibition in the frog's retina* (1953), donde se demuestra que las células ganglionares en la retina servirían como “detectores de bichos”, servirá como base teórica para el trabajo de Hubel y Wiesel (1962; 1968) y su propio trabajo en 1969, donde se demuestra la capacidad de las células Purkinje en el córtex del cerebelo para aprender patrones motores (Marr, 2010 [1982], 12-4). En general, los esfuerzos investigativos de Barlow sugerían que el mero grabado de las actividades de las células en varias partes del cerebro serían suficientes para comprender cómo el cerebro lleva a cabo varias funciones cognitivas, en palabras de tal investigador:

“A description of that activity of a single nerve cell which is transmitted to and influences other nerve cells and of a nerve cell's response to such influences from other cells, is a complete enough description for functional understanding of the nervous system. There is nothing else "looking at" or controlling this activity which must therefore provide a basis for understanding how the brain controls behavior.” (op. cit., 13)

Desde la perspectiva de Marr, este dogma fue perdiendo progresivamente su fuerza al fallar el hacer más progresos en los 70' y es esta falta de una base teórica para salir de tal estancamiento en las neurociencias lo que lleva a Marr a buscar una forma de no meramente describir el comportamiento de células nerviosas, sino que de explicar tal comportamiento (op. cit., 15). En síntesis, saber cuáles células particulares en el cerebro son sensibles a una información sensorial particular no revelaría cómo estas neuronas contribuyen a una función cognitiva, en este caso cómo contribuirían a la visión; se hace imperioso así un *nivel diferente y novedoso de comprensión*.

Este nivel de comprensión novedoso lo obtendrá a partir de la “metáfora del computador”, es decir, a partir de una visualización del funcionamiento sistema nervioso como el funcionamiento de un sistema computacional de procesamiento de la información (op. cit., 4-5). Esta analogía será descompuesta en profundidad por Marr al proponer que las funciones cognitivas, las cuales son entendidas como idénticas al funcionamiento del sistema nervioso central, pueden ser descritas bajo una análisis que las subdivide en tres niveles, el *computacional*, el *algorítmico-representacional* y el *implementacional*. La forma en que

introduce tales niveles es a partir de una descripción de cómo sería el análisis en tres niveles hecho sobre el ejemplo de una caja registradora (op. cit., 22).

Lo que la “teoría computacional” (aquella teoría cuyo nivel de análisis está restringido al “nivel computacional”) de la caja registradora contendría serían proposiciones como “el orden en que los productos son presentados no afectan el total de la boleta (conmutatividad de la suma)” y otras que contendrían: (i) argumentos separados del *qué* es computado y *porqué*, y (ii) las restricciones a las que estarían afectos los objetos de la computación (op. cit., 22-3). Lo que sería la “teoría representacional y algorítmica” de la caja registradora involucraría dos cosas (i) las representaciones de entrada y salida de las entidades que el proceso computacional manipula, y (ii) un algoritmo mediante el cual las transformaciones de las entradas se llevan a cabo (op. cit., 23). Finalmente, la caja registradora puede ser construida sobre diferentes tecnologías, las cuales llevarán a cabo la “teoría computacional” y la “teoría representacional y algorítmica” especificada. Citando a Marr: “The child who methodically adds two numbers from right to left, carrying a digit when necessary, may be using the same algorithm that is implemented by the wires and transistors of the cash register in the neighborhood supermarket, but the physical realization of the algorithm is quite different in these two cases.” (op. cit., 24)

En síntesis, *los niveles de análisis de Marr permiten explicar una función cognitiva mediante un esfuerzo integrativo de tres formas de teorizar sobre ésta* – una forma “computacional”, otra “representacional-algorítmica” y una última “implementacional” –, y el modo de proceder desde, por ejemplo, el nivel computacional al implementacional o viceversa es lo que en ciencias cognitivas distinguirá normalmente a si a una investigación se la llama “top-down” o “bottom-up” respectivamente. Cabe agregar que, aunque tales tres modo teóricos pueden verse desde cierto ángulo como aumentando la complejidad del fenómeno cognitivo, dado que tales tres niveles tienen una suposición de fondo común – la “metáfora del computador” –, entonces la propuesta de Marr puede verse como un modo de tratar de dar respuesta que simplifique la tarea del cómo integrar los múltiples niveles temporales y espaciales que están involucrados en el funcionamiento cognitivo (Bermúdez, 2014, 122-6)

Ahora bien, las estrategias comunes explicativas del análisis funcional o de los niveles de análisis de Marr ¿puede realmente considerarse “comunes” dentro de las ciencias cognitivas? Lamentablemente, *no hay acuerdo ni sobre el uso del análisis funcional, ni sobre el uso de los niveles de Marr*. En el caso del análisis funcional esto es fácil de visualizar, por ejemplo, cuando se le coteja contra trasfondos enactivistas o dinamicistas; son incompatibles por el simple hecho que el análisis funcional se aplica sobre el individuo, mientras que al agregar el cuerpo y el ambiente, sencillamente, se vuelve incongruente un análisis funcional que se aplique al continuo que se establece entre cerebro y ambiente. Por otra parte, en el caso de los niveles de análisis de Marr sucede algo más complejo que con el análisis funcional⁸⁹, ya que existe una clara incongruencia de interpretaciones respecto a lo que implica la propia noción de niveles de análisis. Por una parte el grupo neurocomputacional interpreta a Marr de modo de que el nivel computacional es constreñido “desde abajo”, mientras que el grupo cognitivista lo interpreta como afirmando la autonomía del nivel más alto respecto a los inferiores – este punto se señalará con algo

⁸⁹ De hecho, el análisis funcional podría ser considerado como compartido entre simbolistas y conexionistas (en general por todos los grupos neurocomputacionales) dado cierto individualismo –o internalismo– (Marraffa & Paternoster, 2012, 28) común en ambos.

más de detalle en la sección (2.5.iv.). Una lectura directa de Marr se presta para ambas exégesis y, desde mi perspectiva, simplemente no hay salida si ya se está sesgado desde antes de hacer la lectura.

No obstante, quizás, estoy mirando en el lado equivocado de la historia y la filosofía de las ciencias cognitivas y deba actualizarme con la filosofía de las ciencias cognitivas. Pues bien, contemporáneamente existe otro intento para dar con alguna clase de estrategia común explicativa en ciencias cognitivas desde la filosofía de las ciencias particulares, a saber, la propuesta del mecanicismo de William Bechtel y Adele Abrahamsen, la cual quizás podría servir como una salida hoy por hoy más aceptada por todos y que valdría la pena evaluar.

La definición de mecanismo es “a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena” (Bechtel & Abrahamsen, 2005, 423), a lo cual se agrega que las partes componentes del mecanismo son aquellas que figuran en la generación del fenómeno, que cada operación involucra al menos una parte componente, que las operaciones pueden ser organizadas mediante una secuencia temporal – aunque los mecanismos biológicos tienden a exhibir organizaciones más complejas – y que los mecanismos involucran múltiples niveles organizativos (op. cit., 424).

La explicación mediante mecanismos contrasta con la explicación nomológica en cuanto no busca leyes cubrientes, sino que trata de identificar partes componentes y operaciones de un mecanismo. Incluso un investigador que no observa los componentes, sino que los infiere, está construyéndolos como partes y operaciones de un mecanismo existente de hecho. El proceso de descubrimiento de un mecanismo involucra el proceso de *descomposición*, existiendo dos modos en que los investigadores lo aplican: estructuralmente o funcionalmente, dependiendo si, o bien ellos se enfocan en las partes componentes, o bien en las operaciones de los componentes. Aunque ambas clases de descomposición pueden ser llevadas a cabo simultáneamente para un mecanismo dado por un solo investigador, es usual que diferentes investigadores en diferentes campos sean los primeros en proponer las dos formas de descomposición. La integración de éstas en una explicación completa, se da posteriormente por lo común, y a menudo involucra un período de coevolución, durante el cual las descomposiciones son modificadas repetidamente de modo de volverlas progresivamente más compatibles (op. cit., 432-433).

Las tareas de descomposición de los mecanismos en componentes y operaciones, y la tarea de unir las partes a las operaciones, puede abarcar una variedad diversa de procedimientos experimentales. Estos incluyen, por ejemplo, inhibir un componente de forma de observar los efectos sobre todo el funcionamiento del mecanismo, o bien el grabado de las condiciones internas del mecanismo cuando está operativo bajo diferentes condiciones. Es bastante claro que cuando la meta es el descubrimiento de un mecanismo, y no la de una ley cubriente, entonces el proceso de llevar a cabo el descubrimiento es bastante diferente de la simple articulación de leyes. Buena parte del descubrimiento no sólo está enfocado en las partes componentes y las operaciones del mecanismo, sino que buena parte del trabajo involucra encontrar la organización del mecanismo (op. cit., 435).

La propuesta de Bechtel y Abrahamsen tiene al menos dos ventajas desde la filosofía de las ciencias: el problema de la subdeterminación se presenta con menor fuerza que cuando se testean leyes. Cuando un

investigador se dispone a probar el modelo de un mecanismo, el foco típicamente no está en el mecanismo como un todo, sino que en componentes específicos del mecanismo. Cuando las expectativas fallan, entonces estos resultados apuntan directamente sobre qué revisiones son necesarias, es decir, los resultados del testeo de los mecanismos son frecuentemente mucho más informativos que los de las leyes cubrientes (op.cit, 436).

La otra ventaja del mecanicismo es que provee una mejor caracterización de cómo las explicaciones son generalizadas en muchas ciencias. Las leyes son generalizadas por estar universalmente cuantificadas y el dominio de aplicabilidad está especificado por las condiciones en sus antecedentes, ergo en esta clase de explicación ningún caso ejemplifica la ley mejor que otro. Al contrario, en las investigaciones de los mecanismos, los científicos se enfocan en un ejemplar específico cuando están al principio de sus explicaciones; tal ejemplar será el *sistema modelo* y será elegido por razones que varían (op.cit, 438).

Desgraciadamente, tampoco la propuesta de Bechtel y Abrahamsen ha estado libre de polémica. Stepp et al. (2011) consideran que tal planteo filosófico de las ciencias *excluye buena parte de la investigación actual publicada en ciencias cognitivas*, en particular, al igual que han hecho otros investigadores cognitivos al suponer que la cognición involucra transformaciones de representaciones por definición, deja fuera toda versión de la cognición que vaya contra el representacionalismo y el procesamiento de información, es decir, *deja fuera al dinamicismo no representacionista*. En este último la cognición es un proceso de mantenimiento continuo y robusto de un sistema animal-ambiente que se alcanza mediante la coordinación cercana entre percepción y acción, y va más allá de las áreas “normalmente” asociadas al dinamicismo, como por ejemplo la relación percepción-acción, ya que implicaría las áreas como aprendizaje, razonamiento y lenguaje, es decir tipo “alta cognición”. (Stepp et al., 2011, 431-2)

Para Stepp et al. los modelos dinamicistas cognitivos son genuinamente explicativos, a pesar que el mecanicismo como filosofía de las ciencias normativamente los excluya como meramente descriptivos. Esta exclusión se debe a que las explicaciones dinámicas no proponen mecanismos causales que muestren cómo se produce el fenómeno en cuestión. Más bien, éstos muestran que el cambio en el tiempo de cierto conjunto de magnitudes en el mundo puede ser capturado por un conjunto de ecuaciones diferenciales tipo leyes y, en cierto sentido, las explicaciones dinámicas pueden asemejarse a explicaciones mediante leyes cubrientes. Las explicaciones dinamicistas revelan que un fenómeno podría ser predicho bajo condiciones locales y algunos principios generales tipo leyes, inclusive tales explicaciones tendrían soporte contrafáctico (Stepp et al., 2011, 432). Estas afirmaciones, sin ir más lejos y sin entrar en detalles, Stepp et al. las apoyan en el desarrollo de una explicación dinamicista alternativa para dar cuenta de sistemas circadianos sin la necesidad de apelar ni a mecanismos ni a un sistema representacional, tal como hacen Bechtel y Abrahamsen en *Decomposing, recomposing, and situating circadian mechanisms: Three tasks in developing mechanistic explanations* (2009).

Entonces, si es imposible construir un TC para ciencias cognitivas, ¿qué nos queda?

Para demostrar que el análisis hasta aquí hecho sobre trechos de la historia de las ciencias cognitivas y su filosofía, que se ha ido sofisticando a través del tiempo, es erróneo, solamente bastaría traer a colación algún contraejemplo claro de TC en algún punto de las ciencias cognitivas. Es decir, como se ha postulado en la introducción, ser capaz de *encontrar* un TC sin necesidad de *construirlo*. Sin embargo,

considero que por lo dicho aquello es, tanto empíricamente, como en principio imposible. Por ende, la primera cuestión que surge, por lo menos para el filósofo que se interesa por universalizaciones es ¿qué nos queda hacer si aceptamos la premisa de la *imposibilidad de encontrar estrategias explicativas comunes en ciencias cognitivas*? ¿Sencillamente dedicarnos a hacer “recopilaciones de metodologías”? ¿O dedicarse meramente a hacer “cartografías” de las ciencias cognitivas como la Dennett (1998 [1986]) o la de Varela et al. (1991) – presentada en el curso de la primera parte?

Considero que estas salidas – recopilaciones o cartografías – no serían deshonrosas, más bien al contrario, me parece un buen punto de partida dedicarse a la *recolección esencialmente descriptiva* de, por así decir, “situaciones histórico-filosóficas del cambio conceptual en ciencias cognitivas”, las cuales, con algo de suerte, puedan mostrar cierto patrón que pueda llevar, recién en aquel hipotético punto, a conjeturar sobre la posibilidad real de *normar* respecto al desarrollo de las ciencias cognitivas. En general, considero que esta actitud no es una salida deshonrosa ni para el filósofo especializado en ciencias cognitivas, ni para el que esté especializado en otra ciencia.

No obstante, lo desarrollado hasta acá ha sido guiado por una noción de TC quizás demasiado estrecha, la cual pueda ser reformulada de modo que sea capaz de ser satisfecha por las ciencias cognitivas *reales*, y no por una definición restrictiva de éstas. Por ejemplo, una noción de “terreno común” que no requiera hechos básicos *antes* de la investigación – generalizaciones construidas a priori –, sino que sea totalmente aceptable que todos los hechos funcionen meramente como hipótesis de trabajo o, como he tratado de implicar a partir de la digresión 2 en la sección (1.3.3.), *una interpretación no constreñida ontológicamente de los fenómenos*. Esta estrategia de flexibilidad también puede tomarse con respecto a la búsqueda de “metodologías comunes” y de las “posturas antagónicas”.

Empero, para transformar la noción de TC no basta meramente postularlo, en general requiere cierto trabajo filosófico que dé sentido más claro el *desde dónde* es posible afirmar cuestiones como “*ninguna ciencia contiene alguna generalización ni necesaria ni suficiente que funcione como definitoria de ésta*” o la misma aseveración de la “*imposibilidad de encontrar estrategias explicativas comunes en ciencias cognitivas*”. La clave sobre cuál es el “desde dónde” viene dada por el trabajo de Serban (2014), para quien el mecanicismo en filosofía de las ciencias cognitivas tiene el siguiente problema de fondo:

“Although mechanists acknowledge that models/theories developed at different levels of analysis are governed by different epistemic and ontic principles, they also imply that only concrete how-actually mechanistic models are appropriate candidates for cognitive explanations. This contravenes the alleged *pluralism* entailed by the picture of multilevel mechanistic explanations of cognitive capacities⁹⁰. A genuine pluralist position would attempt to characterize the structure of explanatory models of cognitive capacities in terms that are metaphysically as neutral as possible.” (Serban, 2014, 71).

Aquí se ha tomado el “desde dónde” como *pluralismo*, pero ¿qué tipo de postura filosófica es aquella?

⁹⁰ Maria Serban cita aquí los trabajos de Carl Craver, *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience* (2007) y Carl Craver y Gualtiero Piccinini, *Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches* (2011). Dejo en esta nota tales títulos para el lector interesado en profundizar en la corriente “neo-mecanicista” de la cual el mecanicismo de Bechtel y Abrahamsen es parte.

2.2. Presentación del Pluralismo Explicativo

El “pluralismo”, entendido como *actitud filosófica* puede ser documentado como presente en la física de la primera mitad del siglo XX. Por ejemplo, Bokulich (2008) presenta el pluralismo teórico en el contexto del problema de las relaciones interteóricas de la mecánica clásica y la cuántica, *grosso modo*, que las dos teorías pueden ser tomadas como relacionándose reductivamente, es decir, una teoría puede ser derivada de otra, o pluralistamente, es decir, donde cada teoría tiene su dominio distintivo de leyes, entidades y conceptos, sin posibilidad de reducción (Bokulich, 2008, 1). La filósofa tratará de superar este antagonismo mediante lo que denominará “interestructuralismo”, postura que tomará un poco de cada bando filosófico (Op.cit, 3).

Más allá de adentrarse en la obra, o en el “interestructuralismo” de Bokulich, vale la pena profundizar un poco en cómo la autora define “pluralismo teórico”. El pluralismo teórico sería una postura que trata de evitar la expansión del reduccionismo en las ciencias y que sería la postura de fondo que permite defender la existencia de ciencias autónomas (Op.cit, 5). Dados estos rasgos, el pluralismo teórico puede ser pensado como una familia de enfoques. Como una tesis negativa, *el pluralismo niega que el mundo pueda ser explicado por un conjunto unificado de principios fundamentales o leyes*. La pluralidad de teorías científicas describiendo diferentes dominios no sería, ni una característica temporal de las ciencias, ni una característica permanente que sea mera consecuencia de nuestras limitaciones epistémicas como humanos. Más bien, *la naturaleza en algún sentido demanda en sí misma esta pluralidad de descripciones*, luego todo enfoque que busca reducir o eliminar esta pluralidad está metodológicamente mal guiada y resultará en una mala representación de la naturaleza (Op.cit, 8).

La tesis de la inconmensurabilidad – de Kuhn y Feyerabend, aunque cada uno con diferentes sentidos – *habría jugado históricamente un rol central en el movimiento desde el reduccionismo al pluralismo teórico*. Autores posteriores, por ejemplo John Dupré, llevarán esta postura pluralista a lo que se denominará “realismo promiscuo”, el cual, sintéticamente, postula que el problema no es que las taxonomías científicas fallen en capturar las clases naturales reales, sino que todas ellas lo hacen y que estas taxonomías no son únicas (Op.cit, 9-10). Vale la pena apuntar, que el pluralismo también habría sido defendido en el contexto de la psicología. Bokulich nombra el artículo seminal de Fodor (1974), donde se argumenta que la psicología no puede ser reducida a las neurociencias, atribuyéndose esta irreductibilidad a una manera de ser del mundo, más que a nuestras limitaciones epistémicas. Luego, la psicología es y siempre será una ciencia autónoma sin relación a las neurociencias (Op.cit, 11).

Ahora bien, más allá de si el pluralismo teórico debe considerarse como postura filosófica anterior al autonomismo o como implicando una postura metafísica realista, como deja ver Bokulich (2008), mi interés en la obra de la autora se centra en que a partir de esta definición de pluralismo es capaz de demostrar que en física Werner Heisenberg habría mantenido una actitud pluralista frente a la relación entre mecánica newtoniana y mecánica cuántica. El argumento de Bokulich se basa en una revisión de la propia filosofía de las ciencias de Heisenberg, la cual estaría desperdigada en diferentes obras. Muy resumidamente, el argumento partiría revisando la, fuertemente original, visión de filosofía de las ciencias del afamado físico, la cual se basaría en la noción de “teoría cerrada”, la cual se definiría como un sistema apretadamente tejido de axiomas, definiciones y leyes que proveen una descripción perfectamente acertada y definitiva de un dominio acotado de un fenómeno (Op.cit, 29).

Una “teoría cerrada” se distinguiría de una teoría fenomenológica dentro de la filosofía de Heisenberg en que la primera sería una descripción verdadera de lo que el mundo es, mientras que la segunda no. Luego, Bokulich apunta que Heisenberg no considera ni la mecánica cuántica ni la mecánica newtoniana como teorías fenomenológicas, esto es, adopta una interpretación realista de ambas. El sistema de definiciones y axiomas de la mecánica newtoniana describiría una estructura eterna de la naturaleza. Por consiguiente, *la única manera de interpretar este realismo de Heisenberg sobre ambas mecánicas sería una salida pluralista*, la cual sería apoyada por su idea de diferentes regiones distintas de la realidad correspondientes a cada una de sus “teorías cerradas”. Más exactamente, entonces, la postura del afamado físico podría llamarse algo así como “pluralismo metafísico nomológico” (Op.cit, 38-41).

La segunda mitad del siglo XX, con la embestida historicista de Thomas Kuhn (1970) y anarquista epistemológica de Paul Feyerabend (1975), dará lugar a que el pluralismo *como concepto separadamente postulado y estudiado* en filosofía de las ciencias, ligado, como correctamente apunta Bokulich (2008, 8), al tópico de la “desunidad de la ciencia”. En su discurso presidencial de la *Philosophy of Science Association*, Patrick Suppes emitió un manifiesto a favor del pluralismo (Suppes, 1978). En este postuló que el tiempo de defender a la ciencia contra la metafísica ya habría pasado, siendo esta defensa la causa original por el movimiento que buscaba la unidad de la ciencia (Kellert et al., 2006, vii). El ataque contra la posibilidad de unificación por parte de Suppes tiene un puñado de argumentos bien definidos:

- Primero, no hay posibilidad de encontrar un lenguaje común que englobe todas las ciencias. Esto se puede constatar en el hecho de la especialización y el crecimiento de los términos técnicos de las disciplinas, las cuales tienden a la divergencia más que a la convergencia. De hecho, las ciencias nunca han atravesado un periodo de simplificación del lenguaje, como por ejemplo puede constatar en la más antigua de las ciencias cuantitativas, la astronomía (Suppes, 1978, 5-6).
- Segundo, no tiene sentido una unificación apelando a cierto “substrato común” al cual todo se puede reducir. Aquí se apela a una analogía: así como sucede que dos generaciones alejadas de computadoras pueden llevar a cabo las mismas tareas, pero bajo diferentes implementaciones materiales – por ejemplo, válvulas termoiónicas versus semiconductores –, puede suceder lo mismo con las ciencias, es decir, que más allá de la igualdad en la superficie conceptual, tal reducción sea teóricamente vacía. Yendo a un caso concreto, se aduce que es universalmente aceptado la imposibilidad de tratar de resolver los problemas de la química cuántica apelando a las leyes fundamentales de la mecánica cuántica. En palabras de Suppes, “quantum chemistry, in spite of its proximity to quantum mechanics, is and will remain an essentially autonomous discipline (...) reduction is not only practically impossible but theoretically so as well.” (op. cit., 7-8).
- Tercero, no existe ninguna unidad en los métodos de las ciencias, más allá de lo trivialmente obvio, tal como el uso de matemática elemental – es necesario apuntar que es “elemental” desde la perspectiva de las mismas matemáticas. Esto es sencillo de constatar inter-disciplinariamente, pero incluso yendo más allá, intra-disciplinariamente para los expertos el detalle de los métodos experimentales de una subdisciplina se pueden llegar a volver ininteligibles. Por ejemplo, físicos trabajando en la física de estado sólido, pueden no comprender el trabajo de aquellos que trabajan en teoría de cuerdas. E inclusive en disciplinas menos desarrolladas, como la psicología, lo mismo llega a ocurrir a los expertos (op. cit., 8-9).

- Cuarto, no hay medida física posible de ser tomada con certidumbre cabal. El desarrollo de la mecánica cuántica en el siglo XX llevó a los físicos renuenteemente, pero de modo conclusivo, a reconocer que no tiene sentido afirmar que cualquier cantidad física puede ser medida con un grado arbitrario de precisión en conjunción con la medición simultánea de otra variable física. Este reconocimiento de tal inhabilidad para llevar a cabo mediciones exactas no se debió a inadecuaciones técnicas, sino que está ínsita en la misma teoría física (op. cit., 10-11).
- Quinto, y análogamente a la conclusión que se llegó en mecánica cuántica, puede verse el fracaso de en los intentos de buscar completitud en aritmética con el trabajo de Gödel o de la incompletitud de la teoría de conjuntos, la cual puede ser construida utilizando diversos axiomas independientemente – v. gr. la hipótesis del continuo es independiente al axioma de elección (op. cit., 11-12).

La conclusión a la que llega Suppes la expone mediante una imagen; la ciencia no es una serie convergente, cuyo límite tiende a una teoría final, sino que es divergente. Es una empresa de perpetua solución de problemas, donde no existe área de la experiencia que esté totalmente asentada en un conjunto de verdades básicas, más bien, *estaríamos continuamente confrontados a nuevos problemas y situaciones, los cuales trataríamos con un popurrí de métodos científicos, técnicas y conceptos*. En esta visión de la ciencia la verdad objetiva no habría desaparecido directamente, pero lo que nosotros llamaríamos una “visión global o cosmológica de la verdad” sería mirada con escepticismo (op. cit., 14-5).

Ahora bien, las ciencias cognitivas tienen un periodo de aparición y crecimiento justamente en el mismo periodo en que las filosofías de las ciencias se iban alejando de la unidad e iban *tendiendo a la dispersión* (Ramírez, 2005). Esto es importante notar, ya que un hecho que no ha sido plenamente estudiado en sus últimos detalles es hasta qué punto *desde sus comienzos fue de facto pluralista las ciencias cognitivas*⁹¹ y, por tanto, hasta qué punto tal hecho determina que *la mejor forma de enfoque epistemológico de sus investigadores sea ceñirse al pluralismo*, al contrario de otras disciplinas donde su desarrollo sería más bien de nacimientos de teorías que logran asentarse más fácilmente, lo que significa que una actitud de “sesgo ontológico” o “privilegio ontológico” – como lo denominarán Jilk et al. (2008a) – es más fácilmente adherible sin grandes consecuencias en la investigación⁹².

Esta “sincronía” entre la dispersión de la filosofía de las ciencias y el desarrollo pluralista de las ciencias cognitivas – un periodo amplio que va desde los 50’ a los 80’ – no necesariamente fue de la mano con una actitud epistemológica pluralista en la mayoría de los investigadores. Al contrario, buena parte de la polémica simbolismo-conexionismo, por ejemplo, se fundamenta en una actitud donde, como he detallado en la digresión 1 (sección (1.3.2.)), se da por sentado que existe cierto privilegio sobre la causalidad ínsito en el nivel cognitivo (o, lo que es equivalente en la práctica, el computacional) que por sí mismo daría cuenta del porqué una explicación cognitivista es “realmente” una “explicación”, mientras que la conexionista no lo sería, quedando relegada a ser “mera implementación”. No obstante, a lo largo de este mismo periodo existieron investigadores que, siguiendo la idea de Suppes, ante la complejidad

⁹¹ Esto se delinearé brevemente en la sección (2.4.) y con algo de más detenimiento en (2.5.i.). Sin embargo, tales desarrollos en las secciones señaladas no corresponden a lo debiese ser a un trabajo exhaustivo exclusivo en el tópico, más bien deben ser leídas como una entrada a lo que podría ser un estudio sobre tal pluralidad.

⁹² Este punto será más fácilmente visible cuando introduzca la esquematización mediante fases de los “instrumentos conceptuales” – sección (2.3.2.3.).

de dar cuenta de la cognición, entonces acudieron a “un popurrí de métodos científicos, técnicas y conceptos” en sus propias líneas investigativas. Auxiliándome en lo que he desarrollado en la primera parte de la tesis, es posible encontrar esta actitud pluralista en ciencias cognitivas desde mediados de los 80’ en los trabajos de Marvin Minsky (en su modelo de la sociedad de la mente), en el propio Smolensky al postular al subsimbolismo como mostrando cómo lo “rígido” de la cognición puede emerger de lo “flexible”, también en los modelos híbridos tipo DYNASTY o SAARCS y, además, influyendo a filósofos como Andy Clark quien despliega una postura ecumenista, la cual podría ser considerada *cuasi pluralista*.

¿Por qué considerar al ecumenismo a la Clark como “cuasi” pluralista? Básicamente por el representacionalismo mínimo que se cuela en su corpus teórico, es decir, al aceptarse siquiera la necesidad de tomar la representación como un elemento válido para ser *reformado* o *transformado*. Así pues, tácitamente, se descarta la posibilidad de enfoques o marcos completos que no requieran tal noción, como sería el caso del dinamicismo no representacionalista – brevemente apuntado en la sección anterior como crítico al mecanicismo en filosofía de las ciencias contemporáneas – el cual, hay que decirlo, definitivamente ha llegado para quedarse e inclusive para aspirar a la hegemonía en ciencias cognitivas (Spencer et al., 2011) y dar cuenta de aquellas funciones que eran el último bastión simbolista (Mix et al., 2009). Este *conservadurismo representacional* sería lo que impediría que sea considerado pluralista el ecumenismo. Sin embargo, es totalmente lícito mantenerse ecumenista si se considera la existencia de problemas “ávidos” o “hambrientos” de representación (Clark & Toribio, 1994), los cuales serían *en principio* imposibles de ser solucionados sin la noción de representación. Considero necesario reconsiderar esta cuestión – señalado en el contexto de la sección (1.2.1.) (A1) – y dar por válidas las objeciones contra esta clase de problemas que son expuestas en Degenaar & Myin (2014). Estas son bastante disuasivas y simples en su formulación, por tanto vale la pena detenerse en ellas.

El primer tipo de problema hambriento de representación sería aquel donde la cognición involucra *lo ausente*, es decir “reasoning about absent, non-existent, or counterfactual states of affairs” (Clark & Toribio, 1994, 419). Dado que es posible entender razonamiento en un sentido amplio, puede parafrasearse “razonamiento sobre lo ausente” por *mantener seguimiento de* o *anticiparse conductualmente a algo no presente*.

Considérese el siguiente ejemplo: alguien está viviendo en una casa con una cocina que está separada del comedor por un mesón (una cocina americana) y para llegar al otro lado debe caminar alrededor del mesón. Supóngase que se remueve el mesón y que la persona no sabe de este cambio, y además, por ejemplo, se mantienen las luces apagadas para que no note el cambio del ambiente. La persona, a pesar de la ausencia de una base material para la conducta de caminar alrededor del mesón, lo continuará haciendo. A lo largo del tiempo un patrón conductual ha emergido y esto puede ser explicado mediante representaciones o sin ellas. En este último caso, la persona responde a una situación particular sin representársela, mientras que la “tesis de la ausencia” de Clark y Toribio implicaría que se requiere una representación de “lo ausente” para dar cuenta de esta conducta. En efecto, remover el mesón no cambia la naturaleza de las capacidades de la persona, si eran representacionales antes lo continuarán siendo, mientras que si no lo eran continuarán no siendo representacionales. *La presencia o ausencia de un estímulo ambiental no conlleva un cambio fundamental en la naturaleza del comportamiento, por tanto no parece suficiente para hacerla fundada en lo representacional* (Degenaar & Myin, 2014, 3343)

Sin embargo, se puede ir más lejos, donde por “razonamiento sobre lo ausente” se considere la habilidad de imaginar estados de cosas ausentes, en particular, como se da con la imaginería visual. Supóngase que uno mantiene una postura donde la imaginería visual sea percepción virtual: cuando uno imagina algún objeto o situación, uno estaría en las mismas condiciones del tipo de estar nuevamente confrontado con el objeto o la situación. En su forma más simple, imaginar sería “re-enactuar” las condiciones de la percepción original. Si uno es representacionista sobre la percepción, entonces la imaginería entendida como percepción virtual, indubitablemente involucraría representaciones también. Si, sin embargo, uno combina la noción de percepción virtual con una versión de la percepción no representacional, entonces se tendrá una visión no representacional de la imaginería visual. Por consiguiente, igual que con el mesón, la ausencia de estímulo no posibilita *ipso facto* una interpretación representacional de cierta conducta o capacidad cognitiva (Degenaar & Myin, 2014, 3343-4).

El segundo tipo de problema ávido de representación sería aquel donde la cognición necesariamente deba contener un ítem interno, patrón o proceso, cuyo contenido corresponda a una propiedad abstracta que *aúne cierta disparidad de estímulos*, donde tales estímulos tendrían diferentes manifestaciones físicas con poco en común entre sí.

Tácitamente, en esta clase de problema estaría el supuesto que si diferentes estímulos tuvieran propiedades físicas superficiales comunes, entonces éstas por sí mismas bastarían para que todos los estímulos sean reconocidos no representacionalmente, pero dado que no tienen nada en común, entonces cierto correlato interno debe ser construido mediante algún proceso interno, de modo de dar cuenta de la propiedad abstracta que los conecta. La idea fundamental es que la convergencia de muchos estímulos diferentes sobre uno neuronal con integridad física es la clave para otorgarle su estatus representacional, sin embargo, Degenaar y Myin preguntan ¿por qué la convergencia sobre un elemento interno con integridad física le confiere el estatus representacional? Es fácil imaginar, o constatar naturalmente, muchos procesos en donde causas con diferentes manifestaciones físicas convergen sobre un mismo efecto. Por tanto, no se sigue que de un efecto común se represente sus causas. Considérese un ejemplo simple: el movimiento de una limadura de hierro puede producirse por un soplo, moviéndola con un dedo o atrayéndola con un imán. Ergo, no hay razón para suponer que el movimiento de la limadura representa su causa (Degenaar & Myin, 2014, 3346-7).

Pues bien, si se toman las objeciones de Degenaar y Myin como válidas, entonces *el conservadurismo representacional sería injustificado y el ecumenismo sería no legítimamente pluralista*. No obstante, si el pluralismo dentro de las ciencias cognitivas – en general de las ciencias – no es “ecumenismo” a la Clark, entonces ¿qué es? Hasta aquí he descrito algunas ideas bastante emparentadas entre sí, sin aun desarrollar explícitamente qué entenderé por “pluralismo explicativo”, moviéndome sucesivamente a través de tópicos conectados tales como el “pluralismo teórico” de Bokulich, la existencia de “pluralidad” en las ciencias (tanto cognitivas como en el conjunto total), la existencia de cierta “actitud pluralista” que puede venir de la mano con una pluralidad⁹³, y he llegado hasta el punto del ecumenismo tipo Clark, entonces ¿qué pieza falta por agregar?.

⁹³ Conjeturo como plausible afirmar como una especie de “ley empírica” dentro del campo de filosofía de las ciencias que cierta “conciencia de pluralidad” se da únicamente cuando existe competencia, lo que implica que *la*

Considero que lo último que cabe agregar es una revisión de los artículos centrales que inspiran esta tesis, a saber, aquellos contenidos en *Pluralism and the future of Cognitive Science* y que serán la última pieza para elaborar qué entenderé por “pluralismo explicativo”. A continuación expondré resumidamente los tres artículos “target” de la mentada edición, pero no me detendré a exponer las críticas y sus respuestas que también contiene la misma revista.

The possibility of a pluralist cognitive science, de Rick Dale.

La meta del artículo de Dale (2008) es considerar la posibilidad de un acercamiento pluralista en ciencias cognitivas, acercamiento que implica abrazar la diversidad de las teorías identificándolas por su utilidad única en relación con subdominios del fenómeno cognitivo. Esta posibilidad de pluralismo puede parecer obvia considerando que no sólo es complejo el fenómeno mental, sino que las ciencias cognitivas ya son una “súper-disciplina” al abarcar varias disciplinas diferentes. Esto no ha impedido que en la práctica muchos científicos cognitivos investiguen sin considerarse pluralistas. Ello se debe a que es posible reconocer la pluralidad de esquemas explicativos en ciencias sin adscribir al pluralismo (Dale, 2008, 156).

Lo primero que hará Dale es considerar una serie de hechos bastante claros que motivarían el adscribir una perspectiva pluralista en ciencia cognitiva. Estas observaciones proveen dos modos de tomar el pluralismo seriamente: un modo sobre la naturaleza específica de la ciencia cognitiva y otro a partir de cuestiones generales epistemológicas en ciencias. Estas observaciones, más que funcionar como argumento “probando” la postura filosófica, motivarían *el potencial de adscribir al pluralismo*.

La primera observación sobre la ciencia cognitiva es que aún es una ciencia joven, tomándose el año 1956 como el nacimiento de ella. De hecho, aun cuando el crecimiento de la ciencia hoy en día es mucho más rápido dada la tecnología de la información que permite la diseminación más rápida de la literatura especializada, ello no significa que la juventud de la disciplina haya acabado (Op.cit, 158).

La segunda observación es que la ciencia cognitiva es diversa, fácilmente evidenciándose en textos bien conocidos en el campo. Por ejemplo, la famosa enciclopedia de ciencia cognitiva de Wilson y Keil (1999) tiene una impresionante tabla de contenidos, que, aunque es meramente de “tópicos seleccionados”, abarca desde procesos de bajo nivel, como el olfato, a procesos como la metacognición (Op.cit, 158-9).

La tercera observación es que esta juventud y diversidad contribuyen parcialmente a una tercera característica de la ciencia cognitiva: “la ciencia cognitiva no es física-teórica”. Con esto Dale quiere decir que hoy ni los sistemas simbólicos, ni el PDP, ni los sistemas dinámicos son suficientemente comprensivos para unificar todo el espectro de fenómenos cognitivos. Cada una de ellas tiene su propia región de interés, siendo en los límites donde ocurren disputas entre ellas (Op.cit, 159).

En síntesis, los tres hechos de las ciencias cognitivas se interrelacionarían del siguiente modo: al ser joven, entonces mucho queda por ser descubierto. Al ser diversa, estos descubrimientos inevitablemente tomarán lugar en variedad de esquinas de rango temporal y espacial diverso, a

existencia de teorías en competencia es requisito indispensable para la llegar a la idea de “pluralismo”. Sin embargo, llevar a cabo la tarea de defender esta conjetura implicaría una investigación de una profundidad muchísimo mayor que lo que estoy haciendo en la presente tesis.

metodologías y metas explicativas distintas. Finalmente, estos descubrimientos no vendrían de la mano con una colección de leyes matemáticas unificadas. Estas tres observaciones ya pueden sugerir por sí mismas que la pluralidad es distintiva de las ciencias cognitivas. *El pluralismo, sería un modo de proveer ayuda para la creación de nuevos caminos de descubrimiento e integración* (Op.cit, 160).

Dale agrega a estas tres cuestiones tres más, las cuales no se relacionan exclusivamente con las ciencias cognitivas, sino que con el mismo ser humano y la ciencia. El primer hecho es que el conocimiento humano es inherentemente limitado, los propios hallazgos psicológicos revelan profundas limitaciones en nuestra habilidad para razonar sobre y representarnos el mundo. Dadas estas limitaciones, pareciera que las ciencias cognitivas debiesen contentarse con respuestas parciales sobre los subdominios del cerebro, el comportamiento y el ambiente (Op.cit, 160-1).

Esta limitación es aún más importante cuando se considera complementada con un segundo hecho; los seres humanos son, al igual que muchos objetos de interés científico, extraordinariamente complejos. Aun cuando es posible encontrar descripciones simplistas sobre las conductas, estos patrones requieren involucrar muchas dimensiones explicativas (Op.cit, 161). El tercer punto es que toda teoría está ceñida por su enfoque de aplicación. Buena parte de la filosofía e historia de las ciencias consiste en determinar cómo los científicos escogen el enfoque de su teoría, la forma de tal enfoque y las consecuencias de tales para el desarrollo científico. Dale se basa en el trabajo de Nancy Cartwright, quien llama “máquinas nomológicas” a aquellos sistemas para confiablemente generar ciertas salidas. Las leyes y otras regularidades científicas son descubiertas a través de estas “máquinas”, mediante un control rígido de las circunstancias del mundo y representando las regularidades vistas en ellas (Op.cit, 161-2).

Si se aceptan las seis cuestiones descritas por Dale, entonces se deriva lo que se tratará en la siguiente parte del artículo, donde toma lugar la exposición de las consecuencias del pluralismo. El argumento condensado que traslada las observaciones sobre ciencias cognitivas, seres humanos y ciencias puede ser construido del siguiente modo (Op.cit, 162-3):

Premisa 1: ninguna teoría en ciencias cognitivas tiene aún suficiente comprensión en su aplicación.

Premisa 2: es posible que las teorías cognitivas no tengan los medios representacionales para contactar con las necesidades explicativas en todos los contextos de investigación.

Conclusión: es posible que los científicos cognitivos deban enfrentar la diversidad teórica para tener cobertura de sus dominios de problema.

Para considerar la posibilidad del pluralismo es así necesario tomar como válida la premisa 2, la cual no insiste en suponer limitaciones de ciertas perspectivas en el futuro, sino que solo la posibilidad de que estas perspectivas puedan estar limitadas en su contacto con la diversidad y complejidad de los intereses de las ciencias cognitivas. Esta segunda premisa podría ser cambiada en una que se basara en el fundamentalismo, esto es, que un esquema teórico sea identificado en el futuro (op.cit, 163-4).

Dale tratará de ir más allá de los extremos del fundamentalismo de enfoques versus caer en el nihilismo relativista mediante la presentación de la tesis pluralista como positiva. Esto significa considerar a la ciencia cognitiva más que como un “popurrí”, es decir, tomando el camino de *estrategias integrativas*. Luego, se mostrarán tres cosas: primero, cómo los sistemas complejos implican al pluralismo. Segundo,

cómo marcos meta-teóricos permitirían la coexistencia de teorías aparentemente inconsistentes y, tercero, modos de integrar enfoques teóricos mediante avances metodológicos (op.cit, 164-5).

El primer “modo integrativo” sería el enfoque emergente que trasciende las ciencias cognitivas e incluye disciplinas como física y biología. En efecto, es posible encontrar patrones emergentes coherentes a través de diferentes escalas en los sistemas complejos. Mucho del entusiasmo por el emergentismo deriva de la observación que en sistemas muy diferentes se encuentran principios similares de organización. ¿Cómo esta idea ayuda a las ciencias cognitivas? Principalmente, las características colectivas de los sistemas complejos mejoran el reconocimiento de la importancia del manejo de principios explicativos de múltiples niveles y facetas, principios que son la norma en un sistema complejo como el de los seres humanos y de sus contextos. El emergentismo sugiere que la coexistencia teórica puede ser una consecuencia necesaria de los diversos esquemas de medición en un sistema complejo definido como la sinergia entre cerebro humano, el cuerpo, el ambiente y su interacción (op.cit, 165-7).

Otra forma de “pluralismo positivo” sería a través del, actualmente prominente, marco matemático denominado “dinámica simbólica”, donde confluirían sistemas de medición diferentes, por ejemplo, pertenecientes a enfoques opuestos. Dentro de la “dinámica simbólica”, un espacio de estados de un sistema dinámico es dividido, de modo de asignar símbolos a las diferentes regiones de tal espacio, imprimiendo una secuencia de símbolos a la trayectoria a través de tales regiones. Teoremas bien establecidos mostrarían que ciertas características del sistema dinámico original se mantendrían en la (a menudo infinita) secuencia de símbolos. En otras palabras, las caracterizaciones discretas y continuas tendrían relaciones de equivalencia a través de la “dinámica simbólica”. Así pues, existirían desarrollos teóricos que revelarían que diferencias aparentes serían una cuestión nada más que de elección de medidas espacio-temporales (Op.cit, 168).

Además de la integración mediante el emergentismo y marcos matemáticos, Dale se explaya sobre dos razones “metodológicas” que mostrarían cómo el pluralismo haría valiosas contribuciones a la ciencia cognitiva. La primera razón es que los sistemas que diríamos que por su conducta poseen “cognición” son aquellos sistemas híbridos, entre los cuales SOAR – tradicionalmente simbólico – ha comenzado a integrar más esquemas representacionales y representaciones cuantitativas espaciales del ambiente de trabajo, violándose así la premisa simbólica de usar estructuras informacionales descontextualizadas y descorporizadas. La segunda razón es la riqueza de los datos conductuales que hoy llenan vastos repositorios digitales, la cual lleva a investigadores como Chris Eliasmith a desarrollar herramientas analíticas capaces de lidiar con el funcionamiento cognitivo del cerebro y la gran cantidad de datos en variados niveles, basando estas herramientas en conceptos tanto representacionales como dinámicos en un contexto de modelos neurocomputacionales y teoría moderna de control (op.cit, 169-171).

En la última parte de su artículo, Dale considera la perspectiva pragmática sobre el pluralismo, la cual puede ser tomada como base para tomar en serio al pluralismo. Siguiendo las ideas de Hilary Putnam, el cual defiende una postura que desafía la noción de ontología como útil para ayudar a los problemas de la ciencia humana, la filosofía y la sociedad, *Dale considera innecesaria la postura fundamentalista que busca una ontología esencial para los procesos cognitivos* (Op.cit, 173). Si alguien llegase a decir “lo he descubierto: tengo una teoría de la cognición” la respuesta a tal persona debiese ser “¿cuál clase de

cognición? y ¿qué aspecto?”, ya que la cognición es un sistema extraordinariamente complejo como un río o una selva, de lo que se sigue que es igualmente un sinsentido decir que se tiene “una teoría de la cognición” como decir que se posee “una teoría del río Mississippi” o “una teoría de la selva del Amazonas” (op.cit, 171-2).

On the nature of minds, or truth and consequences, de Shimon Edelman

El artículo de Edelman (2008) parte postulando que el que las mentes sean *fundamentalmente un fenómeno computacional*, parece contemporáneamente poco apreciado. Para Edelman toda tarea perceptual puede ser descrita usando formulaciones computacionales explícitas, ocurriendo lo mismo con la acción. Por tanto, no importa cómo la cognición puede ser descrita, la computación es lo que de hecho es (Edelman, 2008, 181-2). Pero ¿qué se entiende por “computación”? La computación es el proceso de establecer un mapeo entre dos dominios simbólicos, tomando tal mapeo la forma de una función o más generalmente de una relación. Los símbolos pueden como no pueden ser números, por tanto una ecuación diferencial perfectamente puede describir también una computación. Bajo esta definición *una roca rodando cuesta abajo computa su posición y velocidad en exactamente el mismo sentido que un notebook computa la posición y velocidad del cursor del ratón en la pantalla* – estos fenómenos solamente están implementando diferentes mapeos simbólicos. En efecto, el universo estaría implementando una computación, aunque una que va a la basura por la falta de algún proceso externo a éste que le hiciese sentido (op.cit, 182-3).

La noción de que cualquier sistema físico computa lo que sea que uno pueda, plausiblemente, afirmar que computa, es una aseveración fundamental de la ciencia de la computación, así como en neurociencia computacional. Este principio no está exento de críticas, ya que si uno lo sigue al pie de la letra significa que un estado instantáneo de mi cerebro en algún momento dado puede ser isomorfo al conjunto de moléculas de oxígeno en una habitación. Pero algo así sería una rareza, y aunque sucediese en un punto instantáneo, lo que importa es que la descripción computacional que se desenvuelve a lo largo del tiempo de esos dos sistemas sea isomorfa; el cerebro y las moléculas de oxígeno en la habitación serán sistemas que jamás se desarrollarán computacionalmente iguales (op.cit, 183).

Sobre estos puntos iniciales, Edelman se adentrará a la naturaleza de la explicación basada en tales supuestos, es decir, en la explicación que toma por básica la cognición como computacional. Lo primero a tener en cuenta es que los niveles de análisis de Marr, aunque han sido defendidos sobre bases epistemológicas, también han sido favorecidos por consideraciones ontológicas de una perspectiva más amplia. La razón para esto es que las estructuras jerárquicas son características, no solo de sistemas que han evolucionado, o que han sido diseñados para llevar a cabo procesos informacionales, sino que pertenecerían al universo en general. Aquí Edelman cita a Herbert Simon, quien dice (op.cit, 184):

“Scientific knowledge is organised in levels, not because reduction in principle is impossible, but because nature is organised in levels, and the pattern at each level is most clearly discerned by abstracting from the detail of the levels far below. [...] And nature is organised in levels because hierarchic structures – systems of Chinese boxes – provide the most viable form for any system of even moderate complexity”.

Lo segundo a tener en cuenta es que para Edelman, la cognición es un conjunto de fenómenos que, aunque siendo necesariamente computacional, es igualmente compatible a la explicación en términos de un modelo más general de computación – uno dinámico – o uno en términos más intuitivos – tipo Turing (Op.cit, 185).

Ahora bien, *el problema actual es que cualquier explicación que deja fuera el nivel funcional es incompleta en un sentido más profundo que una explicación que pierde alguno de los detalles de implementación*. Es importante notar, luego, que es el nivel funcional el que tiende a quedar fuera en algunas explicaciones de la cognición. La falla de distinguir las descripciones algorítmicas o implementacionales de teorías completas es un problema endémico en la literatura de ciencias cognitivas. Es común considerar que un sistema de producción o un modelo conexionista son “marcos teóricos”, no obstante el primero es un acercamiento algorítmico a una computación tipo Turing, y el segundo es de estilo implementacional. En particular, las contribuciones del conexionismo han sido eclipsadas por las presentaciones de éste como un “modelo milagroso de la mente”, destinado a obviar las necesidades por teorías clásicas de la cognición (Op.cit, 185-6).

Los anteriores apuntes *tornan igualmente inexactos el enfoque simbólico tipo LOT y el más reciente dinamicismo que persigue desplazarlo, en la medida que ambos ignoran los múltiples niveles que requiere una explicación*. Las ambiciones de estos dos enfoques están mal planteadas por dos razones. Primero, una explicación completa precisa tomar en cuenta varios niveles: el funcional, el procedimental y el implementacional. Segundo, la computación es, en cualquier caso, materia de interpretación de un sistema mediante otro – más que una propiedad intrínseca de cualquier sistema tomado aisladamente – y también lo son cualquier símbolo que esté o no esté involucrado. Desde este prisma tenemos ahora un punto de partida más fresco e informado para responder a la cuestión de cuál modelo de computación, discreto o continuo, es mejor que otro en ciertas circunstancias (Op.cit, 187).

Responder esta pregunta lo lleva a plantearse acerca de la complejidad computacional y el mejor modo de lidiar con ella. La posibilidad de transcribir un discurso a símbolos distintos, que puede ser transformado de nuevo en un discurso con muy pocos errores, demuestra que el lenguaje es discreto en un nivel fundamental pragmático. Lo discreto hace posible la dualidad de patrones: el significado en el lenguaje es transmitido en gran manera mediante la combinación de elementos que por ellos mismos son no significativos. Esta característica familiar del lenguaje natural no es una cuestión de necesidad, ya que, en principio, la comunicación puede proceder a través de un intercambio de señales continuas, paralelas y globalmente moduladas, con la información siendo transmitida vía parámetros de modulación en valores reales. En la práctica, sin embargo el lenguaje natural es discreto, serial e informativo localmente. Esto sugiere que hay una ventaja en la comunicación mediante una base discreta, pero ¿qué pasa con el uso de este medio para la coordinación de representaciones internas empleadas por los sistemas que entran en comunicación? La comunicación entre dos agentes es posible en la medida que existiría algún morfismo entre sus estados internos (Op.cit, 188).

Edelman considerará que una buena salida sería la “dinámica simbólica”, sin embargo para llegar a esta respuesta debe argumentar contra una comunicación basada en un sistema continuo, ya que, como se dijo antes, es posible la comunicación mediante un sistema continuo. Sin entrar en mucho detalle un

medio de tipo “dinámica simbólica” tendría ventaja sobre uno continuo en cuestiones como el aprendizaje y la resistencia al ruido (Op.cit, 189-190).

Finalmente, Edelman plantea que el pluralismo, entendido como diversidad explicativa en ciencia cognitiva, es una buena idea, ya que la computación es siempre una materia de interpretación, dado que la interpretación depende del nivel de análisis, además dado que las explicaciones discretas y continuas de computación a menudo pertenecen a diferentes niveles, y porque las explicaciones de un fenómeno complejo tienden a expandirse más que una por cada nivel (Op.cit, 192).

SAL an explicitly pluralistic cognitive architecture

La parte más “técnica” del artículo de Jilk et al. (2008a) ya ha sido expuesto cuando se presentó SAL en la sección (1.2.1.)(B1), y a continuación expondré aquella parte que expone una pequeña defensa en principio del pluralismo en ciencias cognitivas. Esta defensa arguye que las teorías en diferente nivel de detalles y a partir de diferentes perspectivas son mutuamente informativas y, además, que ningún nivel aislado puede capturar la total riqueza de la cognición (Jilk et al, 2008a, 198). Para empezar a discutir su punto, partirán con la declaración pluralista de Richard Feynman:

“..psychologically we must keep all the theories in our heads, and every theoretical physicist who is any good knows six or seven different theoretical representations for exactly the same physics. He knows that they are all equivalent, and that nobody is ever going to be able to decide which one is right at that level, but he keeps them in his head, hoping that they will give him different ideas for guessing.”

Superficialmente, esta declaración no parece controversial. Muchos científicos estarían de acuerdo con Feynman que tener diferentes descripciones de un mismo fenómeno ayuda a progresar la ciencia y tiene numerosos beneficios prácticos. En un nivel más profundo, sin embargo, muchos científicos – en física o ciencias cognitivas – también creen que hay una descripción que es *ontológicamente privilegiada*, la única que captura el modo en que las cosas realmente funcionan. En efecto, esto es lo que muchos científicos persiguen en sus investigaciones. En consecuencia, la perspectiva típica es que, mientras las otras teorías pueden ser *útiles*, la propia es la *verdadera*. Un caso especial de esta perspectiva es aquella del materialista eliminativo, para quien solo una pequeñísima parte de las características de los objetos son reales, mientras las otras son epifenómenos (Op.cit).

La noción que hay una descripción ontológicamente privilegiada de un fenómeno dado sería una extensión de la creencia en la “identidad metafísica”. Si existe una manera en que las cosas son, entonces debe haber una manera correcta o verdadera para describirlas. Pero, para Jilk et al. este no es caso, puesto que una descripción de un fenómeno no es el fenómeno, ni es una implementación de éste; en particular *una descripción es una abstracción*. Cualquier descripción finita debe omitir algunos detalles en favor de otros que serán más predictivos, más reveladores o más importantes para nuestros propósitos. Luego, por cualquiera de tales descripciones, hay otra que elige incorporar algunos de los detalles omitidos y dejar fuera otros previamente incluidos; esta segunda descripción no es menos verdadera, meramente tiene diferentes *prioridades* (Op.cit, 198-9).

Un ejemplo concreto y simple de la validez plural es el comportamiento del agua líquida. En una descripción de nivel fino hay moléculas de agua; en uno más grueso hay ondas. Es fácil ver los beneficios prácticos de ambas representaciones diferentes. Sin embargo, los reduccionistas dirían que las moléculas de agua son ontológicamente privilegiadas. Empero, si uno conociese solo el micro comportamiento de las moléculas, uno ni siquiera podría predecir, ni menos esperar, un fenómeno tal como las ondas. El movimiento de incluso tres partículas – imagínense miles de millones – no es susceptible de solución analítica, y dejando de lado la potencia computacional requerida para una simulación, es improbable que la descripción y comprensión de las moléculas de agua, sin un primer conocimiento de las ondas, sea suficientemente certero para que tal fenómeno de alto nivel se manifieste en el modelo simulado. La única manera de que las ondas apareciesen en la simulación de las moléculas de agua sería si restringiésemos la caracterización de tales moléculas en modos plausibles, hallando aquellas que de hecho produzcan el comportamiento de onda y, a continuación, determinando empíricamente cuál de estas caracterizaciones restringidas de hecho se condice con la física de las moléculas de agua. Este ejemplo mostraría que ondas y moléculas de agua son dos descripciones incompletas y diferentes de la unidad de un comportamiento físico (Op.cit, 199).

Es importante ver que el pluralismo no es lo mismo que el relativismo; primero, nótese que no es el caso que cualquier teoría sea válida, ya que esta debe ser consistente con los hechos. Segundo, aún cuando existiesen varias descripciones diferentes de un fenómeno, todas las cuales sean consistentes con los hechos, no sería el caso que todas ellas sean igualmente buenas. Aunque, la intención es desacreditar el privilegio ontológico, las descripciones pueden ser epistemológicamente privilegiadas. Un estándar para esto es la parsimonia, es decir, preferir descripciones que explican más con menos conceptos. Otro estándar es la coherencia vertical y horizontal. En esencia, el pluralismo es una extensión del instrumentalismo, adhiriendo a éste mediante la aseveración de que diferentes sistemas de conceptos y teoría pueden describir el mismo fenómeno, sin contradicción, pero teniendo diferentes metas y énfasis (Op, cit., 199-200).

La aparentemente sutil distinción entre privilegio ontológico y epistemológico tiene grandes consecuencias para la sociología de la ciencia, y en particular para el pluralismo. A pesar de los mejores esfuerzos de uno por ver el valor del trabajo de otros en el mismo campo o en la periferia, desde la perspectiva del privilegio ontológico es difícil evitar un desprecio disimulado por los resultados ajenos “epifenomenológicos”. Si en vez de ello se mantiene un punto de vista de *privilegio epistemológico*, entonces se seguirá el pluralismo. Sin abandonar las creencias que se tenga sobre la naturaleza subyacente de la realidad, se podría reconocer la inefabilidad de ellas. Así, para Jilk et al. se podría terminar lo que es esencialmente una guerra religiosa sobre lo metafísico verdadero, tomar intuiciones y restricciones de los otros enfoques y, con algo de suerte, mejorar el propio enfoque (Op.cit, 200).

Entonces, finalmente, ¿qué se entenderá por “pluralismo explicativo”?

Una definición totalmente específica de que es “pluralismo” me parece una tarea excesiva, pero creo que es posible elaborar con lo expuesto hasta acá una noción que esté en un lugar intermedio entre Bokulich y Suppes, y que sirva para ir elaborando más el punto:

Pluralista es la postura que acepta como hecho connatural a la investigación de cualquier fenómeno definitorio de una disciplina o subdisciplina el que éste deba ser tratado con un popurrí de métodos científicos, técnicas y conceptos.

Tomando este punto de partida, existirán variados grados, y al menos dos dimensiones (epistemológica versus ontológica), con que esta postura es entendida. Por un lado, y siguiendo la caracterización hecha por Van Bouwel (2014), es posible distinguir epistemológicamente entre cinco clases de pluralismo: el moderado o temporal, el integrativo, el interactivo, el aislacionista o de "niveles de análisis", y el pluralismo "todo vale". Muy brevemente cada uno se define como (Van Bouwel, 2014, 107-9):

(.i) El pluralismo moderado reconoce y promueve la pluralidad temporal de teorías en competencia como medio de alcanzar cierta unidad de las ciencias a largo plazo.

(.ii) El integrativo tendría como principio definitorio la integración de múltiples explicaciones las cuales darían cuenta de un mismo fenómeno en sus diferentes niveles de explicación.

(.iii) La postura del pluralismo interactivo presupone dos puntos, primero que explicaciones satisfactorias pueden ser obtenidas sin la integración de múltiples niveles, luego volviendo el imperativo integracionista innecesario y, segundo, que no se debe desalentar la interacción entre niveles, dado que interacción e integración puede llevar a mejores explicaciones.

(.iv) La clase aislacionista se opone al pluralismo integrativo en cuanto que considera cierto privilegio en algún nivel de análisis frente a otros para responder ciertas preguntas, lo cual implica la no necesidad de integración ni de interacción entre niveles, aun cuando considera válidas el conjunto total de las preguntas. Presume cierta "clausura" o autonomía del nivel de análisis.

(.v) El pluralismo tipo "todo vale" representa una defensa por retener todas las investigaciones, posiblemente inconsistentes, que emergen de una comunidad científica.

Luego, "placed on a continuum going from monism to anything goes pluralism, we thus have monism, moderate pluralism, integrative pluralism, interactive pluralism, isolationist pluralism and anything goes pluralism. This ordering reflects increasing strength of the pluralist position." (Van Bouwel, 2014, 109).

Por otro lado, en la dimensión ontológica, es posible visualizar dos extremos – aunque no he elaborado ni he investigado cuáles serían los puntos intermedios, ya que ello es tarea del filósofo de la ciencia interesado en metafísica, que no es mi caso – los cuales serían, por una parte, que la realidad en sí misma es plural, o *pluralismo ontológico*, postura implicada en el "pluralismo teórico" de Bokulich, y por otra parte el monismo. Este monismo tendría la forma de materialismo eliminacionista (Jilk et al., 2008a), el cual sería idéntico al fisicalismo en caso que por "materia" se entendiera aquel fenómeno que es del dominio exclusivo de la física. Para evitar confusiones se puede denominar "monismo metafísico" al extremo que se opone al pluralismo ontológico, mientras que "*monismo explicativo*" será la postura que está en el extremo opuesto del pluralismo "todo vale".

Los tres últimos artículos expuestos varían en *grado del pluralismo* que proponen. Dale (2008) va en línea con un pluralismo integrativo, el cual se acerca bastante al ecumenismo de Clark (1998). Edelman (2008) es bastante conservador, tomando una postura cognitivista llamativa por su claridad y coherencia interna (v. gr. comprometiéndose con que todo el universo estaría llevando a cabo alguna computación

de modo que no sea “anómalo” solamente dar tal estatus a la cognición) y que, dado que defiende la posibilidad de integrar dinamicismo y simbolismo, entonces puede considerarse “pluralista” en cierto sentido, por ejemplo del tipo (.iv). No obstante, un pluralismo aislacionista en ciencias cognitivas no sería como la propuesta de Edelman, sino que dejaría que cada grupo de enfoques corriera por su propio carril, sin intentar de forzosamente reunirlos bajo un lema como que “la cognición es fundamentalmente computacional”. Luego, considero que la manera adecuada de entender a Edelman (2008) es un caso de pluralismo moderado, que toma el estado de las ciencias cognitivas como transitorio, hasta que se llegue a una teoría que tenga como trasfondo el análisis de Marr y dé por esencial para la cognición su “computacionalidad”. Finalmente, el caso de Jilk et al. (2008a) es el que me parece más interesante, pero que retomaré un poco más adelante para elaborar mi propuesta sobre qué entenderé por “pluralismo explicativo”.

Ahora bien, en esta serie de apuntes falta agregar que lo que se puede decir por *actitud pluralista*, y que he nombrado anteriormente sin definir, no es tanto una postura epistemológica, como más bien es una *forma de actuar dentro de la investigación*. En este sentido, es posible la actitud pluralista, pero sin reconocer una epistemología pluralista del tipo (.i) al (.v). Esto es importante notarlo, ya que en ciencias cognitivas se dan de facto tales casos. Para ello uno puede remitirse al ejemplo de TOTE, donde de manera explícita se construye sobre los elementos de otro “paradigma”, el cibernético, pero se reivindica una forma explicativa como la que regiría el futuro de la ciencia cognitiva. TOTE será entonces un caso de *modelo híbrido*⁹⁴, pero defendido no pluralistamente. Ergo por “modelo híbrido” se entenderá *aquel modelo que está construido a partir de una actitud pluralista*. Una actitud pluralista puede considerarse como un paso previo no necesario para cualquier variante desde (.i) a (.v), así como también todas los grados diversos de pluralismo pueden considerarse paso previo no necesario de alguna tesis pluralista ontológica. Creo que es totalmente lícito apelar a la *separabilidad* de la práctica, con la esfera epistemológica, y a su vez ésta de la esfera ontológica. Esta separabilidad sería justificable basado en que es la interpretación filosófica de un quehacer la que crea las categorías metafísicas, siendo tales categorías difícilmente achacables a un descubrimiento en el mundo.

Con todo lo anterior a la vista, entonces lo que entenderé por *pluralismo explicativo* es una de las variantes sutiles que existirá dentro del rango de pluralismos epistemológicos que se acaba de proponer, y que defino por cinco características clave:

- **Primero**, y la característica más trivial dado que se deduce por su solo nombre, *el pluralismo explicativo sostiene que no hay una única, completa y comprehensiva explicación, teoría o modelo del mundo ni de cualquier fenómeno*. Esto puede derivarse directamente mediante el argumento filosófico quineano de la subdeterminación de las teorías científicas, pero en la presente sección se han desarrollado otra serie de puntos que sirven para reforzar el pluralismo a partir de la falta de unidad concebida de diferentes modos (Suppes, 1978) y una serie de hechos propios de las ciencias cognitivas, la epistemología y el propio ser humano (Dale, 2008). Sin embargo, la primera característica del pluralismo explicativo, no es meramente el ser pluralista, sino que es una postura

⁹⁴ A partir de lo que he expuesto en la primera parte, y en la sección anterior, he delineado que por “híbrido” se entiende la “mezcla de elementos” de diferentes “paradigmas” en un mismo modelo. Sin embargo, el corazón de la tesis es, justamente, desarrollar un medio conceptualmente más pulido para dar cuenta de tal “mezcla de elementos” y de “paradigmas”.

que, puesto que afirma que no hay una única explicación de cualquier fenómeno, entonces normativamente asumiré que *aquellas posturas que legítimamente pueden ser denominadas como “pluralistas” son aquellas que se mueven en los grados (.iii), (.iv) y (.v)*. Esto es importante destacarlo, ya que permite entender desde dónde es posible afirmar que el ecumenismo de Clark no es pluralista. Ergo, el pluralismo explicativo es una postura epistemológica en una zona filosófica delimitada por los pluralismos (.iii) y (.iv), pero que no se expande hasta el grado (.v), ya que no es *conservador de las teorías simplemente justificándose en su diversidad*.

- **Segundo**, el pluralismo explicativo es una variante dentro del espectro de pluralismos epistemológicos que son separables de implicaciones ontológicas, por tanto se aleja de la idea de Bokulich (2008) donde necesariamente el pluralismo se extiende en el terreno ontológico como una postura bien definida. Por otra parte, tampoco considero que el pluralismo explicativo sea necesariamente en su totalidad instrumentalista (Jilk et al., 2008a), ya que cualquier teoría implica un mínimo de ontología (Quine, 1951), por lo tanto el pluralismo explicativo es en cierto punto realista sobre tal mínimo, pero instrumentalista respecto a la permanencia de este piso mínimo de ontología⁹⁵. Considérese, por ejemplo, un mundo posible donde un Newton hiper-instrumentalista simplemente hubiese llenado de fórmulas su *Principia* con cero ontología. Eventualmente, el propio éxito de esta mecánica clásica instrumentalista habría generado suficiente presión para estipular cierta ontología para expandir su éxito, es decir, independiente de si es posible generar una teoría instrumentalista perfecta, ello no asegura que su evolución sea también perfectamente instrumentalista, sino que tendríamos que obcecadamente interpretarla de esa manera. Pero, como contrapartida a esta situación posible, considero legítimamente concebible que en un periodo extensísimo de tiempo absolutamente ninguna categoría perteneciente a cierta ontología de una teoría persista. Es más, basta asumir que, al menos, ha existido un caso de eliminación de una teoría para descartar el realismo extremo respecto de las teorías científicas. En síntesis, *desde una perspectiva temporal absoluta el pluralismo explicativo puede considerarse como instrumentalista* (Jilk et al., 2008a, 200), *pero desde la perspectiva temporal relativa y acotada propia de la investigación científica – y nuestras propias vidas – el pluralismo explicativo es realista sobre aquello que es más resistente a los cambios*
- **Tercero**, los factores de avance técnico determinan a tal grado lo teórico que no tiene sentido tampoco restringirse a una postura que sea pluralista, pero sólo en el nivel de los conceptos teóricos, sino que también *el pluralismo explicativo implica apertura respecto a las metodologías de investigación posibles, lo cual en términos prácticos puede involucrar la colaboración*. Dada la digresión 2 en la sección (1.3.3.), entonces, si esta se aplica más profundamente, entonces siempre existe una metodología que puede ser útil de una manera novedosa para permitir una mejor descripción de los datos, lo cual, subsecuentemente puede generar una mejor explicación. Un ejemplo de la colaboración puede verse en el caso de los modelos híbridos, como SAL, arquitectura

⁹⁵ En la terminología que desarrollaré en la siguiente sección sobre la tesis de los instrumentos conceptuales (IC) esta mínima cantidad de ontología es lo que denominaré “CO”, o conjunto ontológico, y el subconjunto que puede ser sacado sin pérdida de poder explicativo, predictivo, ni pérdida de *potencial evolutivo de la teoría* será el su TF trasfondo filosófico. Véase Def.4.2.

defendida por Jilk et al. quienes consideran que “pluralism is about the practical benefits of open-mindedness and collaboration” (2008b, 267). Esto lleva precisamente a una cuestión que se ha postergado a propósito, a saber, en qué clasificación de pluralismo cae la postura de Jilk et al. Considero que los autores están en una postura entre (.iii) y (.iv), en cierto “territorio” que estipulo se ubica mi definición de “pluralismo explicativo”, ya que, aunque dicen que es deseable la colaboración (pluralismo interactivo), a reglón seguido dirán del pluralismo que “(...) also requires that we accept the possibility of incommensurable complementary theories holding sway simultaneously” (op. cit.). Esto significa que *es totalmente concebible una ciencia cognitiva pluralista que use teorías de diferentes “niveles de análisis” (pluralismo aislacionista) al mismo tiempo, que corran en carriles paralelos*. Por consiguiente, expandiendo lo apuntado en la primera característica del pluralismo explicativo, a saber que se mueve en un territorio amplio desde el pluralismo grado (.iii) a (.iv), la presente definición *no apela a priori ni a la colaboración ni a la no colaboración*.

- **Cuarto**, siguiendo a Jilk et al. (2008b, 266) quienes dicen que “pluralism necessarily incorporates the idea that different theories are incommensurable – otherwise they are either isomorphic or could be merged into a single theory”, *para el pluralismo explicativo la inconmensurabilidad será un hecho constitutivo de cualquier disciplina*, ya que, de modo contrario, entonces sería injustificado mantener una epistemología pluralista – aunque quizás es posible una actitud pluralista “heurística” – si el dominio sobre el cual se está trabajando no es mínimamente pluralista. Así pues, la competencia entre teorías seguirá existiendo, pero de un modo totalmente diferente, ya que del pluralismo explicativo no se sigue un estado final convergente en una disciplina.
- **Quinto**, y finalmente, *el pluralismo explicativo no es anarquismo epistemológico*, dado que como dicen alegóricamente Jilk et al. (2008b, 265) “As in political debate, the commentaries in this issue mostly tend towards the two ends of a similar spectrum: ‘one true path’ versus ‘anything goes’. We hope to persuade swing voters of a more moderate position: *diversity without chaos*”, es decir, no es anarquismo epistemológico, porque aún existe posibilidad de apelar a (algo así como) “patrones de virtud epistemológica” entre dos teorías con exactamente iguales poderes predictivos (Jilk et al., 2008a, 199-200). Por otra parte, el pluralismo explicativo no es conservador de todo modelo o teoría simplemente por su diversidad como puede ser interpretado en el pluralismo tipo “todo vale”, es decir, concuerdo con la opinión de que “(...) pluralism is not a cultural diversity initiative – it does not encourage incommensurable theories for their own sake. Such theories are only valuable to the extent that they contribute novel explanation.” (Jilk et al., 2008b, 267). A esto agregaré que la tesis de los IC que desarrollaré en la siguiente sección (2.3.) visualiza teorías y modelos de manera que es irreconciliable con una visión pura del “todo vale”, puesto que *considero posible construir un criterio de demarcación suave, compatible con la noción de pluralismo* que dio comienzo a esta última parte de la sección (2.2.).

2.3. La tesis de los Instrumentos Conceptuales (IC) y sus consecuencias

Dado que los intentos por encontrar condiciones necesarias y suficientes para dar con una tesis general sobre qué constituye una explicación científica han terminado fracasando a la larga⁹⁶ – o al menos jamás han logrado tener una capacidad general de subsumir todas las ciencias, sino que se contentan con constreñir una de ellas –, entonces no tomaré el camino normativo respecto a las teorías y modelos en ciencias. Por consiguiente, recorreré un camino que, utilizando como trasfondo las ideas del pluralismo, intentará crear un *medio conceptual fundamentalmente descriptivo* que permita subsumir toda forma explicativa-predictiva empírica⁹⁷, pero que aun buscando tal grado de generalidad no llegue a caer en la vaguedad⁹⁸. Una vez hecho esto, entonces mi meta será tratar con un par de ejemplos de su aplicación, al menos en el terreno de ciencias cognitivas (sección (2.4.) y subsecciones respectivas), para luego ver cómo desde esta tesis es posible lidiar con parte de las preguntas esbozadas al final de la primera parte de la tesis. Las posibilidades de normar en la práctica científica a partir de una expansión sobre este medio conceptual fundamentalmente descriptivo serán tratadas en el contexto de las conclusiones y quedarán meramente bosquejadas para un trabajo futuro.

2.3.1. Las metas de las ciencias

En la sección (2.1.) cuando se habló de que los “hechos básicos” de todas las ciencias son las “generalizaciones” empecé a referirme sobre el contenido directo de las *ciencias una vez desarrolladas*, contenido que, nótese nuevamente, *no trata de lo particular, sino que de lo universal* – nunca se es demasiado majadero en filosofía sobre este punto fundamental. Esta situación se extiende más allá de las explicaciones que se han caracterizado como “nomológicas” (o de leyes cubrientes pertenecientes a la *concepción heredada*), ya que, como Bechtel y Abrahamsen (2005) indican, el mecanismo, aun cuando no tiene los mismos rasgos que una explicación nomológica, sí es equivalente a ésta en el sentido que se generaliza o universaliza mecanismos concretos al partir de, o bien un estudio extensivo de casos, o al tomar un caso como representativo – por ejemplo, el caso del calamar gigante como modelo para estudiar la trasmisión de potenciales de acción (Hodgkin & Huxley, 1952).

Dejando de lado, por ahora, la cuestión sobre *la conexión de tales generalizaciones con el mundo y cómo éstas generalizaciones se agrupan en sistemas estables que organizan su relación y que serían comunes a más de una disciplina* – siendo esta, por ejemplo, la finalidad tácita de presentar al mecanicismo abarcando desde la biología hasta las ciencias cognitivas o la concepción heredada que buscaba proposiciones universales o leyes en todas las ciencias – me interesa centrarme en lo que parece ser un acuerdo común en filosofía de las ciencias sobre las metas del quehacer científico, a saber, que *sus metas, en sentido amplio, se reducirían a exclusivamente tres: describir, explicar y predecir fenómenos*. Otras metas podrían ser tomadas como especificaciones de las tres centrales, por ejemplo subdivisiones de la meta de predicción como serían la meta de controlar o retropredecir y, dado que interiorizarme

⁹⁶ Pienso en el positivismo lógico y sus intentos de demarcación.

⁹⁷ Específicamente la noción de “instrumento empírico”, que he implicado en la digresión 1 subsume en sí todas las ciencias cognitivas.

⁹⁸ Pienso aquí en el historicismo del primer periodo de investigaciones en filosofía de las ciencias de Kuhn, donde por “paradigma” se entendían demasiadas cuestiones diferentes. Este punto de partida puede no ser válido para el especialista en el historicismo, pero constituye, a mi modo de ver, un punto de partida suficientemente legítimo.

cabalmente en éstas de modo de demostrar que son subdivisiones sería un trabajo aparte, entonces simplemente continuaré teorizando a partir de este grupo de tres más generales y transversalmente mejor aceptadas.

Pues bien, de estas tres metas, la primera, “describir”, a pesar que pueda parecer similar superficialmente a la práctica *típica* de la interacción normal cuando nosotros “describimos”, por ejemplo, dónde viajamos el fin de semana a nuestros amigos o cómo estuvo el clima de ayer con algún desconocido, las ciencias, al tratar sobre lo general (y no como nosotros que lidiamos con lo particular) tendrá un sentido totalmente diferente. Por consiguiente, para que no exista confusión, la práctica de “descripción científica”, es mejor aludirla como “redescripción”, ya que, por simple constatación es posible verificar, un rasgo trivial a toda teorización empírica es la transformación del lenguaje natural que sirve para hacer referencia a un fenómeno en otro lenguaje que nos sea más práctico para los otros fines de “explicar” y “predecir”. Por ejemplo, si se va a ciencias cognitivas, en buena parte de ella no existe fenómeno representado independientemente de algún formalismo o redescrito en términos de uno; en el dinamicismo se utiliza sistemas dinámicos, lo que implica ecuaciones diferenciales⁹⁹, aunque tal mismo formalismo es útil para introducirse al modelamiento neuroelectrofisiológico (Scott, 2002). En el conexionismo, al menos dentro del grupo PDP, el álgebra lineal es fundamental para tratar con redes neuronales, las cuales se reducen a una serie de matrices que varían a lo largo de un tiempo discreto (Rumelhart & McClelland, 1986, cap. 9). En el caso del simbolismo, lo básico aquí es ser capaz de manejar pseudocódigo o, más concreto aún, algún lenguaje de programación, que sirvan para empezar a lidiar con el nivel algorítmico – ello es evidente a partir de los mismos trabajos de Newell y Simon.

Sin embargo, las redescripciones no se limitan al uso de instrumentos formales. Por ejemplo, el enactivismo utiliza bastante vocabulario de la tradición budista para autodefinirse y definir el fenómeno cognitivo (Varela et al., 1991, parte V). O el propio cognitivismo que, complementariamente al uso del formalismo simbolista, redescibirá el fenómeno mental conservando parte del vocabulario de psicología *folk* (Pylyshyn, 1980). Luego, la redescripción la única condición que debe poseer para considerarse como tal es algo que puede postularse negativamente, a saber, que no es una descripción típica.

Esto es interesante, ya que así como redesciben los instrumentos empíricos, también lo hacen los formales, e inclusive la filosofía a medida que se va especializando y se aleja de las descripciones que se dan en el ámbito típico de la interacción social. Y esto último es importante, ya que posibilita, enunciar más completamente qué trato de decir con la noción de “redescripción” en ciencias; *redescibir consiste en describir algo en cierto lenguaje técnico que no puede ser reducido a una descripción típica*. Esta irreductibilidad de las redescripciones se constata cuando se trata de introducir al lego a cualquier disciplina, donde es a través del lenguaje natural – o mediante metáforas – que se le explica qué significa y en qué contexto se utiliza tal o cual término especializado, pero se le previene de caer en la confusión de que tales nociones no son exactamente lo que se entiende en el lenguaje técnico. Esto es poco visible donde el lenguaje técnico tiene alto traslape con el lenguaje natural, pero se hace patente en instrumentos empíricos bien desarrollados, donde seguir la metáfora al pie de la letra puede llevar a

⁹⁹ Esto es evidente cuando en Thelen & Smith (1996, 57) se presenta la estabilidad dinámica y atractores mediante, por ejemplo, curvas sinusoidales amortiguadas, la cual ha aparecido en el contexto de la digresión 1.

severas confusiones. Por ejemplo, la clásica metáfora pedagógica que dice “la corriente eléctrica es como una corriente de agua”, puede llevar a alguien que toma al pie de la letra tal imagen a preguntar “¿cómo entonces es que no se acaba la electricidad que viene desde los generadores, si un generador sería como una ‘fuente de agua’?”, si no tiene en cuenta que el conocimiento técnico de la gran distancia entre las escalas físicas de “corriente agua” y “corriente eléctrica”.

En las ciencias ya desarrolladas la redescrición en sí misma es una actividad nunca secundaria, sino que complementaria a la actividad explicativa, que está guiada por la actividad de generación de hipótesis (Casadevall & Fang, 2008). Sin embargo, a pesar de la complementariedad, la redescrición de órdenes cada vez más altos no es la meta en sí misma de las ciencias, sino que aquella sería un paso que se va dando paralelamente a la aparición de explicaciones (Ellis, 1956, 505-6). Luego, un proceso es anterior al otro, o dicho de otro modo, *se puede redescibir sin explicar, pero no se puede explicar sin redescibir el fenómeno de interés. Exactamente lo mismo puede decirse de la predicción*, por ejemplo, si nos remitimos a los casos de formalización directa. Éste será un modo de redescrición basado en nuestro éxito histórico que hemos tenido como especie en el uso de las matemáticas, desde el desarrollo de la aritmética, donde es el constructo formal el que funciona cuasi como “instrumento predictivo puro”. Considérese el caso del modelo predador-presa de Volterra, el cual se generó a partir de propiedades relevantes tales como tasas de crecimiento exponencial de presas en ausencia de predadores y tasas constantes de muerte de predadores, a su vez que ignoraba propiedades fisiológicas de las presas reales (Weisberg, 2007, 210-2; Simmons & Robertson, 2002, 457-462). En este modelo son hechas *mínimas presuposiciones explicativas* del sistema real mismo, como por ejemplo que necesariamente un cruce de un predador y una presa incide en la población de ambas especies, sin entrar a detallar cómo es que los predadores tienen tal tasa extraordinaria de éxito en la caza.

Esto lleva a una siguiente fase dentro de esta sucinta investigación filosófica sobre las tres metas generales de las ciencias; *la redescrición sería un rasgo trivial de todo instrumento empírico, no así la explicación y la predicción, las cuales sólo son satisfechas en la medida que la redescrición es suficientemente útil, y que pueden darse en grados distintos dentro de un mismo instrumento empírico. Luego, la meta común a todo instrumento empírico es una combinación variable en grados de explicación y de predicción*. Esto a primera vista puede ser chocante, ya que cómo ejemplifiqué justo arriba “explicar” y “predecir” parecen actividades bien delimitadas que serían mutuamente excluyentes y que podrían no darse a la vez en un mismo instrumento empírico.

Pues bien, el que la distinción de las nociones entre predicción y explicación pueda ser hecha conceptualmente no implica una refutación *a priori* de lo que trato de mostrar, a saber, que ambas metas pueden darse conjuntamente, solo en una variación independiente en grados cada una en todo instrumento empírico. En el fondo, lo que trato de mostrar es que filosóficamente es tan válido suponer que la meta de cualquier instrumento empírico será *a priori* la explicación, como suponer que es la predicción, como suponer que es la *explicación-predicción*. En general, que es igual de válida una abstracción filosófica que dé preponderancia a cada una por separado de las metas nombradas como a las dos en conjunto. Para hacer esto me explayaré un poco más sobre algo que dejé momentáneamente de lado al principio de esta sección, la cuestión de *la conexión de las generalizaciones con el mundo*.

La concepción heredada para tratar con esta cuestión hacía una separación taxativa entre enunciados observacionales versus enunciados teóricos. La forma en que se conectaban generalizaciones y mundo era sencilla: el mundo era constatado mediante enunciados observacionales, los que posteriormente eran organizados mediante los teóricos, estos últimos siendo el contenido mismo de las teorías. Sin embargo, existen problemas con esta visión que se empeña en separar categóricamente lo observacional de lo teórico, ya que depende de los “dogmas” de la distinción analítico/sintética y el reduccionismo verificacionista (Quine, 1951) y, aunque el tratar de rescatar versiones más moderadas de lo observacional y lo teórico es aún una disputa en curso (Rottschaefer, 1976), la moraleja de tal historia es que lo metafísico “se nos cuele por todos lados”, es decir *las redescpciones en sí mismas y en su grado más mínimo ya son en parte enunciadas teóricamente*¹⁰⁰. Ergo, dado que la conexión generalizaciones-mundo ya heredaría problemas si se sistematiza desde una filosofía de las ciencias demasiado positivista lógica en su inspiración, entonces vale la pena inclinarse por una mirada un poco más descriptiva del desarrollo de las ciencias, de modo de saber hasta qué punto ésta puede sistematizarse filosóficamente.

Si miramos el desarrollo desde lo “no científico” a las “científico”, constatamos que, dado un constructo teórico altamente explicativo – entendido “altamente explicativo” como postulando prodigiosamente estructuras causales en el mundo –, siempre existirá otro ante el cual palidecerá en términos de predicción, en tanto que este otro estará mejor vinculado al mundo pragmáticamente mediante su método confirmativo, por consiguiente teniendo más poder predictivo a costa de sacrificar pretensiones explicativas/causales, es decir, a costa de utilizar un método – el de la experimentación empírica – que restringe en demasía la cantidad de mecanismos causales que pueden ser investigados a lo largo del tiempo. Por ejemplo, las explicaciones religiosas pueden llegar a configurarse en redes intrincadas de mecanismos causales trascendentes y postular su propio método de verificación (y confirmación), por ende siendo altamente “explicativas”, pero pobremente predictivas bajo la luz de las *primeras teorías científicas modernas*. En efecto, y contrariamente, las explicaciones científicas usualmente son parcas en su ontología de causas, privilegiando casi siempre la capacidad predictiva. El que al día de hoy sean sendos castillos metafísicos que compiten en su inextricabilidad con la más compleja de las ontologías religiosas no significa que puedan ser comparados, ya que tal semejanza es sólo superficial.

Ahora bien, aunque sea concebible acercarse más y más a mejores predicciones, lo que a su vez volvería potencialmente irrelevante buscar explicaciones del porqué de los fenómenos (considérese el caso que fuésemos perfectos demonios de Laplace), las explicaciones nunca cesarán de existir, *ya que su esencia está ligada a la capacidad de organizar nuestra experiencia, permitiendo el uso eficiente de nuestras capacidades cognitivas finitas*¹⁰¹. Esto último se puede desprender de una revisión del paso de lo “no científico” a lo “científico” como la hecha justo antes, donde no es solamente la capacidad predictiva lo que permite a lo “científico” terminar de cavar su nicho, sino que es una combinación de un uso

¹⁰⁰ Nótese que son las redescpciones y no toda descripción, como sería el caso con las descripciones típicas en la interacción común, donde asumiré sin entrar en justificaciones que no nos comprometemos ontológicamente del mismo modo que cuando se redescrive en contextos científicos. Esta falta de compromiso me parece evidente meramente por el hecho de la facilidad con la cual podemos corregir o rectificar nuestras descripciones, corregibilidad que contrasta con la de la redescpción que, de ser modificada, significa un golpe crítico a la teoría.

¹⁰¹ Esta idea está inspirada directamente de Douglas (2009, 454), pero, creo, el uso que doy aquí es diferente al original, es por ello que lo acoto como nota.

adecuado de ontología explicativa y de predicción. Y aquí está el pequeño truco: si se acepta lo anterior, a saber, que algo para que cuente como “explicativo” debe ser esencialmente una herramienta cognitiva que nos ahorra “costo de procesamiento cognitivo”, entonces cualquier constructo que sea capaz de predecir ya constaría de un componente explicativo – v. gr. cualquier formalización directa – y cualquier instrumento que sea explicativo, por el mismo desarrollo científico, será también en parte predictivo, ya que de lo contrario sus generalizaciones no estarían justificadamente conectadas al mundo.

En síntesis, aunque la tesis de la simetría hempeliana no sea estrictamente cierta (Douglas, 2009, 457-8), hay en la ingenuidad positivista lógica algo que se conserva luego de infructuosos intentos de tratar de capturar unívocamente el “espíritu” de “lo científico”: *la mejor testeabilidad de las predicciones de una teoría es lo que más naturalmente le otorga el estatus suficiente a una explicación como competidor por ser la mejor para capturar cierto fenómeno* y, por otra parte, *la misma historia de las ciencias previene que castillos puramente explicativos con nula capacidad predictiva sean parte de las disciplinas científicas*. Ergo, la meta común a todo instrumento empírico es una combinación variable en grados de explicación y de predicción y es lícito suponer que, a priori, una unidad de análisis en filosofía de las ciencias se plantee como diciendo *que la meta de un instrumento empírico es la explicación-predicción*.

Aceptando los argumentos ya explicitados antes, a saber, que, primero, la redescrición es un rasgo trivial y básico de todo instrumento empírico y, segundo, que explicación y predicción pueden ser disociados conceptualmente, *mas no cuando se mira a las ciencias en su desarrollo y en sus resultados, donde serían procesos complementarios*, entonces extraigo que lo *mínimo que puede ser dicho de cualquier instrumento empírico, sin conocer exactamente su aplicación, ni su contenido, es que siempre es el caso que es una redescrición cuya meta es explicar-prededir*. Esto es un buen progreso, ya que esta noción mínima intenta seguir una estrategia lo más descriptiva posible, tratando de restringir juicios normativos, de modo que sea fácilmente compatible con la definición de “pluralismo explicativo” que fue descrita en la sección (2.2.). Con esto en mente es posible trabajar en el problema de cómo las generalizaciones científicas se agrupan en sistemas estables que organizan su relación y que serían comunes a más de una disciplina. Mi propuesta que comenzaré a detallar en lo que sigue se basa en el *meta fundamental de producir un sistema estable que sea tan abstracto que, cualquiera sea la disciplina, ésta siempre contendrá constructos que se puedan descomponer o interpretar según mi propuesta de análisis*, de modo de, al final, lograr una descripción más pulida sobre las relaciones interteóricas entre simbolismo y conexionismo – y, presumiblemente, de la relación de cualquier instrumento empírico con cualquier otro instrumento empírico.

2.3.2. La abstracción de los IC y sus principios de construcción

Si se toman como válidos los apuntes que se han hecho hasta ahora, entonces una primera definición sobre qué constituye un *instrumento conceptual* será:

Def.0.1. Un artefacto o instrumento conceptual (IC) es todo aquel constructo conceptual que mediante redescrición, cumpla las metas de explicar y predecir un fenómeno empírico.

Esta definición fundamental está sustentada en las nociones de “redescrición” y “explicar-prededir” defendidas en la sección (2.3.1.) y he transformado la noción de “instrumento empírico” que fue presentada intuitivamente en la digresión 1 – sección (1.3.2.) – en la de “instrumento conceptual”, cuyo

contenido será desarrollado en el plano filosófico en la subsección que sigue. En general, la noción de IC es creada como una variable que no especifica ni la noción antigua de “teoría”, ni la de “modelo”, que han sido siempre parte del vocabulario clásico en filosofía de las ciencias, sino que genéricamente pretende subsumirlos a todos aquellos, puesto que ya pertenecerían a la categoría intuitiva de “instrumentos empíricos”. Esta abstracción que trata de ir más allá de la dicotomía que existe entre “teoría” y “modelo”, tiene en cuenta que, el segundo habría sido dejado lado en la tradición positivista, el cual privilegió a la teoría como unidad de análisis, aun cuando desde épocas tempranas se acusó tal sesgo importante en el contenido por parte de científicos interesado en filosofía (Hutten, 1952).

No obstante, se hace preciso un último supuesto para generar el panorama total de las bases de la abstracción, a saber, dar por hecho que *la filosofía de las ciencias es progresiva*¹⁰². Siguiéndose de este supuesto una serie de conclusiones básicas constituirán principios de construcción de la tesis de los IC y un trasfondo algo más completo que estaría siendo implicado por la Def.0.1.:

- (.A) “Pragmáticamente” guiada

La digresión 2, a saber, la declaración de que la salida a la paradoja de Smolensky no deviene de tomar partido ontológico, sino que deviene de la suspensión del juicio con respecto a un enfoque sobre el qué de la cognición, lo que permitiría una interpretación puramente epistemológica de ésta, es exactamente lo que Jilk et al. (2008a) postulan al insistir sobre dejar de lado los privilegios ontológicos. Al dejarse de lado los privilegios ontológicos, entonces las discusiones pueden centrarse en tomar lo mejor de cada lado – obviamente si es promisorio hacerlo de ese modo, como resultó con la arquitectura SAL, la cual, por ejemplo, sería una buena salida al problema del *symbol grounding* – para lidiar mejor con el fenómeno que se está investigando. Dado que el foco está en el fenómeno y considerar cualquier enfoque solamente en su faceta epistemológica, entonces con “pragmáticamente” guiada, quiero decir que toma en cuenta un aspecto predominantemente frente a los otros, el cual ya ha sido detallado en la sección (2.3.1.), a saber, el aspecto combinado de la tarea de predecir-explicar. Luego, estoy descartando la parte de la filosofía de las ciencias que trataba con las estrategias genéricas dentro de las disciplinas en general y en particular, para centrarme en algo más sencillo y común a lo que se ha conceptualizado hasta hoy: que todo sistema contenido en la ciencia redescubre y, en menor o mayor medida, predice-explica.

Luego, (.A) implica para Def.0.1.

Que un “IC” no puede ser separado de las metas de explicar y predecir. Sin explicación ni predicción un IC se transforma en un concepto vacío de contenido empírico. Tal tipo de IC “vacío” debiese ser considerado como parte de una categoría distinta de los IC bien definidos.

Que un IC es cualquier cosa que explique y prediga, por tanto, no hay diferencia entre las teorías y los modelos, ya que un IC es una abstracción que subsume parte del contenido de lo que se entiende por una “teoría” o un “modelo”.

¹⁰² El especialista ya habrá notado que he hecho uso de este supuesto aun antes de enunciarlo al, por ejemplo, desestimar la concepción heredada y comenzar construyendo teóricamente desde el historicismo y Quine. Sin embargo, lo he apuntado aquí explícitamente para que no quepa duda de mi posición.

- (.B) Pluralista explicativamente:

Con este trasfondo estoy volviendo a plantear exactamente lo que sintetizo en la recta final de la sección (2.2.), a saber, – y muy sintéticamente aquí puesto –: Primero, que el pluralismo explicativo es una postura epistemológica que reivindica que no hay una única, completa y comprehensiva explicación, teoría o modelo del mundo ni de cualquier fenómeno. Segundo, desde una perspectiva temporal absoluta el pluralismo es instrumentalista, pero desde la perspectiva temporal relativa y acotada propia de la investigación el pluralismo explicativo es realista sobre cierto “piso mínimo ontológico” que es más resistente a los cambios. Tercero, la postura pluralista no restringe las metodologías de investigación posibles, lo cual en términos prácticos puede involucrar la colaboración, aunque tal colaboración está sujeta a criterios pragmáticos. Cuarto, que la inconmensurabilidad es un hecho normal a cualquier disciplina. Quinto, y finalmente, el pluralismo explicativo no es anarquismo.

Luego, (.B) implica para Def.0.1.

Que no existe restricción respecto al cómo la redescipción de cualquier IC es formulada. En general, a partir de la Def.0.1. se sigue que cualquier relato, por muy absurdo que llegase a ser, es en potencia (obviamente si cumple con redescibir y predecir-explicar) un IC.

- (.C) Imposibilidad de “reconstrucciones racionales”:

En su trabajo clásico *Naturalized Epistemology* (1969), Quine considera que la epistemología tradicional puede ser dividida en dos programas (que son esencialmente análogos al trabajo de reducción de la matemática a la lógica), uno doctrinal y otro conceptual, teniendo ambas en común el intento de justificar el conocimiento sobre una base definitiva. Sin embargo, el programa doctrinal falla por el problema humeano de la inducción, mientras que el programa conceptual es criticado por el intento de trasladar conceptos teóricos a términos sensoriales. En particular, Quine critica el proyecto de “reconstrucción racional” de Carnap, donde justamente se persigue reconstruir las teorías sobre la base de términos observacionales, lógica y teoría de conjuntos. La alternativa quineana es la “epistemología naturalizada”, la cual es bastante amplia en su formulación, ya que simplemente afirma que “epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science” (Quine, 1969, 82). Por lo tanto, actualmente existen autores diversos que consideran estar tomando un punto de partida “epistemológico naturalizado”, pero que en el detalle de sus propuestas las divergencias son profundas – considérese Thagard (1988) versus Nersessian (2008). Considero válido el punto de vista quineano sobre las “reconstrucciones racionales” tipo Carnap, sin embargo, dado que “naturalizar”, en general “naturalismo”, tiene demasiadas versiones sobre qué significa (Flanagan, 2006, 430-1), entonces exclusivamente me quedaré con la idea de no cometer el mismo error de *propósito* de Carnap y, en general, del positivismo lógico.

Luego, (.C) implica para Def.0.1.

Que un IC no puede ser considerado un objeto filosófico válido para servir como principio general que separe entre “lo metafísico” y “lo científico”.

Que los IC no tiene como propósito la fundamentación del conocimiento. En general, es una herramienta epistemológica, sin ningún problema epistemológico clásico – origen del conocimiento, posibilidad, esencia, criterio de verdad, etc. – como meta.

Que los IC no son “naturalistas”, “materialistas” o se adjudican alguna etiqueta específica.

- (.D) Historicismo e integración entre el contexto de justificación y el de descubrimiento:

El positivismo lógico, específicamente Hans Reichenbach, hacía una distinción taxativa entre el contexto de justificación y el de descubrimiento. Según tal postura, la evaluación empírica de las teorías se atenía al análisis lógico, con las condiciones específicas bajo las cuales una teoría podía ser confirmada o descartada. Las teorías vendrían generadas axiomáticamente con una base de proposiciones observacionales y leyes. El descubrimiento científico, en contraste, no sería susceptible de análisis lógico y, por tanto, sería tarea de mera psicología estudiarlo sin consecuencias necesarias para la racionalidad científica. El descubrimiento sería un proceso psicológico y la justificación solamente sería posible de análisis lógico (Bechtel & Richardson, 2010, 3-4). En otros términos “(...) discovery is involved with the genesis of beliefs and scientific theories and hypotheses, whereas justification is involved with their evaluation: while the notion of discovery is descriptive, the notion of justification is normative” (Vassallo & Bianchi, 2010, 313). La filosofía de la ciencia posterior a Popper (para quien todavía es válida tal distinción), y el historicismo que tomará protagonismo en la segunda mitad del siglo XX, pondrá en duda tal distinción y la presente tesis considerará tal postura que no supone una diferencia taxativa entre ambos contextos como un principio sobre el cuál trabajar.

Luego, (.D) implica para Def.0.1.

Que para entender qué se entiende por “fenómeno empírico” se hace imperativo tanto un estudio de los contextos de justificación como del contexto de descubrimiento.

- (E) Continuidad las explicaciones de sentido común y las científicas:

Esta idea intuitiva puede encontrarse en manuales de filosofía de la ciencia clásicos (Nagel, 1961) y actualmente subyace a parte de la investigación en psicología centrada en el estudio del cambio conceptual, es decir cómo tales “explicaciones de sentido común” conformarían variadas *folk sciences* que darían lugar a las teorías científicas que van contra precisamente las intuiciones que tenemos del mundo (Keil, 2010). Considero que esta clase de evidencia, que muestra cómo esta dicotomía entre lo folk y lo científico que persiste al “adiestramiento” en ciencias (Shtulman & Vancarcel, 2012), puede ser considerada como sentando las bases de un programa investigativo que muestre cómo ambos fenómenos tomados como una continuidad darían sentido a cómo se ha ido desarrollando la ciencia.

Luego, (.E) implica para Def.0.1.

Que el IC más simple concebible es aquel que se da en el ámbito casi puro de un lenguaje natural con apenas un lenguaje técnico y el más complejo pareciera no tener un tope en su uso de lenguajes especializados.

- (.F) Subdeterminados por la evidencia:

El holismo confirmacional u holismo de Duhem-Quine, “(...) dice que los enunciados científicos no son vulnerables a las observaciones adversas por separado porque sólo conjuntamente, como una teoría, implican sus consecuencias observables.” (Quine, 2001, p. 55-6) Esto quiere decir que una observación afecta a la conjunción de proposiciones dentro de una teoría científica, por lo tanto para salvar la teoría basta retirar cualquier enunciado dentro del conjunto y no necesariamente aquel que parece directamente refutado por la evidencia empírica. A partir del holismo confirmacional, se obtiene lo que Quine denomina “subdeterminación de la teoría científica” que significa que si todos los acaecimientos observables pueden explicarse mediante una teoría científica global, entonces podemos esperar que resulten igualmente explicados mediante otro sistema del mundo en conflicto con el anterior. Todo esto viene dado porque los científicos inventan hipótesis que hablan de cosas que están más allá del alcance de la observación, por lo cual las consecuencias observables de las hipótesis no implican, inversamente, las hipótesis. Luego, en el presente trabajo la existencia de subdeterminación no será considerado un problema que solucionar, sino más bien un hecho de las ciencias con el cuál lidiar descriptivamente.

Luego, (.F) implica para Def.0.1.

Que dados dos IC que expliquen el mismo fenómeno, éstos tendrán el mismo comportamiento descrito por Quine para las teorías, es decir serán subdeterminados.

2.3.2.1. Definiciones sobre la estructura de los IC y su funcionamiento

Los principios de construcción sin definiciones claras y más apegadas a los objetos que se quiere describir, a saber, los instrumentos empíricos, no dejan de ser una mera idea preliminar y general, así como también la propia Def.0.1. de cómo funcionaría el medio conceptual de los IC para tratar con las relaciones interteóricas. Por ejemplo, de la cuestión (.A) la pregunta más inmediata que surgiría es ¿cómo toman los IC contenido empírico? Aunque he enfatizado que un IC parte de las generalizaciones constreñidas por cualquier grupo de métodos confirmatorios – dado el trasfondo pluralista explicativo – no es aún claro cómo se conecta un IC con un fenómeno o lo que usualmente se entiende por “fenómeno” que es el “aquello” que *a priori*, uno intuitivamente diría trasciende un IC.

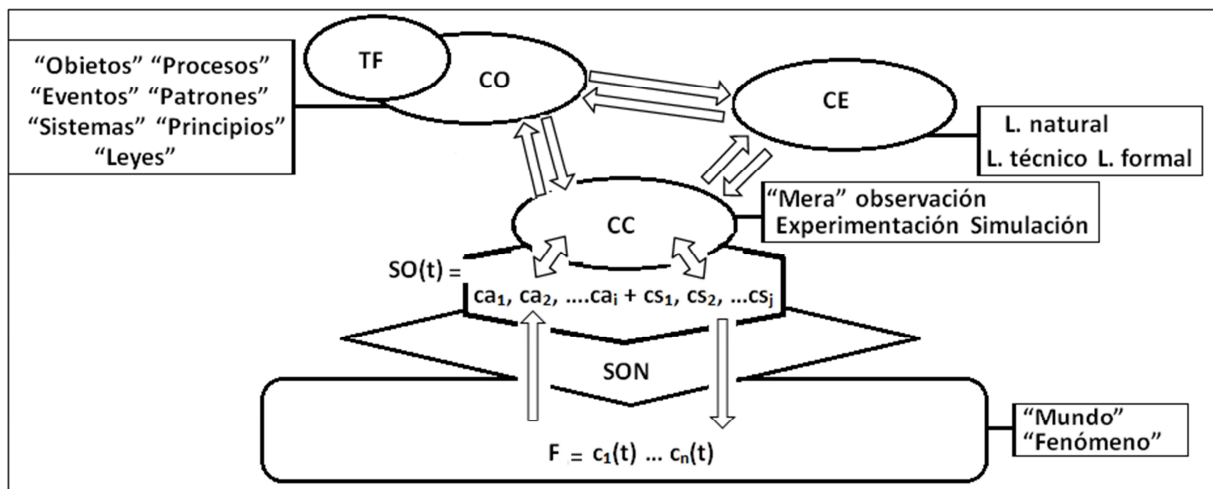
Del principio (.B), a saber, del propio pluralismo, en conjunción con el principio (.C), es decir, la afirmación de la imposibilidad de “reconstrucciones racionales”, la tarea más difícil es cómo dar con un criterio (o conjunto de criterios) que permita demarcar aquellas redescripciones que uno consideraría “científicas” de las no científicas. Ya que no es el fin invocar a lo normativo como una salida al sempiterno problema demarcatorio, entonces definiciones más claras sobre la tesis de los IC permitirán encontrar un mínimo prescriptivo que parece difícil de lograr. Sin embargo, como mostraré, no es imposible hallar una salida “suave”, aunque quizás la solución no sea del todo satisfactoria para quienes aún esperan una salida del tipo soñada dentro de la tradición de la concepción heredada.

Del principio (.D), a saber, al tomar el historicismo como base para el análisis, el lector avisado habrá caído en cuenta que, dado que los IC son históricamente situados, entonces sus redescripciones y contenido global no se pueden extraer unívocamente, sino que están sujetos a cierto grado de

hermenéutica. Esto es lo considero indudablemente cierto, y más que un obstáculo que deba ser salvado, considero que es un hecho que debe ser domesticado. Por tanto, como se verá en las secciones de ejemplificación (sección (2.4.) y sus subsecciones), mostraré como tal hermenéutica, gracias a la propia esquematización de los IC, estaría bien restringida.

Con respecto al principio de construcción (.E) su aplicación directa será postergada para futuras líneas de investigación que permitan fundamentar más exactamente un patrón de comparación normativo, que implique que entre el simbolismo y el conexionismo, el primero posee limitaciones que conllevan una menor profundidad explicativa-predictiva relativa que el segundo. Estas futuras líneas de investigación serán delineadas en la última parte de la tesis, dentro del sector que corresponda a las conclusiones (sección (3.2.)). Finalmente, la subdeterminación de las teorías científicas, o el principio (.F), que parece ser un problema dentro del ámbito de las filosofías de las ciencias desde que se postuló para quienes están ligados a la concepción heredada, considero que, si es tomado como un hecho descriptivamente útil, entonces me servirá de base para dar cuenta más exacta de una noción tal como “revolución”.

Entrando definitivamente en materia, y ya estipulados lo más explícitamente posibles, tanto mis supuestos más intuitivos (digresiones 1 y 2 en la parte I de la tesis), como el trasfondo filosófico de las ciencias en el cual pretendo encajar (sección (2.2.)) y los principios de construcción que me guían – y cuyo origen es una actualización de lo que se ha aprendido en filosofía de las ciencias – la imagen que servirá de apoyo visual para el lector para entender la relación de los conceptos que se irán presentando en la presente sección es la siguiente (FIGURA 13):



• 0. Definiciones especificando qué es un “IC”

Def.0.1. *Un artefacto o instrumento conceptual (IC) es todo aquel constructo conceptual que mediante redescrición, cumpla las metas de explicar y predecir un fenómeno empírico.*

Def.0.2. *Un IC es todo aquello que puede ser trasladado a una Estructura tipo IC*

0.2.1. *Un IC es todo aquello que pueda ser interpretado como poseyendo una Estructura tipo IC.*

0.2.2. *Todo aquello que tenga una Estructura tipo IC puede ser interpretado como un IC.*

Las Def.0.1. y Def.0.2. (con sus consecuencias) permiten entender que, a partir de aquí, al término “IC” daré dos acepciones diferentes: la Def.0.1. implica que una teoría o un modelo es un IC, por lo tanto decir “IC” reemplaza referirse al objeto “real” que es la teoría en conexión con una práctica científica y el mundo. La Def.0.2., 0.2.1. y 0.2.2. implican que por “IC” se aludirá a cualquier cosa que pueda ser interpretada como tal, por tanto aquí un “IC” es un objeto “ficticio”, en cuanto depende fuertemente de la subjetividad de alguien para ser construido, utilizado y comunicado al resto. La segunda acepción de “IC” utiliza a las teorías y modelos como objetos sobre los cuales hablar o, lo que es muy cercano, como objetos sujetos a *interpretación*, mientras que la primera acepción es un mero reemplazo que generaliza cualquier teoría, modelo, o grupos de éstos. El contexto a lo largo del presente trabajo bastará al lector para distinguir entre ambas acepciones. En general, la que usaré de modo más intensivo es la segunda, donde los “IC” servirán como una herramienta analítica de alcance *micro, meso y macroscópico*.

Def.0.3. Una Estructura tipo IC requiere los siguientes conjuntos componentes: {CO, CE, CC, SO, SON}

¿Cómo asegurarse que un IC siempre consta de estos componentes? Por los supuestos que constituyen la base que he construido para dar contenido a la Def.0.1., a saber, en tanto un IC es redescrípción, entonces ya contendría un lenguaje técnico no reducible al lenguaje natural (sección 2.3.1.). Por ende, me aseguro así que, si un cierto grupo de conceptos no cuenta suficientemente como redescrípción, entonces no podrá ser trasladado a una estructura tipo IC. En caso de existir redescrípción, entonces debe agregarse que tiene la meta de explicar-predecir, y recién entonces, con esas condiciones sería lógico que cumplirá la existencia del conjunto de componentes CO, CE, CC, SO y SON.

Es importante notar que cuando se diga que algo es una “redescrípción” no necesariamente será un IC, ya que puede estar construido sin las metas. Es por ello que la Def.0.3. no debe leerse olvidando que un IC es “más” que una redescrípción en tanto que está pragmáticamente guiado (principio (2.3.2.)(A)), pero es “menos”, en cuanto es un subconjunto del conjunto más amplio de las “redescrípciones”. En ciencias, es común el uso peyorativo de decir que “X es meramente descriptivo”, implicando justamente esto recién apuntado, a saber, que “X pertenece al conjunto de las redescrípciones, pero no de los IC”.

Def.0.4. Todos los conjuntos componentes de un IC son separables de tal IC

El pluralismo explicativo (sección (2.2.)), sintetizado en (2.3.2.) principio de construcción (.B)) implica la multiplicidad de conjuntos confirmatorios y cruzamientos entre ellos. La posibilidad de cruzamientos generará en sí mismos cambios “hacia arriba”, es decir, hacia el conjunto expresivo y el conjunto-ontología. La forma de compatibilizar tal cruzamiento existente en la práctica científica¹⁰³, y estipulado como definitorio del pluralismo explicativo, con un objeto abstracto que tenga como meta fundamental (me auto-cito del final de la sección (2.3.1.)) “*producir un sistema estable que sea tan abstracto que*

¹⁰³ Creo que es la norma, y no la excepción, lo que describí en la sección (1.2.2.) (B1), a saber, la existencia de hibridación, la cual he descrito brevísimamente también en (2.1.) con TOTE. Luego considero la hibridación o la pluralidad de hecho como extendiéndose en la historia de las ciencias cognitivas, no solamente post-polémica simbolismo-conexionismo, sino que desde su concepción como una “súper-disciplina” (Dale, 2008). Queda a un estudio histórico de las ciencias más detallado si el caso las ciencias cognitivas es extrapolable, no necesariamente al nacimiento de todas las disciplinas científicas, sino que al desarrollo y perfeccionamiento de los IC que componen cada una de ellas.

cualquiera sea la disciplina, ésta siempre contendrá constructos que se puedan descomponer según mi propuesta de análisis” es suponer la separabilidad de todos los conjuntos componentes. De no suponerse tal separabilidad, entonces me comprometería demasiado normativamente de un modo que no parece justificado si uno observa detenidamente cómo progresan los campos científicos.

0.5. *Un IC existe como una estructura esencialmente inestable.*

Este punto no es una definición propiamente tal, sino que deriva de la propia separabilidad, ya que, si no es posible establecer que cierto conjunto componente pertenece a un IC dado, entonces ¿qué es un IC? La respuesta es una “estructura esencialmente inestable”, lo cual sería equivalente a decir que es un “esquema de variables que es saturada por cualquier cosa que cumpla desde la Def.0.1. hasta la Def.0.4”

0.6. *La individuación mediante fijación de conjuntos componentes en un IC siempre será convencional.*

Esto traduce el principio (D), o sea, que un IC pueda ser descrito depende de lo que puede ser interpretado desde la práctica situada históricamente, desde un intervalo de tiempo largo, o una época completa, a un periodo estrechísimo de tiempo tomando en cuenta un puñado aislado de investigadores. La consecuencia 0.6. es la que subsume lo que he adelantado como la presencia de cierta “hermenéutica” tácita en mi tesis de los IC y esta consecuencia es, obviamente, una extensión de la consecuencia anterior, a saber, el que un IC no tiene contenido estable a través del tiempo. Así pues lo que aquí estoy desarrollando, más que servir como guía para “encontrar”, es una guía para “interpretar”.

- **1. Definiciones ligadas al “SO” y el “F”**

Def.1.1. *Un Sistema-Objetivo (SO) es el conjunto de características abstraídas (ca) y supuestas (cs) desde un Fenómeno (F).*

La forma de mostrar que existen tales dos tipos generales y básicos de características definitorias para la noción de SO es apelando a dos hechos transversales en la investigación científica: Primero, todo modelo que uno pueda encontrar *abstrae y descarta características debido a la meta/enfoque del investigador*. Y segundo, dadas las limitaciones materiales inherentes al desarrollo histórico contingente científico, entonces todos los modelos *suponen características*.

Def.1.2. *Un Fenómeno (F) es cualquier cosa que funcione como dominio de características ca y cs.*

Esta definición y la anterior suponen que un SO metafísicamente accede a cierto grupo de características propias del fenómeno que permite diferenciar a las “abstraídas”¹⁰⁴ de las “supuestas”, aun cuando es perfectamente sensato filosóficamente decir “¡cuidado! las abstraídas son también ‘supuestas’ en cierto modo, ya que nunca accedemos al fenómeno”. Empatizo con una opinión así, pero tal postura siempre

¹⁰⁴ “Abstraer” usaré genéricamente para todas aquellas características que surgen mediante simplificaciones, abstracciones, idealizaciones, sesgo, etc. Existe muchísima literatura que trata de delinear las diferencias entre estas diferentes acciones. Aquí estoy sacando solamente el factor común a todas éstas, es decir, el que quienes “simplifican, idealizan, etc.” *suponen que sus características abstraídas realmente se corresponden con alguna característica del fenómeno que tratan de capturar mediante su SO perteneciente a su IC en construcción.*

estará situada desde una perspectiva de $(t+1)$ donde es posible decir algo así. Luego, estoy justificando este acceso sólo en la medida que tomamos un periodo pequeñísimo t , en el cual tiene sentido diferenciar entre abstraídas y supuestas, donde las primeras tienen cierto “estatus ontológico especial” ante las segundas, aunque en la práctica sirvan de igual modo para redescubrir y predecir-explicar.

Mirando específicamente F en el esquema al principio de la sección, metafísicamente es posible plantearse preguntas como: Dado cualquier F las características o propiedades ¿son estables o no en el tiempo? Si la respuesta es sí, entonces no tiene sentido decir que $c_1(t) \dots c_n(t) \neq c_1(t') \dots c_n(t')$ Si la respuesta es no, entonces tal desigualdad es válida y estaría comprometiéndome con asumir que varía en el tiempo. Puede extenderse la duda metafísica diciendo “pero, quizás, no es el fenómeno el que cambia, sino que nuestra percepción sobre éste ¿cómo saber si es el fenómeno el que cambió o más bien nuestra percepción?” Piénsese la cuestión desde este ángulo, todas las personas que hoy viven en el mundo no son las mismas personas que hace 100 años habitaban el planeta – suponiendo que el promedio de expectativa de vida no sobrepasa tal límite – ¿cómo podemos entonces estar seguros que el mundo es el mismo? Podemos estar medianamente seguros por la cadena de interacción entre nuestros antepasados y nosotros, pero mientras más extendemos el intervalo temporal, es meridiano que nuestras categorías metafísicas han mudado sus elementos de modo que no podemos estar seguros de cómo ir desde un punto muy alejado en el tiempo hasta el presente respecto a su contenido.

La única salida a ésta y otras clases de problemas metafísicos derivados, es suponer que, primero, – y de manera básica –, tienen sentido tales preguntas, aun cuando no puedan dirimirse empíricamente, y, segundo, adhiriendo a cierta ontología lo más cercana a lo que hemos aprendido del mundo y de nosotros mismos que nos permita seguir haciendo ciencia tranquilamente como la continuamos haciendo, es decir sostenida por perogrulladas metafísicas tales como:

- Que es suficientemente estable el mundo como para no tener que buscar todos los días donde estaban las mesas y sillas o un papel tornasol para constatar que el H₂O mantiene el mismo pH. Y que es suficientemente dinámico como para que tenga sentido preguntarse todos los días cómo va estar el clima mañana o preguntarse si el agua que bebimos el mes pasado mantiene las mismas propiedades químicas que este mes.
- Que somos lo suficientemente estables como para asumir que si ya adquirimos cierta habilidad cognitiva, como que somos capaces de leer y escribir hoy, entonces seremos capaces de hacerlo mañana. Y que somos lo suficientemente dinámicos como para no estar seguro si quienes son nuestros amigos más íntimos continuarán siéndolo.

Esta breve digresión filosófica muestra que cualquier F , por tanto, es *siempre discutible conceptualmente* (o lo que considero sinónimo “filosóficamente”), empero, asumiré sin mayor fundamentación que su contenido exacto no puede ser determinado. Además esta muestra de divagaciones la presento para exponer aquellas cuestiones que *no pretendo ni plantear tácitamente en mi tesis de los IC, ni responder mediante su exposición más detallada*. La razón fundamental de distinguir entre ca y cs es que son útiles para cuantificar la capacidad de adelantarse de un IC y su respectivo SO a que ciertas de las cs se transformen en el futuro en ca , o sea, en características que pueden ser extraídas desde un F por un IC futuro donde el progreso científico permita confirmarlo. Por ejemplo, se alaba el modelo de Hodgkin y

Huxley (1952) precisamente porque, en un periodo donde nada se sabía de los mecanismos de excitabilidad de la membrana neuronal, fue capaz de, exclusivamente basado en la “fenomenología neuroelectrofisiológica”, guiar a la predicción sobre la existencia canales iónicos sensibles al voltaje diferentes para sodio y potasio (Catterall et al., 2012). En este caso clásico por “características supuestas” se trata de “mecanismos” eléctricos, pero una *cs* o *ca* puede ser *literalmente cualquier cosa*. Puede ser un patrón supuesto, una ley abstraída, etc. Para visualizar esto, considérese el modelo de Volterra, donde las *ca* son la cantidad de individuos y las tasas de nacimiento y “consumo del predador”. En este modelo *F* es una relación entre dos especies, mientras que el *SO* que se construye lo hace a partir de los *datos abstraídos* – si es que se basan en estadísticas del fenómeno respectivo – o bien supuestos – en caso que no se conociese en ese momento los datos, dado que se tratase de una especie nueva con muy poquísimos datos – y los primeros funcionarían como “características abstraídas”, mientras que los segundo, valga la redundancia, funcionarían como “características supuestas”.

Esto lleva una idea central: puesto que la tesis de los IC es interpretativa, luego un SON (Cfr. Def.5.1.) variará considerablemente aun cuando se trate del mismo modelo. Por ejemplo, cierto SON puede buscar predecir tanto la evolución de la relación ecológica, como tratar de encontrar el valor de una tasa desconocida. Luego, al ser ambos SON diferentes, darán lugar a IC diferentes, aun cuando deriven de un mismo modelo matemático. Esto, no solo muestra que un modelo matemático no es idéntico a un IC, sino que muestra que el análisis mediante IC es adaptable para dar cuenta de este fenómeno de ligeras variaciones “dentro” de un mismo “modelo-marco” individuándose IC de distintos grados de abstracción.

Def.1.3. Una característica abstraída (ca) es cualquier concepto trasladado a partir de lo que se considera que es F en cierto momento dado.

1.3.1. Una ca es cualquier concepto relevante para un SO.

La noción de “relevancia” solamente tiene sentido completo cuando se ha desarrollado mejor la noción de Sistema Objetivo Nicho en las definiciones número 5. Un buen *SO* es aquel que logra capturar aquel conjunto de *ca* y *cs* de modo de “producir” el SON deseado o, lo que es equivalente decir, un conjunto de *ca* y *cs* son relevantes cuando generan un SON deseable. La noción de “deseabilidad” también será mejor comprendida cuando se trate SON (Cfr. consecuencia 5.1.1.)

Def.1.4. Una característica supuesta (cs) no puede ser atribuir como perteneciente a F en cierto momento dado.

La diferencia entre “que es *F* en cierto momento” (Def.1.3.) y “no atribuible en cierto momento dado” (la presente definición) es graficada en el esquema al principio de esta sección en que las *ca* “suben” desde *F* al *SO* y las *cs* “bajan” desde el *SO* a *F*, es decir son expresamente adscritas como hipótesis.

1.4.1. Una cs es cualquier concepto relevante al mismo nivel que una ca para SO.

1.4.2. Una cs queda ligada a una o un conjunto de ca en cuanto componen un mismo SO.

Esta consecuencia es importante dejarla derivada explícitamente y es ejemplificada con el mismo caso del modelo H-H, donde los *cs* no pueden ser disociados de los *ca*, ya que ambos pertenecen a un mismo

SO. Si se variaran, de un modo en que un cs se eliminase, entonces se transformaría el SO. No es así el caso cuando un cs pasa a ser una ca como he ejemplificado antes.

1.5. *Un F se relaciona mediante un SO con el resto de los conjuntos componentes del IC.*

1.5.1. *Un F nunca es saturado por sólo un IC.*

Estas dos consecuencias implícitas en Def.1.1. y Def.1.2. permiten cerrar una comprensión más detallada de aquello que es confundido fácilmente en contextos donde no se percibe la sutileza que constituye la generalización científica, a saber, la creencia profundamente arraigada de que la relación desde un fenómeno a una teoría es de 1 a 1, donde se produce una teoría para luego ser refutada si no se ajusta al fenómeno. En realidad, *la "teoría" no constituye un cuerpo estático ni homogéneo, sino que ocurre en términos de un proceso abierto que convencionalmente es reificado y homogeneizado de manera de hacerlo pedagógico al lego y trasladable desde su contexto de nacimiento a otros donde se quiera extender.* Luego, dentro de la tesis de los IC, la existencia de "anomalías" en la predicción se supone como presente *ab initio* – solamente que no inmediatamente descubiertas –, justamente porque *la ciencia es en sí misma anómala*, es decir, está construida sin lograr saturar el fenómeno, explicándose tal no saturación en que un IC se levanta sobre un SO, el cual es suma de abstracciones y supuestos.

- **2. Definición ligada a "CC"**

Def.2.1. *Un Conjunto Confirmatorio (CC) toda colección de procedimientos para conectar un SO con un Sistema-Objetivo Nicho (SON).*

¿De qué modo un IC cualquiera hace contacto con la realidad? La cuestión de la confirmación ha pasado por las etapas donde lo científico era definido por un artefacto conceptual único, la teoría, y por una definición estrecha de ésta, un sistema deductivo de proposiciones, hasta las etapas donde la proposición dejó de ser la unidad mínima de confirmación y se llegó a afirmar que "the unit of empirical significance is the whole of science" (Quine, 1951, 39).

Dado que no requiero distinción analítico-sintético, puesto que no persigo reconstrucción axiomática, entonces *cualquier cosa arbitrariamente puede considerarse "observacional" y cualquier otra "teórica"*, preservando la posibilidad de análisis de las teorías, que en este caso han derivado en la abstracción de los IC, los cuales poseen conjuntos componentes en una relación cercana entre ellos, sin embargo aceptando su individuación mediante interpretación. Luego, más allá de si, a partir de una metáfora donde los artefactos conceptuales son análogos a una red que intuitivamente hace contacto con puntos-amarre de la "realidad", propongo precisamente que lo que funciona como "realidad-dominio" es un SO, el cual es claro y distinto en cada caso, dado un margen donde se asume lo inexacto como hecho. A su vez lo que funcionará como "realidad-recorrido" o *explanandum* será un *SON bien delimitado*, ya que de no ser así, entonces la empresa explicativa-predictiva misma no cumpliría el requisito básico de establecerse adecuadamente, generando que el IC no se constituyese como tal al fallar la Def.0.1. El modo de conectar cada SO a su SON es cualquier CC, el cual es específico a cada IC y, por tanto, por el término "procedimiento" presente en la Def.2.1., no necesariamente apelo a un procedimiento material, sino que *puede ser cualquier operación o método de cualquier tipo sobre un SO para llegar un SON.*

3. Definición ligada a “CE”

Def.3.1. Un Conjunto Expresivo (CE) es el conjunto de expresiones formales y no formales (en cuyo caso constituirían un lenguaje técnico) que permiten referirse al SO y al CC.

Al contenido “duro” de cualquier IC se accede mediante la comprensión de su CE, siendo éste lo que permite “manipular”, tanto el SO, como hablar del CC. Un IC idealizado en el modo de la concepción heredada es puro CE sin CO, pero tal cosa no existe, ya que cualquier lenguaje, por más técnico que sea, requiere cierta ontología mínima que, subsecuentemente se transformaría en el conjunto componente CO dentro de tal IC. En este caso específico las flechas de ida y vuelta que esquematizan la relación entre CE y CO no deben ser tomadas como equivalentes a las flechas ida y vuelta que caracteriza las relaciones CE-CC y CE-SO, ello porque algo desde la pura ontología no puede intervenir directamente en el CE, sino es mediante una relación que va a través del SO, es decir, en cuanto un elemento foráneo de CO se integra y plantea la necesidad de tantear una variación en el SO que está previamente hecho. Luego, si llegase a ser necesario, tal cambio se traducirá en un cambio en CE.

- **4. Definiciones ligada a “CO”**

Def.4.1. Un Conjunto-Ontología (CO) es el grupo de expresiones formales y no formales que refieren a todos los conjuntos componentes dentro de un IC.

Ya que un IC contiene cierto compromiso metafísico al redescibir un F, entonces tal redescipción puede ser traducida en parte a lo que denominaré CO. Este compromiso puede deducirse, por ejemplo, si se piensa en qué consistiría una descripción “pura”; “It is doubtful that pure description is possible. Presumably pure description would consist in presenting nothing but sense data each in complete isolation from every other.” (Miller, 1947, 311). Aun cuando fueran concebidos tales *sense data* como capturados perfectamente mediante lenguaje natural, resulta contraproducente a estas alturas de la ciencia suponer que ésta utiliza el lenguaje natural para conformar su CE en sus respectivos IC, sino que en mayor o menor grado un CE se compone por cierto lenguaje técnico o formal correspondiente a cierta disciplina, de modo que tal redescipción encaje en un IC cuyo desarrollo esté guiado por hipótesis

Este CO serán aquellas entidades (o “meta-”entidades – principios universales, leyes, patrones, sistemas, etc.) cuya tarea sea unificar la *fenomenología* que se deriva de los otros conjuntos componentes y que se vayan definiendo y confirmando a un nivel *operacional*. Por “operacional” pienso en la noción de Bridgman (1991 [1927]), donde se desprende que, en jerga metafísica, no existe al final ninguna cosa “en sí” o “esencial”, sino que todo se define *relacionalmente* siempre en “respecto a”, y, en el caso operacionalista, se define mediante la obtención metódica a través de una operación bien definida una magnitud que saturará una variable dentro de una función o ecuación en un modelo o teoría. Tal nivel operacional es el límite máximo al cual aspira la ciencia. Por ejemplo, cuando se revisan manuales introductorios a ciencias de los materiales es posible observar que todas las propiedades de los materiales son definidas operacionalmente (v. gr. propiedades térmicas, mecánicas, etc.) y que, a su vez, aunque composición y estructura parecen no definidos operacionalmente, ello es porque a nivel de materiales, los átomos, moléculas y sus arreglos son tomados como “esencias-primitivas”, las cuales en otras disciplinas a su vez son operacionalmente definidas (Callister & Rethwisch, 2014; Tilley, 2013).

4.1.2 *Un CO también es el grupo de expresiones formales y no formales que refieren al F que es dominio del SO.*

Mientras el grupo de expresiones CO que refieren al “interior” del IC, parecen susceptibles de individualizar caso a caso, dado que consistirían en aquellas que se subentienden son mínimamente necesarias unificar la comprensión de los conjuntos componentes de un IC, existe otro grupo de entidades del CO que no se tiene, dentro de la tesis de los IC, por muy seguras – ya siendo esta tesis bastante aproximativa en su naturaleza –, a saber, aquellas expresiones que tratan de F. El problema con éstas es que, dado que por definición no encajan dentro de la estructura del IC, entonces no hay forma de establecer hasta qué punto dos o más IC están tratando de lo mismo, al no existir la posibilidad de mapeo entre ambas estructuras de IC al estar los respectivos F “afuera” de tales estructuras.

Def.4.2. Un Trasfondo Filosófico (TF) es aquella parte de un CO cuya presencia no agrega nada a la conexión entre los conjuntos componentes de un IC, excepto SON, y SON.

Quizás el lema sea aquí para diferenciar TF de CO sea *dudar de todo aquello que se pueda dudar, pero no dudar de aquello que pragmáticamente no podamos desechar*. Aquello que no se puede desechar, es precisamente lo que fue apuntado en el comentario de la Def.4.1. y lo que puede ser desechado, o separable, es fácilmente visible en los trabajos de filosofía de las ciencias, no necesariamente de los filósofos, sino que también de los mismos científicos, donde se suponen que ciertos principios de amplio alcance son siempre necesarios para el funcionamiento explicativo-predictivo acotado de sus IC. Por consiguiente, la existencia de un TF estará siempre presente sobre cualquier IC, ya que se trata de una empresa que acometen también los propios especialistas cuando tratan de dar con una imagen completa del mundo desde su propio IC, extendiendo innecesariamente las conclusiones que, en otro contexto, están bien delimitadas. Sin embargo, al lema que sirve para el análisis práctico de separar aquello que es TF de CO, se puede agregar un rasgo de aparición mayoritaria encontrado en aquello que constituye parte de un TF, a saber, el *apriorismo* en la formulación del contenido del TF.

No obstante, entre la separabilidad y el apriorismo, la separabilidad es el rasgo más interesante, ya que permite distinguir aquello que sería algo así como el “mito útil”¹⁰⁵, de aquella parte que resulta inútil en la práctica, ya que, puede darse el caso, que un TF sea útil parcialmente en cuanto permite al investigador cognitivamente dar con el contenido definitivo de un IC, por ejemplo con su CE y CO, pero luego de ello deja de ser relevante. En otras palabras, lo que se asume es aquello que redescubre y nos da las mejores chances de explicar-redecir, pero tal CO, aun cuando posee una historia y un TF del cual deriva, no se necesita asumir homogéneamente con el mismo grado de compromiso. *Los conceptos en un IC “concreto puntual”¹⁰⁶ funcionan ahistóricamente y, por lo tanto, un TF puede ser descrito como adscrito heurísticamente (grosso modo “epistemológicamente”¹⁰⁷) o fundamentalmente (grosso modo*

¹⁰⁵ La expresión de “mitos útiles” nace en el contexto de hablar de, por ejemplo, la clase de objetos ideales dentro de las teorías físicas o químicas (“gas ideal”, “superficie sin roce”, “choques elásticos”, etc.). En términos de Quine, son “deliberadamente un mito, útil por la plasticidad, la belleza y la corrección sustancial con las cuales retratan ciertos aspectos de la naturaleza, aunque leídos radicalmente falseen en otros aspectos dicha naturaleza. [Son útiles] también por la simplicidad que permite a los cálculos.” (Quine, 1968 [1960], p. 258-9)

¹⁰⁶ Tales nociones se definirán en la sección siguiente, la (2.3.2.2).

¹⁰⁷ No toda adscripción epistemológica es heurística, pero sí toda adscripción heurística es epistemológica.

“ontológicamente”), siendo en este segundo caso que tal adscripción fuerte termina restringiendo la flexibilidad definitoria de los conjuntos componentes del IC, eventualmente estancando la investigación.

Dos casos de TF bien diferenciados en ciencias cognitivas – resabios de la filosofía pasada –son, por un lado el innatismo “platónico” y, por el otro, el empirismo “aristotélico”. El primero es un innatismo extremo que nadie hoy en día defiende, pero que si se lee con cuidado la literatura cognitiva muchas veces se utiliza como hombre de paja por parte de los enfoques alternativos, mientras que el otro es un tipo de empirismo que tampoco nadie defiende, achacándose tal TF a investigaciones contemporáneas.

5. Definiciones ligada a “SON”

Def.5.1. Un Sistema-Objetivo Nicho (SON) es cualquier cosa sobre lo cual se evalúa el éxito del resto de conjunto de componentes del IC.

5.1.1 SON está determinado por el enfoque de los investigadores el cual construye SO (el conjunto de ca y cs) sobre cierto F.

5.1.2. SON contiene ca y cs de algún modo que permita conectarlos (“de regreso”) a F.

Vale la pena repetir en este contexto, lo dicho al cerrar el grupo de definiciones número 1 que versaban sobre SO y F, a saber, que la existencia de errores en la predicción de un fenómeno se produce desde el comienzo justamente porque el SO no logra saturar el fenómeno, explicándose tal no saturación en que un IC se levanta sobre un SO que permite la construcción del SON respectivo, funcionando este último, recurriendo a anterior terminología, como *explanandum* (Hempel & Oppenheim, 1948, 136-7).

Por ende, lo que se entiende usualmente por “explanans”, en el contexto de la tesis de los IC se transforma en un conjunto de componentes interconectados, a saber, un SO, un CC, un Conjunto Expresivo (CE) y un Conjunto Ontológico (CO). No obstante, y aquí hay otro truco, que es importantísimo notar: *el SO juega una posición doble, es decir puede ser tanto explanandum, en cuanto sus componentes son parte del SON, pero también es parte del explanans, en tanto que ca y cs son la base para que el resto de componentes – CC, CE y CO – “jueguen” con ellos de modo de llegar al SON. SON será un explanandum que, contrario a la noción más ingenua, ya estaría tácitamente “dentro” del IC y no “afuera” de éste, como tradicionalmente se espera en las descripciones más simplistas de las ciencias, donde el fenómeno se mantiene impertérritamente aparte de las teorías y modelos que lo explican-predicen.*

Por lo tanto, SON está incrustado en la estructura del IC cerrando *un proceso altamente circular* si se lo mira con detenimiento; F es abstraído en ca y transformado mediante cs en un SO, el cual es “manipulado” de modo que CC, CE y CO, cada uno de los conjuntos de mayor a menor relevancia, permiten llegar al SON, el cual funcionará como cuasi-F. SON será aquel que usualmente presentará “leyes” o “principios”, mientras que F será el que presentará “regularidades” y, al decir esto, nótese que estoy asumiendo el carácter instrumental de fondo que está entre las características conceptuales del pluralismo explicativo definido en (2.2.). No obstante, hasta aquí he supuesto algo que atraviesa todo el conjunto de definiciones desde 1. a 5.; el *enfoque de los investigadores*, el cual, en sí mismo, es invisible desde la perspectiva interna del IC, es visible desde un análisis hermenéutico como el que estoy

haciendo, y solamente puede ser puesto en entredicho, no directamente, sino que frente a otro IC, el cual tácitamente contendrá otro enfoque, al contener otro modo en que el SO y SON son pensados¹⁰⁸.

2.3.2.2. Identidad, dinámica y categorías de los IC

La subsección anterior solamente ha especificado una noción de los IC que está “abierta” o no restringida, es decir donde su identidad va mutando de acuerdo a un contenido variable sobre una estructura cuya interacción interna permite que la conexión entre SO y SON se cumpla. Los IC planteados como están hasta la sección (2.3.2.1.) son procesos discretos cuyos elementos varían de acuerdo a las necesidades que se van generando en el tiempo, *donde intervienen con influencia desconocida variables externas e internas*. Las necesidades externas tendrían un cariz tanto social como psicológico, además de las propias necesidades empíricas, o aquellas que derivan de la no saturación de los fenómenos que se intentan ir explicando-prediciendo de manera más exitosa. Las necesidades internas serían aquellas presiones teóricas propias de la misma “inercia” de los IC más desarrollados, por ejemplo, surgidas de la necesidad de transformar características supuestas a características abstraídas.

La forma de individuar un IC – de modo de otorgarles una identidad convencional que permita su estudio meta-científico – es, dado el principio constructivo (.D), recurrir a la historia misma de cómo se desarrollan los IC precursores y elegir arbitrariamente un espacio de tiempo para acotarlos. Cuando se recurre a esta estrategia, entonces lo primero que se observará, naturalmente es la consecuencia:

6.1. Dada Def.0.5. y Def.0.6. la inestabilidad de un IC transcurre según fases que constituirían la dinámica de los IC (La FIGURA 14 abajo esquematiza tales fases)

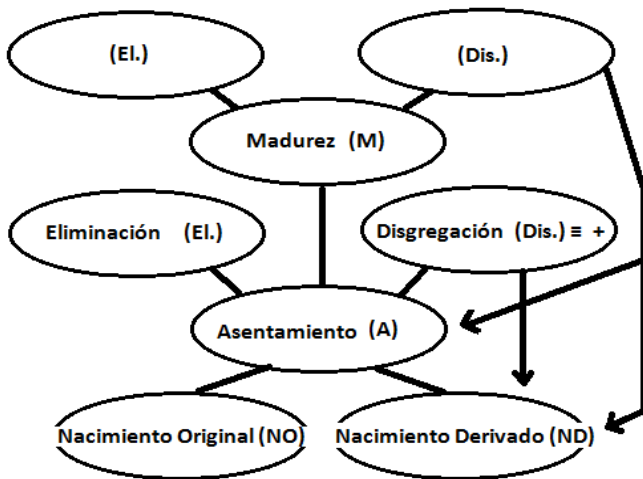


FIGURA 14; Una observación importante para el especialista: las fases kuhnianas, que tienen a la noción de paradigma como central, tienen al todo de una disciplina como foco de análisis. Mi foco trata a la disciplina como algo que es a posteriori de los IC, que emerge a partir de estos (véase la Def.7.3. y consecuencia 7.3.1.), por lo tanto, conjeturo que es imposible generar una traducción exacta de lo que presento aquí y la filosofía de la ciencia historicista como la de Thomas Kuhn. Esta conjetura requiere un trabajo aparte que no llevaré a cabo en la presente tesis.

¹⁰⁸ Creo que un caso concreto puede verse en, por ejemplo, cuando se dice que la mecánica newtoniana es desafiada en su noción de tiempo absoluto frente a la teoría relativista. Sin embargo, el modo en el cual se presenta este contraste muchas veces recurre a la oposición directa de los supuestos que guiaron respectivamente a Newton y Einstein. Esto sería un error, ya que, trasladándolo a la tesis de los IC, tal contrastación utiliza solamente los componentes CO y TF, cuando el modo de contrastarlos es desde SO y SON, mostrando como el SON relativista modificaría el SON clásico, compartiendo elementos de éste. Luego, si se observa estrictamente en estos niveles, entonces, podemos extraer cómo se enfrentan los enfoques indirectamente. Esto es una conjetura, ya que para demostrar como válido este punto se requeriría desarrollar con detalle SO y SON de cada teoría, sin embargo, el punto me parece intuitivamente correcto.

6. Definiciones ligada a la dinámica de los IC

Def.6.2. Existe una fase de nacimiento original definida cuando los elementos que constituyen los conjuntos componentes de un IC son utilizados por vez primera.

Def.6.3. Existe una fase nacimiento derivado definida cuando los elementos que constituyen los conjuntos componentes de un IC ya eran parte de anteriores IC.

El caso más intuitivo y emblemático de nacimiento cuasi-original es el de la mecánica newtoniana, no tanto por la originalidad de la ontología – ya que, por ejemplo, es afirmado que Kepler estuvo muy cerca de llegar a plantear la gravedad exactamente como Newton lo lograría hacer (Gentner, 2002, 26) –, sino que por su CE que involucraría el desarrollo del cálculo infinitesimal, el cual es perfeccionado independientemente a Leibniz.

Un nacimiento derivado se ha dado en casi todos los casos donde el investigador ya está circunscrito a un círculo académico, por ende donde su labor ya tiene un camino relativamente prefijado por los IC en boga. Es por esto que, en la práctica, gran parte de los IC son de nacimiento derivado, ya que es bastante difícil dar con un IC que cumpla todas las definiciones básicas, y que no se tenga vestigios en su contenido que puedan ser ligados a IC anteriores¹⁰⁹. Conjeturo que, a causa de la globalización y las cada vez más comunes tecnológicas de la información, el hallazgo de nacimientos originales de IC pertenecerá exclusivamente a la historia de las ciencias y ya no al ámbito de la investigación en curso.

Def.6.4. Existe una fase de asentamiento definida por la aparición de IC abstractos-generalizados.

Es la fase que ya presupone un éxito básico de cierto IC para explicar-describir, dado que si no fuera así el caso, entonces no se podría hablar siquiera que atravesó exitosamente la fase de nacimiento. En esta fase de asentamiento el énfasis está en cuánto impacto social logra tener tal IC específico, de modo de generar otros IC sobre éste y, por otra parte, de cuánto logre expandir y prolongar su éxito empírico.

Si el arrastre es suficientemente alto, entonces el asentamiento generará una fase intermedia entre el asentamiento y la madurez donde se producirá el primer intento de *naturalización, tanto del contenido del IC abstracto derivado del IC concreto originario como del IC concreto originario en la disciplina*¹¹⁰. Esta fase puede tardar unos pocos años en completarse – como fue caso de los polos eléctricos que rápidamente tomó el lugar de hecho natural definitorio de la electricidad (Steinle, 2006, 192) – puede tardar un poco más, pero ser fácil debido a cuestiones sociológicas, como la falta de competencia (como pasó con la noción de gravedad newtoniana) o puede nunca terminar de darse debido a la presión de otros IC que tratan de asentarse sobre el mismo F generador de SO pertenecientes a los diferentes IC en competencia, siendo éste el caso del simbolismo. Más sobre esto último se tratará en la sección (2.5.i.)

¹⁰⁹ Decidir si un elemento de un conjunto componente de un IC está en continuidad con otro es altamente hermenéutico, es por ello que me he cuidado al afirmar que la mecánica newtoniana se trataría de un nacimiento “cuasi-original”.

¹¹⁰ Dada la interdependencia fuerte en la construcción teórica de la tesis de los IC es difícil transformar ciertas ideas sin terminar deformando la presentación. Es por ello que en esta frase he adelantado definiciones que serán tratadas en el punto 7.

La dinámica es el lugar definitivo dónde se produce el tamizado que permite a “lo científico” ser reconocido como tal. Específicamente, al atravesar la etapa de nacimiento, etapa esencial para que un IC demuestre cierto éxito explicativo-predictivo, y completar también exitosamente la fase de asentamiento lo que permite dirimir lo que es científicos de aquello no científico, aunque, de una manera mejor dicha, es lo que permite dirimir aquello que se integra a una disciplina, de aquello que infructuosamente puede terminar siendo subsumido por otros IC. Luego, al no haber un IC unívoco para generar conocimiento útil y novedoso que trascienda a las diferentes disciplinas, entonces lo máximo a lo que se puede apuntar en tal aspecto es a ciertos métodos comunes (CC) a subdisciplinas científicas y esto que es limitado, se complementaría con el tamizado que se deriva de la dinámica de los IC que deben atravesar desde el nacimiento a un asentamiento completo.

Def.6.4. Existe una fase de madurez definida por la naturalización de los elementos del CE y el CO de un IC.

Quizás la manera más intuitiva de entender con esta fase es suponer que un IC ha logrado cierta hegemonía en cierta subdisciplina, al punto de subsumir a ésta en un IC “abstracto generalizado”. Sin embargo, este aspecto es secundario al de la *naturalización definitiva*, la cual muchas veces deriva en el fenómeno de *popularización*, generando que la competencia contra un IC maduro sea muy complicada para IC que esté en las fases más tempranas como nacimiento y asentamiento.

Ahora bien, que un IC esté maduro dentro de una disciplina, no implica que la disciplina funcione de modo maduro. La condición para lo segundo tiene que ver *con investigadores de actitud pluralista lo cual permita la “colaboración”, una “competencia constructiva”* o, como mínimo, cierto agnosticismo saludable al monismo explicativo. Una disciplina puede tener un IC maduro, pero ello puede ser contraproducente a largo plazo, por ejemplo, como sucedió con Le Verrier, donde parte de sus premisas de razonamiento habría sido que la mecánica newtoniana es incuestionablemente cierta, lo cual le llevó a plantear la existencia de un octavo cuerpo, Neptuno (Hanson, 1962, 363-4) el cual fue confirmado después. Sin embargo, tal misma premisa llevó a Le Verrier a postular un planeta intramercurial, Vulcano que explicase irregularidades de la órbita de Mercurio, pero en este caso el tiempo no le dio la razón, haciéndose patente la necesidad de un nuevo IC. Sería la aparición de la teoría general de la relatividad lo que se diera cuenta de las anomalías observadas en Mercurio (Clemence, 1947) sin expandir la cantidad de planetas. Considero que una dosis de pluralismo *podría haber servido* para que esto no se hubiese producido (nótese una pequeña dosis de normatividad aquí).

Existe hasta aquí una incógnita que debe ser resuelta exactamente en este punto, a saber, una noción más concisa de “naturalización”. Definida de manera amplia existe cierto consenso que “naturalizar” es *el paso de una noción desde lo no científico a su conceptualización dentro de un IC asentado en una disciplina científica*, habiendo al menos tres tipos de naturalización:

1. El paso desde el LN (o desde el “sentido común”) a ser parte del criterio de un instrumento empírico o IC. El ejemplo clásico es el de “rayo”.
2. El paso de una noción de tipo formal o de un instrumento formal a una de instrumento empírico. Ej. “vector” como se apuntó en el transcurso de la primera parte de la tesis.

3. El paso desde el LN a un instrumento formal. Ej. “infinito” (Stahl, 1970). Este tipo no necesariamente se etiqueta como “naturalización”, sino más precisamente define lo que sería “formalización” de un concepto. No obstante, siguiendo la definición amplia de “paso desde lo no científico a lo científico”, entonces este tercer tipo sí es una forma de “naturalización”, ya que es el paso desde un término de LN a disciplinas como la matemática, la cual se considera parte también de las ciencias *grosso modo*.

Hay un segundo proceso que sigue a la naturalización, definida como está hasta aquí, que se da en las teorías muy exitosas, a saber, la *popularización* de la noción naturalizada al punto de permear hasta el LN y el sentido común. Luego, por ejemplo, un concepto como “gravedad” pasa a ser la causa indiscutible de todos los casos concretos en la vida real donde vemos un objeto caer, y si uno llegase a afirmar que “la gravedad no existe”, entonces tal afirmación se entendería como negando los casos concretos de que las cosas caen, o aseverando que algo inerte que deje caer pueda salir volando, cuando tal afirmación debe ser entendida como negando cierto elemento de un CO de un IC, pero no aquellas características (el que las cosas caigan) del F del cual se abstraigo en un primer momento el SO.

Pues bien, la naturalización es el rasgo distintivo del completamiento de la fase de asentamiento de un IC llegando a la madurez, pero al ser a la vez entendida “naturalización” como el “paso de una noción desde lo no científico a su conceptualización dentro de un IC asentado en una disciplina científica”, entonces ¿cómo superar esta circularidad evidente? En realidad, no hay nada que superar, ya que la circularidad, sería problemática si no se tuviera en cuenta que al completarse cada círculo entonces lo que es “lo científico” ha variado sensiblemente a lo largo del devenir histórico científico. Es decir, la posibilidad de naturalización de elementos de CE y CO de cierto IC, depende de lo que ya los IC previamente maduros o bajo una etapa intermedia de asentamiento-madurez han naturalizado. Una vez que el competidor ha logrado naturalizar su CE y CO – o al menos parte de éstos – a la vez de haber transformado el panorama de los conjuntos componentes posibles (mediante eliminación, subsunción y agregación de lo propio), entonces se vuelve parte del tamiz de aquello que nuevamente trata de ser naturalizado.

Si un IC no logra atravesar la etapa intermedia asentamiento-madurez es, precisamente, porque la competencia frente a otros IC no permite una naturalización indudable del contenido específico de tal IC. Esta dificultad de la naturalización se da, no meramente por una neofobia de los IC existentes, sino que es una consecuencia natural del principio (.F) o subdeterminación de las teorías científicas, ya que ¿por qué naturalizar el concepto de IC1 y no el de IC2, si ambos están subdeterminados por los mismos datos empíricos? Es por ello que lo que he descrito, que en una primera mirada puede sonar una tesis fuertemente guiada por lo sociológico, tiene en cuenta en su trasfondo el valor empírico mínimo obvio que debe tener un IC novedoso.

Def.6.5. Existe la posibilidad de que en cualquier fase un IC sea disgregado, entendiéndose por “disgregación” el que algún sector de su estructura o su estructura completa sean subsumidos o reducidos por otros IC que trataban un F o SON cercano al IC disgregado.

Def.6.6. Existe la posibilidad de que en cualquier fase un IC sea eliminado, entendiéndose por “eliminación” el que algún sector de su estructura o su estructura completa desaparezca del grupo válido de los F que trata una disciplina.

Es importante distinguir entre lo que sería el eliminacionismo como postura y afirmar la eliminación como un hecho; el eliminacionismo es una postura donde la competencia entre dos o más IC es llevado al punto límite de afirmar, además del obvio monismo explicativo y privilegio ontológico, que el o los IC contrincantes deben desaparecer para el progreso sobre la investigación de cierto F. La afirmación de la eliminación como un hecho, más que una postura, en realidad es una sencilla constatación, a saber, que existen conjuntos componentes de ciertos IC que, con la suficiente cantidad de tiempo, terminan por diluirse en el desarrollo de las investigaciones. Luego, *es posible afirmar la existencia de la eliminación, sin por ello tener ningún ápice de eliminacionismo tal afirmación*. Esta distinción sutil, quizás dentro de los eliminativistas más clásicos (Churchland, 1984; Churchland, 1989) no ha sido suficientemente bien destacada, al punto que si alguien afirma la eliminación como un hecho dentro de las ciencias, se tiende a asociar a tal persona de inmediato una postura eliminativa. Es más, el eliminacionismo puede ser entendido como una postura en filosofía de la mente, en filosofía de las ciencias, en ambas o, como haré en mi propio contexto, como la afirmación de un hecho independiente de posturas filosóficas.

Por otra parte, por “disgregación” estoy apuntando a un fenómeno de dos etapas: una primera donde son extraídos los elementos de los conjuntos componentes de cierto IC, para en una segunda etapa integrarse tales elementos en un IC que sea en parte híbrido del primero. Luego la etapa de que he denominado “disgregación” debe ser más exactamente entendida como “disgregación + subsunción” o “disgregación + reducción”. *Luego, se puede, al igual que con la eliminación, ser reduccionista y afirmar el hecho de disgregación más reducción, o simplemente afirmar que tal cosa existe como parte del desarrollo normal de los IC sin comprometerse con una postura reduccionista*.

7. Definiciones ligadas a las categorías de IC

7.1. *Dada Def.0.5., Def.0.6. y 6.1. un IC puede tener diferentes estatus convencionales para su análisis.*

Def.7.2. *Un IC es concreto si cumple totalmente con poseer Estructura tipo IC.*

Es imposible que un IC sea construido en el aire, dadas sus restricciones por definición, por ejemplo de SO y SON, donde la ligazón mínima que se impone entre IC y F viene dado por ca y cs, los primeros variando en grados de abstracción. Si es el caso que un IC tiene una estructura completa tipo IC, entonces es concreto en un sentido que permite rastrear, inclusive, cada detalle del qué, el cómo y el porqué del mismo IC. Dado que la tesis de los IC es un instrumento de intenciones fundamentalmente descriptivas que depende fuertemente de cierta dosis de interpretación, entonces transformar un IC “real”, en una estructura tipo IC, de modo de darle la categoría de IC concreto, depende de la pericia del investigador en lograr capturar el mínimo de detalles necesarios para dar contenido a cada uno de los conjuntos componentes de la estructura IC.

Def.7.3. *Un IC es abstracto si cumple parcialmente con poseer Estructura tipo IC.*

Un IC abstracto también puede ser denominado un “pseudo-IC”, dado que no cumple las condiciones explicitadas en Def.0.2. y Def.0.3. Sin embargo, aun así es bastante útil de otras maneras que tocaré con cierto detalle en la sección (2.5.iv.)

7.3.1. *Un IC es abstracto de dos modos: abstracto-generalizado y abstracto-incompleto.*

Un IC abstracto es generalizado cuando su aparición se debe a una abstracción de rasgos comunes de diferentes IC concretos y este modo de IC abstracto es característico de la etapa de asentamiento, la cual es posterior a la etapa de nacimiento donde, típicamente, se puede identificar con un IC concreto. Un IC es abstracto-incompleto cuando es construido sobre un F, pero no llega a ser concretado, es decir, no llega a cumplir una estructura tipo IC. El análisis basado en IC abstracto-generalizado es de utilidad dentro de la fase de asentamiento de los IC (Def.6.3.) y el IC abstracto-incompleto puede llegar a tener su utilidad dentro de la fase de nacimiento (Def.6.1. y Def.6.2.). Por igual, ambas clases de IC pueden ser pedagógicas de algún modo, por ejemplo, para investigadores interesados en los IC concretos que son generalizados, o bien por el IC concreto que puede llegar a surgir en caso de que el IC abstracto-incompleto sea completado.

Específicamente, la clase de IC abstracto-generalizado es, en principio, *suficiente para subsumir cierto subcampo dentro de una disciplina*, pero, dado el estado altamente especializado de las disciplinas hoy en día, sería raro encontrar un IC abstracto-generalizado suficientemente significativo que sirva para capturar una disciplina completa; ello queda a la evidencia que se logre reunir. Sin embargo, no está implícita en la misma definición de IC abstracto-generalizado el que *a priori* sea capaz de dar cuenta de todas las subdisciplinas o una disciplina. Conceptualmente aquello puede ser estipulado, no obstante, dejo la noción de “subdisciplina” y “disciplina” a la intuición de cada lector, aunque soy de la opinión que lo que estoy desarrollando aquí puede servir como una base que permite definir el *fenómeno emergente de las disciplinas científicas*.

Def.7.4. *Un IC es cerrado o puntual si es tomado como estrictamente ligado a un instante específico de su desarrollo histórico.*

Def.7.5. *Un IC es abierto o extendido si es tomado su desarrollo en un intervalo de tiempo.*

La diferencia entre ambos tipos se debe sencillamente a la variable histórico-temporal, sin embargo no he desarrollado ninguna heurística (como en el caso de los TF) que me permita claramente dirimir entre ambos, a pesar de que son intuitivamente diferenciables dado el contexto de estudio de un IC. Por ejemplo, cuando se hace historia de las ciencias, lo que se hace es tomar un IC concreto o abstracto combinadamente con la categoría de ser extendido (Def.7.5.). En caso de hacer ciencias, entonces lo que se hace es tomar un IC concreto puntual (Def.7.4.) y “extenderlo”, no temporalmente, sino que transformando sus conjuntos componentes, o, como se ejemplificó con los modelos híbridos en ciencias cognitivas, se toman dos o más IC concretos y puntuales, de modo de combinarlos para incrementar la explicación-predicción de características de cierto F.

7.6. *En consecuencia, existen cuatro posibles tipos de IC abstractos: abstracto-generalizado cerrado, abstracto-generalizado extendido, abstracto-incompleto cerrado, abstracto-incompleto-abierto.*

7.7. *Y existen dos posibles tipos de IC concretos: concreto cerrado y concreto abierto.*

Cada una de las posibilidades enumeradas en (7.6.) y (7.7.) tienen distinta labor en el campo amplio de las filosofías de las ciencias y la historia de las ciencias. Lo que he hecho en 7. es dar un nombre descriptivamente guiado de los objetos cuyas categorías convencionales sirven para ambos campos.

8. Definiciones ligadas a la competencia/colaboración y conmensurabilidad

Def.8.1. Dado traslape entre F y SON de dos (o más) IC, éstos pueden relacionarse compitiendo o colaborando entre sí.

Esta definición, en conjunción con la Def.5.1.1., significa que debe existir un mínimo traslape entre ca y cs de dos IC para que estos entren en una relación que va, desde la competencia total, donde los defensores de un IC abogan por la eliminación del otro IC, a otra de colaboración, usualmente motivada, o bien por el pluralismo, por una actitud pluralista o bien por algún reduccionismo suave. No es parte de este trabajo individuar todas las potenciales actitudes que puede tener un investigador respecto a las relaciones interteóricas que establece su propio IC respecto a sus contrapartes, empero, es posible extenderse en este punto utilizando la propia tesis de los IC. Véase, por ejemplo, las Def.6.5. y Def.6.6.

8.1.1. Competencia y colaboración pueden darse entre IC en cualquiera de sus fases.

Es parte del esquema (**FIGURA 14**) que en cualquiera de las fases hay posibilidad de eliminación y disgregación. Tales dos procesos *siempre* van ligados a los más amplios de competencia y colaboración, al punto que se inter-definen. Suponer, por ejemplo, que existe algo así como una competencia sin perspectivas de alguna variante de eliminación y/o disgregación es imposible, ya que la competencia requiere al menos una contraparte, y al existir un IC de contraparte, entonces la propia dinámica de los IC los obligará, eventualmente, a transformarse.

Def.8.2. La colaboración se define como la complementación de las estructuras de los IC de modo de lidiar con problemas de conexión entre sus respectivos SO y SON.

8.2.1. En caso de colaboración dos IC pueden llegar a ser analizados como uno solo.

El mejor caso de colaboración que sirve para ilustrar la Def.8.1. y su consecuencia 8.1.1., y que ha sido presentado en dos secciones separadas de la presente tesis (secciones (1.2.2.) y (2.2.)), es la arquitectura SAL, la cual es desarrollada para complementar los IC concretos que pueden extraerse de ACT-R y Leabra. SAL puede analizarse como siendo un IC concreto, pero también se presta su caso para analizar cómo los IC originarios se despliegan dentro de SAL y cómo se agregan, por ejemplo, nuevas ca o cs, que no habrían surgido dentro de un desarrollo aislado de ACT-R y Leabra.

Def.8.3. La competencia depende de la noción de conmensurabilidad, ya que es esta la que determina la posibilidad de competencia.

Esta definición es, sencillamente, la explicitación de un lugar común en filosofía de las ciencias con el cual concuerdo, excepto por el hecho que tal lugar común me sirve exclusivamente como un punto de partida para una definición propia de conmensurabilidad. La existencia de competencia es un hecho normal en el desarrollo de los IC desde la modernidad, no así la conmensurabilidad, la cual fue

“descubierta” con el historicismo kuhniano, permitiendo una visualización de las empresas científicas mejorada, la cual, inclusive para los propios científicos, era novedosa. Aquí sería el lugar donde desarrollar las anteriores nociones de conmensurabilidad en filosofía, para después hacer una comparación con la propia, sin embargo, dado que mi trabajo tiende hacia la filosofía de las ciencias cognitivas, desarrollar tal contrastación me desviaría demasiado del punto.

Def.8.4. La conmensurabilidad entre dos IC se da con dos condiciones necesarias: ambos deben ser concretos puntuales y ambos deben compartir SON.

La conmensurabilidad se da, si y solo si, se ignoran las actualizaciones naturales por el progreso científico de elementos de los conjuntos componentes que se dan en el tiempo o, lo que es equivalente en la práctica del análisis de IC, se los supone como dando lugar a un IC concreto y cerrado que convencionalmente es susceptible de ser aislado. La noción de conmensurabilidad surge en el contexto del análisis de las relaciones interteóricas en las ciencias empíricas, por lo tanto un IC abstracto, por definición parcialmente completo en sus rasgos necesarios y suficientes para transformarlo en uno concreto, simplemente se mantiene en el terreno especulativo y no cabe como un IC, es decir, como parte de los instrumentos empíricos. De esto se sigue que propiamente se puede hablar de conmensurabilidad bajo el requisito indispensable de hablar de IC concretos puntuales, entre los cuales deben existir, *al menos de forma preliminar, intersección entre los SON de cada uno.*

Def.8.5. La conmensurabilidad aumentará en grado entre dos IC, X e Y, dependiendo de si se cumplen de menos a más las siguientes condiciones (desde 8.5.5. a 8.5.1.):

8.5.1. Equivalencia entre $CO \in X$ y $CO \in Y$

8.5.2. Equivalencia entre $CE \in X$ y $CE \in Y$

8.5.3. Equivalencia entre $CC \in X$ y $CC \in Y$

8.5.4. Equivalencia entre $SO \in X$ y $SO \in Y$

8.5.5. Equivalencia entre F que da lugar al $SO \in X$ y F que da lugar al $SO \in Y$

En cada una de las condiciones entenderé por “equivalencia” como el mapeo entre conjunto que cumpla inyectividad sobre el $IC(t')$ y no necesariamente sobreyectividad. Es decir, para cada elemento en cierto conjunto componente de cierto $IC(t)$ debe existir, al menos, un elemento que le corresponda en $IC(t')$, pero pueden existir elementos en $IC(t')$ que no les corresponda un elemento en el dominio $IC(t)$. Esta definición permite visualizar en qué sentido expande $IC(t')$ a $IC(t)$. Es decir, veladamente dentro de la presente definición existe, no solamente una forma más exacta de definir conmensurabilidad, sino que hay, desde lo más puramente pluralista y libre de normatividad, *un criterio de decisión que permite dirimir entre dos IC conmensurables en competencia.* Esto será de especial trascendencia en la sección (2.4.2.), donde se mostrará cómo funciona este principio de equivalencia.

Sin embargo, dado que *siempre* habrá al menos un elemento en $IC(t)$ que no tenga una contrapartida en $IC(t')$, entonces, es permisible acotar los elementos de cada uno de los conjuntos componentes del $IC(t)$

de modo de establecer cierta conmensurabilidad. Luego, la pregunta obvia es ¿hasta qué punto se puede acotar IC(t) para que sea mapeable en IC(t')? Y en responder esta pregunta entra fuertemente el factor interpretativo, el cual, empero, permite por sí mismo ir abiertamente a la cuestión de cuánta conmensurabilidad es posible; si se debe acotar demasiado, entonces la conmensurabilidad será de un grado menor que si es posible acotarlo hasta cierto punto arbitrario, pero aun así encontrar posteriormente elementos en IC(t) que no era necesario acotar, dado que poseían imagen en IC(t').

Def.8.6. Dependiendo el grado de cumplimiento de las condiciones desde 8.5.1. hasta 8.5.5. se hablará desde una Conmensurabilidad Local (CL) Completa a una Conmensurabilidad Mínima o Virtual (CMV).

En caso de cumplirse desde (8.5.1) hasta (8.5.5.) existirá conmensurabilidad local completa, en caso contrario será incompleta, con grados que van desde lo cuasi-completo a lo virtual. Existe una restricción que impone la Def.8.4. atingente a la presente Def.8.6., a saber, que se debe cumplir traslape de SON, el cual se daría simplemente por el hecho de cumplirse la condición 8.5.4., dada la consecuencia 5.1.1. Luego, si una competencia puede cumplir con la Def.8.4., entonces ya será una CL cuasi-completa.

Sin embargo, la Def.8.4. es bastante restrictiva si uno se pone a pensar detenidamente sobre ello, ya que fácilmente puede no cumplirse tomados dos IC completos, donde sea el caso que cierto SO sea totalmente desigual al SO rival. Aun así, es posible tomar solamente la condición 8.5.5., dando lugar a una CMV, la cual desde Def.8.4. no sería ya parte del grupo de CL, pero que parece ser interesante, ya que implica solo equivalencia entre F que da lugar al $SO \in X$ y F que da lugar al $SO \in Y$. En vista de esta situación es que he apuntado en el desarrollo de la misma Def.8.4. que la “intersección” de dos SON debe ser meramente preliminar. Esto es lo que permite suponer que en un comienzo dos IC son susceptibles de ser conmensurados, aunque, finalmente, se llegue a una conclusión medianamente negativa con el caso de CMV. Esta situación es necesaria tenerla en mente cuando sea se trate con la relación de IC concretos locales del simbolismo versus los del conexionismo (sección (2.5.iv)).

Def.8.7. Existe una Conmensurabilidad Global (CG) completa definida como la máxima conmensurabilidad entre dos IC concretos y abiertos.

Esta definición, la única utilidad inmediata que presenta, es dar una noción más exacta a la conmensurabilidad máxima concebible entre dos IC concretos. Dado que, siguiendo las nociones presentadas en 0., la “naturaleza real” de los IC es el ser “abiertos” (quizás con un punto de inicio, pero no siempre con uno de término), entonces es necesario presentar aquel tipo de conmensurabilidad que les sea más preciso. Un IC concreto abierto es posible de estudio en aquellos casos donde es posible historiarlo desde su nacimiento hasta su eliminación – no su disgregación, ya que entonces quedaríamos atados en la investigación del IC que lo subsumió –, sin embargo tales casos son poco frecuentados por filósofos e historiadores de las ciencias que dedican más su energía y atención a las historias de éxito que a las de un fracaso de largo aliento como sería un IC concreto abierto. Por cuestiones netamente prácticas he definido las conmensurabilidades locales (Def.8.6.) y he dejado como una pregunta abierta cómo tratar con aquellas globales, las cuales, conjeturo, son intratables, a menos que se logre de algún

modo dar sentido más exacto a aquello que dejé como intuitivo, es decir, la distinción entre lo puntual o cerrado y lo abierto o extendido (véase el comentario en la Def.7.5.)

Def.8.8. Una conmensurabilidad es abstracta o pseudo-conmensurabilidad sysss se comparan dos IC abstractos. La conmensurabilidad abstracta depende de dos condiciones mínimas:

8.8.1. *Que se compartan los conjuntos componentes no existentes.*

8.8.2. *Que se compartan los F.*

Esta definición es indiferente de si se tratan abstractos-generalizados o abstractos-incompletos. Los IC abstractos son todos, por definición, cierta *forma de especulación* (véase Def.7.3. y 7.3.1.), por lo que una conmensurabilidad construida entre dos especulaciones, será claramente otra especulación. Por ello, considero adecuado referirme a ella como “conmensurabilidad abstracta” o, dado que también un IC abstracto es un pseudo-IC, llamarla también “pseudo-conmensurabilidad”. No obstante, aun cuando pareciera esta forma de conmensurabilidad no ser útil para dirimir de forma concisa entre dos IC, la conmensurabilidad abstracta jugará un rol importantísimo en el caso de la polémica simbolismo-conexionismo (sección (2.5.iv.).

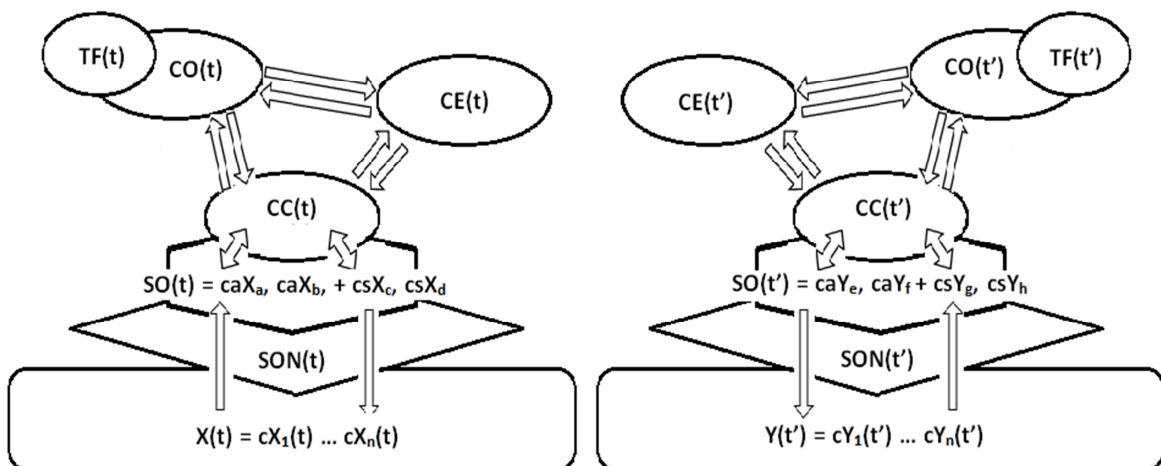
Def.8.9. Por “Inconmensurabilidad” se pueden tener diversas definiciones dentro del marco de los IC, las cuales pueden ser:

8.9.1. *No existe CL cuasi-completa.*

8.9.2. *No existe CMV.*

8.9.2. *No existe Conmensurabilidad Abstracta.*

FIGURA 15; la figura a continuación representa un caso de inconmensurabilidad de dos IC concretos puntuales, ya que cada uno de ellos cumple los requisitos para ser concreto puntual, poseyendo cada uno su propio SON, los cuales abstraen y suponen características de F diferentes y, además, en diferentes espacios temporales. (“F = X(t)” versus “F = Y(t’)”).



2.3.2.3. Revoluciones y la unidad de las ciencias

Antes de cerrar la exposición de la tesis de los IC en la presente subsección, la cual se abocará en tratar el tópico de las “revoluciones”, es preciso mostrar cómo hasta aquí se han cumplido buena parte de las preguntas que se hicieron al comenzar en la subsección (2.3.2.1.):

Primero, a la pregunta de cómo los IC delineados en (2.3.2.) con únicamente la Def.0.1. se “cargan” de contenidos, la respuesta será que el contenido pasa a través de la estructura que involucra CC, CE y CO, conectándose específicamente CC a un SO, con lo cual se cierra el círculo de SON. A través del sesgo de los investigadores en su enfoque se determina, no solamente SO, sino que también hasta qué punto se considera que el SON se logra completar o no. Luego, el fenómeno no es aquello que carga de contenido un IC, sino que está implícitamente desde los conjuntos componentes-base del IC, y un IC tendría cierta identidad convencional que lo diferencia de otros en cuanto que el conjunto de ca y cs difiere de otros. Así pues, F puede seguir siendo tan trascendente y tan discutible filosóficamente o por pura conceptualización como siempre, pero se concilia esta imagen con un SO que deja ya de ser asunto puramente conceptual, en tanto CC ya es una operacionalización sobre SO y no sobre F.

Segundo, a la cuestión de la demarcación, problema permanente en filosofía de las ciencias, la respuesta se reduce a las mismas definiciones de estructura tipo IC y a cómo esta estructura en conjunto con la propia dinámica de los IC tamiza el rango de posibilidades válidas que se terminan naturalizando definitivamente. Empero, *a priori*, no existe un conjunto predeterminado ni de redescpciones posibles, ni de CC posibles, ni de CE posibles, etc. Por eso la historia de las disciplinas tendrá un mismo rango de importancia que la práctica experimental de la investigación para comprender el porqué de la elección de ciertos IC por sobre otros; no existe una visión pura desde sólo el contexto de justificación. Por consiguiente, la cuestión de la demarcación al ser liberada de la búsqueda monista explicativa de sistemas normativos sería solucionable mediante la combinación de aquello que empíricamente se ajusta caso a caso y la misma historia de cómo se va cambiando en el tiempo tales cuestiones.

Y, tercero, la subdeterminación de las teorías científicas que aún no ha sido explícitamente usada en las definiciones sobre la estructura de los IC y su dinámica, a continuación será descriptivamente útil como hecho básico para construir la noción de “revolución”, la cual, al ser detallada como será a continuación, podría generar una salida divergente del problema de la unidad de las ciencias, la cual desde el pluralismo se postula como *a priori* irresoluble. Considérese la siguiente definición:

Def.9.1. Una revolución científica se define a partir de las tres condiciones siguientes:

9.1.1. Que un IC esté en su madurez mientras que el competidor esté entre nacimiento y asentamiento.

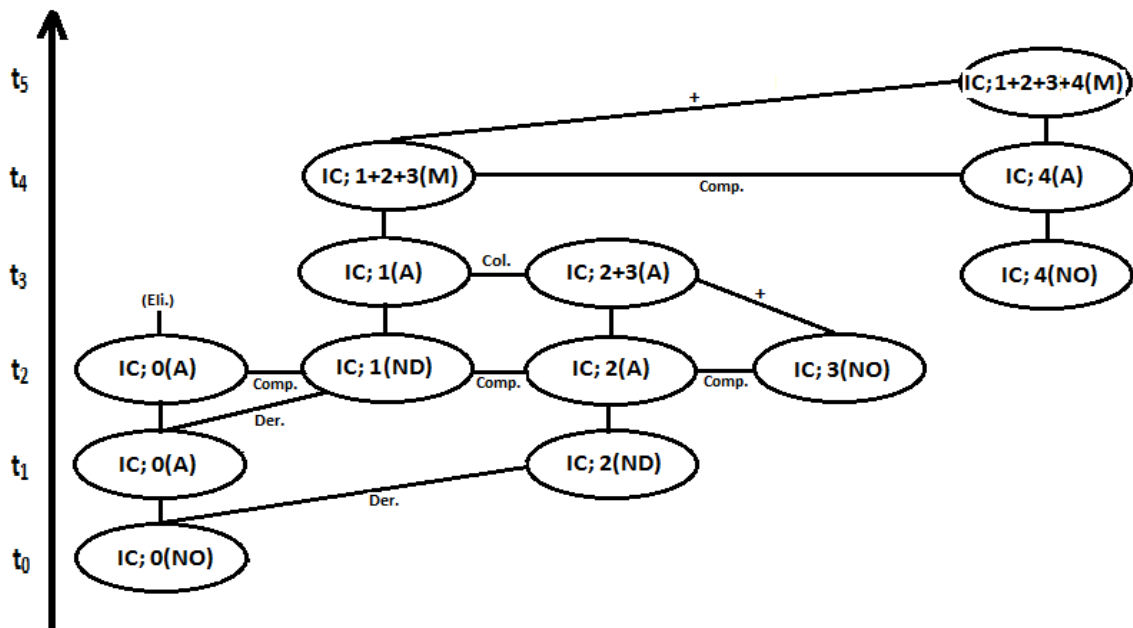
9.1.2. Que exista, mínimo, subdeterminación sobre el mismo F.

9.1.3. Que el IC competidor subsuma conjuntos componentes de la estructura del IC maduro.

Dadas estas tres características (medianamente restrictivas), entonces ciertos casos clásicos de revoluciones, como, por ejemplo, la revolución copernicana o la teoría de la relatividad, se mantienen

como ejemplo perfectos¹¹¹. Sin embargo, en particular en ciencias cognitivas nunca ha existido una revolución, con excepción de su propio nacimiento interdisciplinario¹¹², al amparo de la confluencia entre la teoría de la información, la naciente computación y la cibernética – esta última que vendría a ayudar a desarrollar la moderna teoría de control (Goodwin et al., 2000, 9) –, confluencia que dio lugar a muchos IC, los cuales entraron casi simultánea y paralelamente en una fase de asentamiento notoriamente alargada, donde los nacimientos derivados posteriormente se transformaron en la norma.

Una revolución idealizada sería la que muestra el siguiente esquema (FIGURA 16) en fases de diferentes IC en relaciones de competencia (“comp.”), colaboración (“col.”) y disgregación-subsumción (“+”):



El esquema en la FIGURA 16 divide el tiempo en 5 intervalos arbitrarios desde t_0 a t_5 y se considera el desarrollo de cinco IC, numerados desde el 0 al 4, en tal intervalo de tiempo. En este esquema se grafica, no solamente el acaecimiento de una forma de revolución con el IC;4, apareciendo primero aislado de los otros en t_3 – sin entrar en relaciones ni de competencia ni de colaboración – para luego evolucionar rápidamente al punto de subsumir a los anteriores en t_5 , sino que se grafica también la eliminación del IC;0, el cual, a pesar de la *fertilidad* para dar a lugar a otros IC derivados (“der”), no logra sobrevivir a la fase de asentamiento de modo que sea considerado como parte de los IC subsumidos por el IC;4.

Cabe agregar, que la revolución aquí esquematizada es ideal al punto de que IC;4 tiene un nacimiento no derivado y sería suficientemente fuerte para quedar como el único IC válido sobre cierto F. La condición

¹¹¹ En este sentido tengo exactamente la misma opinión de Paul Thagard “Kuhn's general view of scientific revolutions and his accounts of particular scientific episodes must be questioned, but his basic claim that the development of scientific knowledge includes revolutionary changes can be sustained.” (Thagard, 1992, 4). Es decir, los ejemplos de Kuhn (1970) *deben ser* considerados como ejemplares casi indiscutibles de “revolución”, y la búsqueda de una especificación de qué se entiende por “revolución” debiese subsumir tales ejemplares, o debiese darnos muy buenas razones para considerar que algunos calzan, mientras que otros no.

¹¹² En las conferencias Macy *in nuce* estaban la mayoría de los “paradigmas” que hoy pueden ser distinguidos en ciencias cognitivas. Esto se señalará con algo más de detención en la sección (2.4.).

de ser una IC de nacimiento original podría ser agregada como característica 9.1.4. en la Def.9.1. de modo de hacer más restrictiva la noción de revolución, pero creo que ello traería consecuencias indeseables, como, por ejemplo, terminar dejando fuera la revolución copernicana por no ser suficientemente “original” al reflotar el heliocentrismo originalmente griego – a pesar de que este último estuviese históricamente muy distante en el tiempo. Otra condición que serviría, también para tornar más estricta la Def.9.1., sería mediante una condición tipo “aniquilación de los IC competidores”, como se puede visualizar en el esquema antes colocado.

Ahora bien, dado que el pluralismo explicativo es incompatible con la unidad de las ciencias basada en el monismo explicativo – *traducido ahora a la tesis de los IC, el “monismo explicativo”, es asumir que existe contenido necesario y trascendente para algún conjunto componente que sea parte del “real” IC abstracto-generalizado que capture el todo de las ciencias* – y es neutral respecto al monismo metafísico, entonces, especulo sobre la base que se ha aceptado el medio conceptual descriptivo expuesto hasta aquí, quizás la salida al problema de la “unidad de las ciencias” es volver a preguntarse sobre cuál sería la noción que tendría una versión pluralista explicativa de la “unidad de las ciencias”.

Creo que la manera de superar el historicismo kuhniano, el cual es la base del quiebre que desembocó en el abandono del proyecto unificador, es con más historicismo; si vamos a la actividad de los científicos en cada época vemos que sus hipótesis son construidas, no necesariamente como una continuación de las teorías pre-existentes (considérese el último esquema presentado y nótese la relación de derivación entre IC;0 y los IC;1 e IC;2 la cual es esencialmente diferente a la de disgregación-subsunción y a la de propia evolución de un IC que cambia de fases), sino que sobre la base de CO, CE y CC ya existentes en su época, es decir sobre cierta, por así decir, “continuidad a trozos”.

Así pues, los IC poseerían una contigüidad de los elementos constituyentes de cada uno los conjuntos componente, de manera que poseerían cierta “continuidad” debida a cómo son reemplazados paulatinamente unos elementos por otros. Metafóricamente, piénsese en la paradoja del barco de Teseo: claramente no es el mismo barco si nos alejamos bastante entre dos puntos del tiempo, pero de dos maneras puede existir como una unidad: primero, *en cierto instante específico en el tiempo* y, segundo, *como proceso ininterrumpido*, aunque cada una de las partes del barco sea discreta.

Luego, la unidad de las ciencias tiene una imagen, que aunque no sea la que soñaron los positivistas lógicos, es desde mi perspectiva, suficientemente satisfactoria. Las ciencias sería una especie de torre en permanente construcción y reconstrucción que va creciendo hacia arriba y hacia los costados, es decir, en tanto la ciencia es acumulativa avanzaría verticalmente, y en tanto revisionista de sus propios fundamentos generaría una reconstrucción horizontal. No sería ni una torre de arquitectura preestablecida a la cual se tiende, ni tampoco una multiplicidad de ladrillos sin organización alguna.

En síntesis, y para cerrar esta exposición, todas los IC describen y tienen por meta explicar-predicir mediante su propia estructura cierto F. En un primer momento un IC puede ser conjeturado a partir del sentido común (principio (.E) en (2.3.2.)), sin embargo es menester la existencia de un lenguaje técnico mínimo (un CE mínimo) que permita separar conceptualmente al IC de ser una mera descripción sofisticada utilizando el lenguaje natural.

Las entidades que componen CO de un IC, estarán definidos directa o indirectamente por el CE que se conecta con el CC, el cual permite que el SO – que es conjunto de características abstraídas y supuestas de F – llegue a conectarse con SON. Las entidades de CO y el lenguaje de CE deben lograr exitosamente un proceso de naturalización para que tal IC que las contiene sea considerado como llegando a la madurez. En caso contrario, la fase anterior de asentamiento se prolongará indefinidamente.

La competencia es el espacio dónde se dirime, no solamente cuál naturalización llega a consumarse concluyentemente, sino que también cuáles conjuntos o elementos de cierto IC terminan por ser desechados y cuáles se rescatan. La competencia depende de la noción de conmensurabilidad, pero tal conmensurabilidad, en tanto ha sido restringida a los IC concretos puntuales, implica que no existe conmensurabilidad alguna sin supuestos precisos con los cuales trabajar, más bien al contrario, se parte de que debe existir alguna equivalencia entre dos IC para que haya discusión.

Por consiguiente, no hay punto de vista neutral, sino que siempre relativo a tales equivalencias contingentes que dependen de cierta hermenéutica, pero ello no es caer en una subjetividad, sino que en una *objetividad pluralista explicativa*, en tanto se asume un mínimo donde los “supuestos de comparación” o equivalencias pueden (y deben) ser explicitados *públicamente* (cfr. Def.8.5.). Este mínimo es el que permite la objetividad pluralista, y no caer en la subjetividad, donde se comparan dos IC. Las equivalencias debiesen ser debatidas abiertamente y “lo intuitivo” las reconstruye. Aquí lo intuitivo queda por parte de quien establece la interpretación mediante estructuración de IC de dos constructos explicativo-predictivos en competencia y es posible discutir si tal intuición es correcta o incorrecta en la medida que siempre se deben mostrar entre cuáles elementos se produce las equivalencias, y cuáles elementos representan una “expansión” o “mejora” desde cierto IC a otro IC.

2.4. Aplicación de la tesis de los IC al análisis de modelos aislados en Ciencias Cognitivas

La pluralidad de las ciencias cognitivas es un hecho no lo suficientemente destacado desde la perspectiva de los filósofos de las ciencias cognitivas quienes, normalmente, tienden a la separación entre “paradigmas”, mientras olvidan que tal pluralidad puede ser enfocada sin prismas monistas explicativos – y, asociados a ellos, privilegios ontológicos – y puede ser interpretada de otro modo, la pluralista. Luego, aun cuando la pluralidad sea un hecho, no implica que sea obvio enfocarse en ella pluralistamente, aunque esa sea la estrategia que he tomado en esta tesis.

Quizás, antes de entrar en las ejemplificaciones de cómo se aplica en análisis basado en IC, deba detenerme un momento de mejor modo a tal tópico, el de la pluralidad desde *el punto de comienzo de las ciencias cognitivas*, extendiéndome un poco más en tal cuestión que solamente ha sido apuntada muy a la pasada con el modelo TOTE y sus equivalencias respecto a un modelo generalizado cibernético.

“El comienzo de las ciencias cognitivas” es un punto arbitrario dependiendo la meta historiador, por lo tanto no puede ser “demostrado” como el único concebible, más bien puede ser “mostrado” como relativamente mejor o peor dependiendo el propio objetivo de cada autor. Por ejemplo, Gardner (1985) tiene como punto de partida el simposio Hixon en 1948, donde presentaron sus trabajos autores como Karl Lashley o Warren McCulloch, ambos aportando claves para la superación del conductismo, siendo en las conferencias Macy (primera conferencia en 1946) el lugar previo donde confluían autores como Wiener y Von Neumann, cuyas ideas ya serían parte del trasfondo común de Hixon y el posterior

“nacimiento consensuado” de las ciencias cognitivas en las conferencias en el MIT sobre teoría de la información en septiembre de 1956 (Gardner, 1985, cap.6). Esta elección no es al azar, ya que en el contexto de estas conferencias se presentarían los trabajos clásicos de Allen Newell y Herbert Simon (1956), George Miller (1956) y Noam Chomsky (1956).

En otro caso, en Boden (2006), la búsqueda de un inicio consensuado de las ciencias cognitivas le toma un trecho muchísimo más largo y detallado que Gardner – lo cual se agradece bastante, ya que presenta una imagen bastante más realista del modo en el cual se desarrolló la investigación de lo mental –, retrotrayéndose desde la antigüedad, la filosofía y lógica moderna, hasta fines del siglo XIX, para luego pasar a la primera mitad del siglo XX, donde, al igual que Gardner, coincidirá que es en la década de los 50’ cuando se produce la sincronía de diversas cuestiones que consistirán en “el comienzo de las ciencias cognitivas” (Boden, 2006, cap. 6); Primero, el uso de la teoría de la información – inclusive antes de la desarrollada por Claude Shannon – por parte de George Miller en psicología, específicamente en el desarrollo de explicaciones sobre el lenguaje, a las cuales se agregarían nociones computacionales a éstas de origen puramente informacional con la irrupción de Noam Chomsky estudiante del propio Miller. Segundo, el desarrollo de la psicología “New Look”, de la mano del ya conocido Jerome Bruner, la cual enfrentaba al conductismo terminando por desembocar en una psicología que haría uso metafórico de nociones computacionales. Tercero, los trabajos en inteligencia artificial simbólica de Newell y Simon. Cuarto, una serie de reuniones que permitirían cierta “consciencia común” de los investigadores, tales como las conferencias Macy, el verano del 56’ en Dartmouth y el seminario en noviembre de 1958 en el Laboratorio Nacional de Física (NPL) de Londres – donde se presentarían modelos como el Pandemonium de Selfridge, el Perceptrón de Rosenblatt, las especulaciones de McCarthy sobre dotar a programas con sentido común (investigación que llevará al desarrollo del lenguaje LISP) o la crítica de Yehoshua Bar-Hillel a McCarthy que sería clave para el llamado “frame problem”. Quinto, y, finalmente, la aparición de un texto-declaración de intenciones que sería la obra *Plans and the Structure of Behavior* (Miller et al., 1960), que serviría como libro guía para expandir los horizontes y la idea central de lo que sería a partir de entonces la ciencia cognitiva.

En estos dos ejemplos pareciera existir cierto consenso sobre colocar “el punto de comienzo de las ciencias cognitivas” a partir de la década de los 50’. Sin embargo, creo que es posible mover el inicio unos cuantos años antes, a las míticas conferencias Macy¹¹³, las cuales a partir de 1946 estarían centradas en la cibernética, siendo esta última, al menos si se recurre a la obra clásica de Norbert Wiener (1948), la corriente meta-teórica que contiene el germen completo de las características de absolutamente todas las vertientes de las ciencias cognitivas que han transcurrido en el tiempo (simbolismo, conexionismo, enactivismo, dinamicismo). Por ejemplo, en su obra *Cybernetics*, Wiener especulando sobre la relación entre aquellas condiciones psicopatológicas sin un correlato material conocido en aquella época, por ejemplo, la esquizofrenia, considera que la cibernética es un avance para explicar tales males, ya que permite la diferenciación entre la base material y su aspecto funcional. En sus palabras (Wiener, 1961, 146-7):

¹¹³ Lo de “míticas” alude a que las primeras cinco conferencias Macy carecen de *proceedings* donde se consignen el contenido de las presentaciones y las discusiones que se suscitaban posteriormente (Dupuy, 2009, 73-4).

This distinction between functional and organic disorders receives a great deal of light from the consideration of the computing machine. As we have already seen, it is not the empty physical structure of the computing machine that corresponds to the brain (...) but the combination of this structure with the instructions given it at the beginning of a chain of operations and with all the additional information stored and gained from outside in the course of this chain. This information is stored in some physical form - in the form of memory - but part of it is in the form of circulating memories, with physical basis which vanishes when the machine is shut down or the brain dies, and part in the form of long-time memories, which are stored in a way at which we can only guess, but probably also in a form with a physical basis which vanishes at death. (...) There is therefore nothing surprising in considering the functional mental disorders as fundamentally diseases of memory, of the circulating information kept by the brain in the active state, and of the long-time permeability of synapses.

El detalle de las relaciones entre los diferentes “paradigmas” con la cibernética queda a una investigación ulterior¹¹⁴. No obstante, este párrafo inmediatamente trae a la mente que, primero, la suposición de la existencia de “instrucciones” resuena con la apelación de “reglas” en el simbolismo. Segundo, las “memorias en circulación” tienen una contrapartida obvia con la formulación de las redes neuronales recurrentes. Tercero, y no presente en el párrafo, pero sí en el mismo contenido de la obra, el copioso uso de ecuaciones diferenciales de inmediato recuerda al enfoque dinamicista. Y, cuarto, la misma noción de causalidad circular es fundacional de la cibernética y común al enactivismo.

Si se acepta esta premisa, a saber, que “el comienzo de las ciencias cognitivas” se da en el contexto de las conferencias Macy a partir de 1946, donde se delinearía parte de lo que posteriormente se divulgaría en *Cybernetics*, entonces, no solamente la defensa de los “enfoques alternativos” frente al cognitivismo sería innecesaria, sino que *la pluralidad de las ciencias cognitivas sería fácilmente constatable* en la variedad de tópicos presentes a lo largo de las conferencias Macy de las que se guardan registro (Von Foerster, 1955).

Pero uno puede ir más lejos y aseverar que, más allá del comienzo de las ciencias cognitivas, *la pluralidad de las ciencias cognitivas es un hecho constante dentro de su historia*. La actitud pluralista sin implicar pluralismo de ninguna clase es otro hecho poco estudiado, pero que considero intuitivamente válido afirmar como norma dentro de las ciencias cognitivas desde sus comienzos (esto será mejor desarrollado en (2.5.i.)). El pluralismo epistemológico en sus diferentes variantes ha sido abrazado con menor o mayor fuerza dependiendo la época, sin embargo, es altamente probable rastrearlo como postura más común en el periodo de auge de la *primera* cibernética¹¹⁵ (décadas del 40’ y 50’, quizás hasta principios de los 60’) y a partir de la segunda mitad de la década del 80’ hasta la actualidad con la polémica del simbolismo-conexionismo. El pluralismo explicativo es una versión epistemológica bastante más refinada, y considero que *Pluralism and the future of cognitive science* es un hito que marcaría una nueva actitud emergente sobre la pluralidad inherente de las ciencias cognitivas.

¹¹⁴ Aunque algo de ello será muy sucintamente bosquejado dentro del apartado (2.5.i.) que puede verse como una continuación de esta parte.

¹¹⁵ La existencia de diferentes “cibernéticas” es un hecho bien conocido para el especialista en tal metateoría (Novikov, 2016), sin embargo para el lego no lo es. Es por ello que especifico a que me refiero aquí, y en general siempre a lo largo de toda la tesis, a la primera cibernética.

Ahora bien, la tesis de los IC, antes de aplicarla en la sección final de la presente tesis directamente a la polémica entre simbolismo y conexionismo, considero que es adecuada utilizarla en el marco de un par de hitos clásicos dentro de ciencias cognitivas, quizás no con todo el rigor que necesita un constructo filosófico original ser ejemplificado, pero con un piso mínimo que permita comprender lo que he desarrollado en la sección (2.3.) y sus subsecciones. A continuación, secuencialmente, en las siguientes dos subsecciones desarrollaré:

- Primero, en (2.4.1) se presentará la neurona de McCulloch y Pitts de manera de mostrar cómo ésta puede ser interpretada como un IC concreto puntual. Especialmente, me interesa mostrar cómo un IC abstracto generalizado puede transformarse en un IC concreto puntual sin necesidad de la intervención de “lo empírico” entendido como un marco experimental desarrollado, sino que meramente mediante la saturación de los conjuntos componentes.
- Y, segundo, en (2.4.2.) se presentará brevemente la teoría y supuestos del perceptrón de Rosenblatt para pasar de inmediato a contrastarlo con la neurona de M&P. Esto tendrá la finalidad de ejemplificar cómo es la competencia directa entre dos IC, producida por la conmensurabilidad local al ser ambos IC concretos cerrados y constar de varios elementos equivalentes.

2.4.1. La neurona de McCulloch-Pitts

Ampliamente es considerada la neurona de Warren McCulloch y Walter Pitts (M&P), que es presentada en el artículo *A Logical Calculus of the Ideas Immanent in Nervous Activity* (1943), como el “origen” del conexionismo (Gardner, 1985, 319-320; Abdi, 1994, 247-8; Parks et al., 1998, 8-9), no obstante también como un punto de nacimiento común entre el simbolismo y el conexionismo (Piccinini, 2004, 204)¹¹⁶. Tratando de mantenerme ortodoxo respecto a lo que se dice dentro del contexto de la historia de ciencias cognitivas, supondré la neurona de M&P como representante de la prehistoria del conexionismo tipo PDP – y sólo de modo muy general como siendo predecesor del simbolismo –, pero de modo más amplio, puede ser considerado como dando lugar a buena parte de lo que será parte de la literatura de las redes neuronales artificiales (ANN) al indicar la posibilidad de ANN recurrentes (ejemplos paradigmáticos son la red de Hopfield y la máquina de Boltzmann).

La originalidad de la neurona de M&P radica en su *intención explícita de idealizar computacionalmente el umbral de activación de la neurona, la cual es considerada como una estructura que consta tanto de dendritas como soma y axón, pero que no reduce su análisis al comportamiento eléctrico de la membrana*. En términos más mundanos, la neurona de M&P no es una mera abstracción *eléctrica* de la neurona, sino que es una abstracción *electrónica* de ella. En general, la importancia del artículo de M&P radica en, al menos, tres motivos:

(1) Pone de manifiesto, inclusive antes que Turing (1950), Von Neumann (1958) y Wiener (1961 [1948]), que el nivel *relevante* de la cognición humana es, *grosso modo* dicho, el computacional, y muestra un modo concreto de cómo el cerebro sería una máquina computadora, siendo el pensamiento

¹¹⁶ Un tratamiento exacto del impacto del artículo de M&P daría lugar un estudio transdisciplinario que abarcaría, al menos, tres campos: teoría de autómatas, inteligencia artificial (específicamente ANN y *machine learning*) y ciencias cognitivas. Sin embargo, tal tratamiento más adecuado del impacto del artículo daría lugar a un trabajo de una envergadura que va más allá del foco del presente escrito muchísimo más acotado y “práctico”.

esencialmente el funcionamiento de tal circuito abstracto. Sin embargo, aunque M&P no harían la distinción hardware/software, hecho que permite ligar su propuesta teórica más cercanamente al conexionismo que al simbolismo, es menester notar que utilizan la noción de “implementabilidad” en un sentido bastante cercano al que lo hace el simbolismo, el cual utiliza explícitamente la metáfora del computador (McCulloch & Pitts, 1943, 119).

(2) Conceptualmente, contiene los dos componentes generales que conforman una neurona conexionista genérica, es decir, una neurona de M&P posee un “integrador” y una “no linealidad” (Anderson, 2007, 71-6), siendo tal “no linealidad” en esta neurona equivalente a una función de escalón (Scott, 2002, 236-7). Estos dos componentes generales son vitales para entender la continuidad que se establece entre el modelo de M&P, el perceptrón y las redes *feedforward* entrenadas mediante la regla delta generalizada del grupo PDP.

(3) Es el modelo más simple de la diversa gama de ANN, las cuales se diferencian primordialmente de la familia de redes neuronales biológicamente inspiradas (BINN) en que no toman en cuenta la estructura temporal de los trenes de potenciales (Maass, 1997, 1661; Gerstner, 1998, 3-4; Smith, 2006, 440-1). La neurona de M&P es el elemento básico de las redes de primera generación, las cuales, junto a las de segunda generación – como por ejemplo las redes con activación sigmoide y algoritmo de aprendizaje por retropropagación del grupo PDP –, presentan cierta continuidad que permite separarlas de la tercera generación de redes neuronales, estas últimas estando fundamentalmente conectadas a la visión de la neurona con su comportamiento eléctrico o, lo que es equivalente, a su modelamiento neuroelectrofisiológico (Lapicque, 1907; Hodgkin & Huxley, 1952; Knight, 1972).

Ahora bien, aunque el artículo de M&P es bastante sintético, por lo cual pareciera ser un punto de partida sencillo para ejemplificar el modo en que es susceptible cierto modelo de ser trasladado a una estructura tipo IC (Def.0.2., 0.2.1. y 0.2.2.), el formalismo del artículo original de M&P utilizado es el de autores como de Carnap, Russell, Hilbert, Whitehead, Ackermann, por lo tanto su lectura y exposición directa no resulta muy adecuada para entender cómo se llega desde tal formulación lógica a la neurona considerada en su interpretación más “estándar”, es decir como “(...) an elementary arithmetical element of computation which computes a Boolean threshold function.” (Van de Vijver, 1992, 107). Es por ello que trataré de dar una salida expositiva que tome algo, tanto de la formulación original, como de la posterior o más “estándar” para facilitar posteriormente su análisis.

El artículo de M&P se divide en cuatro partes: una introductoria donde se explicita cuál es el objetivo y el alcance de lo que se está presentando, importante en tanto justifica por qué los principios que son detallados en la segunda parte pueden ser justificados neurofisiológicamente. La segunda parte es el desarrollo de la teoría sobre las redes sin círculos, es decir aquellas que constan de algún bucle en su procesamiento. La tercera parte investiga algo sobre las redes con círculos, y es interesante notar que es dentro de este contexto que se hacen las comparaciones de poder computacional con las máquinas de Turing¹¹⁷. Y la cuarta parte son algunas consecuencias que permiten identificar las esperanzas puestas

¹¹⁷ De hecho es la parte que recibe más observaciones críticas, en particular Stephen Kleene, además de considerarla “obscura” (Kleene, 4, 1951), a lo largo de su reconstrucción nota que la fórmula a continuación de (9) en la página 126 del artículo de M&P no es correcta, ya que es posible un contraejemplo (op. cit., 35-6).

por M&P sobre el bosquejo teórico que presentan en su artículo. De estas cuatro partes, la tercera no será delineada, mientras que las otras tres sí.

La historia del artículo de M&P surge de una idea que habría estado siendo pensada por el propio McCulloch que dataría al menos desde el año 1929, cuando concibió que los que él denominaría “psicones”, o los elementos psíquicos más simples concebibles, – análogos al átomo para la química o el gen para la genética, pero sin una perduración como éstos – que podrían ser ideados como los impulsos todo-o-nada de las neuronas (Tara Abraham, 2002, 7). Paralelamente en esta época, a mediados de los 30’, Walter Pitts habría empezado a ir a clases en la universidad de Chicago, sin registrarse como estudiante, y habría estado estudiando lógica con Rudolf Carnap y biofísica con Nicolas Rashevsky, uniéndose al grupo de investigación de este último, comenzando a producir trabajos originales sin haber siquiera obtenido algún grado académico (Piccinini, 2004, 184). Es importante aquí señalar que la figura de Rashevsky es considerada pionera en la investigación estrictamente teórica de las redes neuronales (Mehrotra et al., 1997, 5-6; Cowan, 1998, 105), quien, trabajando sobre supuestos acerca de la posibilidad de factores excitadores e inhibidores, la difusión de tales sustancias y sus gradientes electroquímicos (Abraham, 2004, 345), presentó una teoría de las redes neuronales como sistemas dinámicos no lineales (Haykin, 2008, 673).

Por consiguiente, en el momento en que McCulloch y Pitts colaboraron para la redacción del artículo fue, muy probablemente, McCulloch quien traía la idea de considerar al cerebro como una máquina lógica y Pitts quien con sus conocimientos en matemática y lógica llevó a cabo la tarea de formalización (Piccinini, 2004, 185). Es por esto que, aun cuando el objeto de estudio del artículo de 1943 es el sistema nervioso y el funcionamiento de la neurona, M&P están fuertemente motivados por temáticas de filosofía, lógica y matemática.

La primera parte del artículo, que es la introductoria al tema, sitúa de inmediato al trabajo en lo que es la “neurofisiología teórica”, donde existen una serie de supuestos ya aceptados de cómo es en general el sistema nervioso, por ejemplo, que se compone de una red de neuronas estructuradas en soma y axones, que se conectan entre sí sinápticamente, que los impulsos sinápticos deben superar un umbral para transmitirse y que la velocidad de tal impulso está en proporción al diámetro del axón (McCulloch & Pitts, 1943, 115). Pues bien, el fenómeno neurofisiológico central para M&P será la ley “todo-o-nada” de la actividad nerviosa la cual “(...) is sufficient to insure that the activity of any neuron may be represented as a proposition. Physiological relations existing among nervous activities correspond, of course, to relations among the propositions; and the utility of the representation depends upon the identity of these relations with those of the logic of propositions.” (op. cit., 117).

Sin embargo, dos dificultades aparecen preliminarmente a esta consideración de la actividad neuronal como proposicional. La primera es sobre la facilitación y la extinción, fenómenos donde la actividad antecedente temporalmente altera la capacidad de respuesta a la estimulación subsecuente de una y la misma parte de una red. La segunda es sobre el aprendizaje, fenómeno donde la actividad concurrente en algún punto temporal pasado ha alterado la red permanentemente, luego incidiendo en que el estímulo antes inadecuado es ahora adecuado (o viceversa). Empero, ni McCulloch ni Pitts conciben la equivalencia formal como una explicación factual, al contrario, consideran facilitación y extinción

presumiblemente dependiente de cambios eléctricos y químicos en el umbral, a su vez que el aprendizaje el cambio duradero que resiste a sueño, anestesia, convulsiones y coma. Para M&P la importancia de la equivalencia formal que ellos proponen radica en que: “the alterations actually underlying facilitation, extinction and learning in no way affect the conclusions which follow from the formal treatment of the activity of nervous nets, and the relations of the corresponding propositions remain those of the logic of propositions.” (op. cit.) En otros términos, la idea de M&P es la creación de una teoría que establezca el conjunto de redes posibles que sean *implementadas* por las redes reales. O, también en otras palabras, la idea de la teoría encontrar un modo de establecer una línea base que estime lo que las redes neuronales pueden hacer (Scott, 2002, 235)

La segunda parte del artículo se aboca a los principios para la creación de tal conjunto de redes posibles y a estudiar una serie de hechos generales postulados mediante teoremas sobre el conjunto de redes sin bucles. Los principios son originales de M&P son (op. cit., 118):

1. La actividad de la neurona es un proceso de “todo o nada”.
2. Un número fijo de sinapsis debe ser excitada dentro del período de adición latente de modo de excitar una neurona en cierto momento dado, y este número es independiente de la actividad previa y la posición en la neurona.
3. El único retraso significativo dentro del sistema nervioso es el retraso sináptico.
4. La actividad de cualquier sinapsis inhibitoria previene absolutamente la excitación de una neurona en cierto momento.
5. La estructura de la red no cambia con el tiempo.

La idealización crucial es el principio 1 enunciado por M&P, que es la ley del “todo-o-nada”, la cual mantiene que la relación entre un estímulo neuronal y la actividad que produce se da o no se da. Toda neurona tiene un umbral finito y, una vez producido el impulso, la excitación procede independientemente de la intensidad del estímulo. Después de enunciar estos principios M&P pasan a la introducción de su notación y a la demostración de sus teoremas relativos a redes sin “círculos” o sin realimentación. Dado que su formalismo es infrecuente y difícil de interpretar, acudiré aquí a Piccinini (2004) quien hace una introducción idiosincrática del primer trecho de la segunda parte del artículo.

Las neuronas de una red N se denotan por c_1, c_2, \dots, c_n . Una expresión primitiva de la forma $N_i(t)$ significa que la neurona c_i dispara en el tiempo t . Expresiones de la forma $N_i(t)$ pueden ser combinadas por medio de conectivas lógicas para formar expresiones complejas que describen el comportamiento de diferentes neuronas en cierto instante. Por ejemplo, $N_1(t) \& N_2(t)$ significa que la neurona c_1 y c_2 disparan en un cierto instante t . Otra expresión como $N_1(t-1) \vee N_2(t-2)$ significa que, o bien c_1 dispara en el instante $t-1$, o bien c_2 dispara al instante $t-2$, o bien ambas cosas. Estas expresiones complejas pueden a su vez ser combinadas por las mismas conectivas lógicas. Como expresiones bien formadas M&P permiten solamente el uso de la conjunción ($A \& B$), disyunción ($A \vee B$), conjunción y negación ($A \& \neg B$) y una conectiva especial S que cambia el índice temporal de una expresión hacia atrás en el tiempo, de modo que $S[N_i(t)] = N_i(t-1)$. Una expresión compleja formada a partir de un número de expresiones primitivas $N_1(t), \dots, N_n(t)$ por medio de las conectivas antes dichas es denotada por *Expresión* $_j[N_1(t), \dots, N_n(t)]$. En

cualquier red sin círculos hay algunas neuronas que no reciben input desde otras neuronas, que serán llamadas *neuronas aferentes*.

Con estos elementos en mente, los problemas técnicos centrales por M&P formulados y por resolver serán “to calculate the behavior of any net, and to find a net which will behave in a specified way, when such a net exists.” (M&P, 119). Aquí Piccinini (2004, 184) considera que estos problemas son análogos a los formulados por Householder – uno de los miembros del grupo de investigación de Rashevsky – y Pitts dentro de sus teorías más tempranas. Householder se preguntaba cuáles proposiciones generales expresan la actividad de cualquier complejo de fibras nerviosas en términos de la dinámica de las fibras individuales y de sus constantes estructurales. Pitts habría adjuntado a éste el problema de la red inversa, es decir, dado un patrón de actividad en el tiempo, entonces construir cuándo una posible red neuronal tendría este patrón. Sin embargo, dentro de la teoría de M&P, el patrón de actividad de una red ya no sería descrito como una frecuencia de impulsos, sino más bien como un patrón preciso neuronal de impulsos descritos por una fórmula lógica. En términos de la teoría de M&P, estos dos problemas pueden ser formulados de la manera siguiente:

Primer problema: dada una red, encontrar la clase de expresiones C tal que para cada neurona c_i , en C hay al menos una expresión verdadera de la forma

$N_i(t)$ si y solo si $Expresión_j[(N_{i-g}(t-1), \dots, N_{i-2}(t-1), N_{i-1}(t-1))]$

Donde las neuronas c_{i-g}, \dots, c_{i-2} y c_{i-1} tienen axones entrando a c_i

El valor de esta expresión es que describe el comportamiento de cualquier neurona no aferente en términos del comportamiento de las neuronas que aferentes a ella. Si una clase C de tales expresiones es encontrada, entonces la lógica proposicional puede describir el comportamiento de cualquier neurona no aferente de la red en términos del comportamiento de las neuronas aferentes a ella.

Segundo problema: Dada una expresión de la forma

$N_i(t)$ si y solo si $Expresión_j[(N_{i-g}(t-1), \dots, N_{i-2}(t-1), N_{i-1}(t-1))]$

Encontrar una red para la cual ésta es verdadera

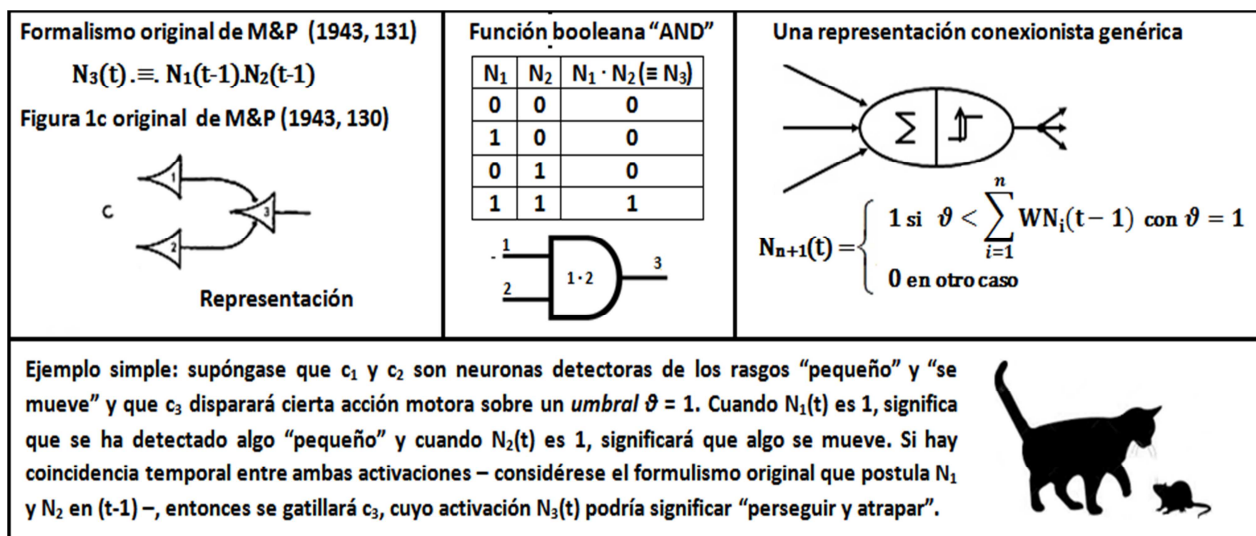
M&P exponen que en el caso de las redes sin círculos, estos problemas son fácilmente solucionables. Para resolver el primer problema, muestran cómo escribir una expresión describiendo la relación entre el disparo de cualquier neurona dentro de una red y los input que recibe desde sus neuronas aferentes. Para resolver el segundo problema, muestran cómo construir redes que satisfacen sus cuatro esquemas combinatorios (conjunción, disyunción, conjunción con negación, y predecesor temporal), dando diagramas que muestran las conexiones entre neuronas que satisfacen cada esquema. Luego, por inducción sobre el tamaño de las redes, todas las expresiones formadas por estos esquemas combinatorios son satisfechos por redes de neuronas M&P.

Vale la pena notar que las figuras de las neuronas van adjuntas al final del artículo de 1943 son acompañadas de una descripción en el formalismo lógico desarrollado por M&P y que sigue la sintaxis de Carnap. Como apunta Abraham (2003), las neuronas en sí mismas son representadas de una manera altamente abstracta, careciendo de detalles histológicos o anatómicos. Cada neurona es

estructuralmente idéntica. Sin embargo, la forma triangular no es totalmente arbitraria, claramente las ya famosas investigaciones de Ramón y Cajal que arrojaba luces sobre la histología de las neuronas piramidales influenciaron estos esquemas (Abraham, 2003, 417-8).

Dejando hasta aquí una primera exposición suficientemente apegada a las bases propuestas por el propio dúo de M&P, es preciso presentar cómo es que es posible decir que las neuronas de M&P son susceptibles de ser interpretada como *implementaciones de funciones booleanas*. La forma de llegar a esto no es complicada, basta considerar que, según M&P, una neurona encarnaría una proposición y que su actividad representa el resultado de aplicar la conectiva a las proposiciones que están presentes en las neuronas aferentes, es decir, una neurona al estar dejando o no pasar el impulso de sus representaría la verdad o falsedad de cierta proposición, esta última compuesta de las actividades de las neuronas aferentes. Sin embargo, esto aún no es exacto, ya que las redes no operan sobre fórmulas, sino que meramente sobre “apagados” o 0 y “encendidos” o 1, luego la actividad de las neuronas no encarnarían conectivas, sino que funciones booleanas, operando no sobre fórmulas, sino que, en el mejor de los casos, sobre valores veritativos de fórmulas; así pues se ha “(...) finally reached the proper description of McCulloch-Pitts nets as logic machines: they embody, or realize, or instantiate Boolean functions.” (Andler, 1992, 131)

A partir de esta lectura, a saber, aquella que denominaré *estándar* de la red de M&P, puede aducirse que los autores obtendrán dos resultados: primero, M&P muestran que su modelo neuronal podría representar los tres elementos fundamentales de los circuitos de la ingeniería computacional o de la electrónica: las compuertas AND, OR, NOT. Y, segundo, apelando al álgebra de clases, demuestran que cualquier función booleana puede ser modelada por una o más de sus redes, y que tales redes pueden corresponder a una o más funciones booleanas (Scott, 2002, 235-6). Lo único que diferenciaría a las neuronas de M&P de un formalismo estrechamente reducido a las compuertas lógicas sería que cada neurona poseería un umbral (usualmente denotado por ϑ), que determinaría posteriormente a la integración de las activaciones si cierta neurona no aferente se activa o no (véase la formulación de la neurona M&P en Anderson, 2007, 66-7) Una síntesis de expuesto hasta acá a continuación (**FIGURA 17**):



El lector avisado notará en el cuadro anterior (**FIGURA 17**) una letra mayúscula “W” en la sumatoria presentada en el extremo derecho que pareciera no hacer referencia a nada explicitado hasta ahora. En realidad sí hace referencia a algo que ha sido apuntado en la primera parte de la tesis, en el contexto de los elementos que componen el marco general conexionista, a saber, a los “pesos” (*weights*) de las conexiones. Sin embargo, este concepto aparecerá recién con el perceptrón y en este caso puede tratarse como una constante 1 o simplemente omitirse. Preferí no hacer la omisión, ya que de ese modo es más fácil visualizar cómo se produce la continuidad y “superación” de la teoría de M&P por parte del perceptrón de Rosenblatt (sección (2.4.2.) a continuación).

Pues bien, como se ha apuntado antes, la tercera parte del artículo de M&P no se revisará. Finalmente, y muy sintéticamente, la cuarta parte, además de traer anexada una serie de diagramas como el 1c (véase la **FIGURA 17**), contiene las consecuencias del artículo que son de un espectro amplio metafísico y epistemológico. Particularmente, una consecuencia epistemológica es que, dado que para M&P es aparentemente obvio que cualquier idea y cualquier sensación es implementada por la actividad dentro de una red compuesta por los elementos por ellos especificados, y además que dado que estas redes son estáticas según sus propios principios especificados al comienzo de la parte segunda del artículo, entonces es la estructura fija de una red la cual determina lo que un sujeto puede extraer del mundo exterior mediante la actividad de la red actual (McCulloch & Pitts, 1943, 129-131).

Por otra parte, M&P señalan algunos puntos psicológicos: primero, que todo lo que podrá ser alcanzado en psicología se restringe a la especificación de las redes. Segundo, que aquello, aunque pueda parecer una postura reduccionista, no lo es, ya que la actividad de tales redes es inherentemente proposicional y que todos los eventos psíquicos poseen un carácter intencional o "semiótico" (op. cit., 131). Por lo demás, la teoría de M&P arrojaría luces para la neurología, ya que la propuesta acentúa la distinción entre las redes necesarias o meramente suficientes para ciertas actividades, así clarificando las relaciones de estructuras perturbadas a funciones trastornadas (op.cit.).

Ahora bien, con lo expuesto hasta aquí, ¿es suficiente para decir que la neurona y las redes de M&P son un IC? Dejando de lado un uso de “IC” como reemplazo de “teoría” o “modelo”, sino que como un modo de interpretar cierto sector restringido de una, entonces lo primero, y más básico a llevar a cabo dentro de la tesis de los IC, es encontrar si resulta posible trasladar un cierto grupo de nociones teóricas a una estructura tipo IC (Def.0.2. y 0.2.2.). Esta estructura consta de los conjuntos componentes CO, CE, CC, SO y SON (Def.0.3.), por lo tanto es menester encontrar elementos para cada uno en lo descrito hasta aquí y, tan importante como ello, preguntarse cuál es el F que sirve como dominio de características abstraídas (ca) y características supuestas (cs) que compondrán el SO (Def.1.1. y Def.1.2.).

Lo más sencillo – pero lamentablemente lo menos exacto y sujeto a controversia – es especificar F, el cual en este caso serán dos elementos: primero, es *el sistema nervioso y su funcionamiento*, los cuales son caracterizados a partir de algunos elementos de la neurofisiología de la época. Y, segundo, es *toda la gama de fenómenos cognitivos*. Empero, difícilmente esta interpretación de F está lejos de la polémica; por una parte hay opiniones como la de Abraham, quien dirá que “McCulloch and Pitts saw their theory simply as a tool, one that would enable investigators to develop a formal, symbolic account of the activity of known arrangements of neurons” (Abraham, 2003, 220), la cual contrasta con la de Piccinini

quien dirá que “the original purpose of McCulloch and Pitts’s technique for designing nets was to explain mental phenomena” (Piccinini, 2004, 196), es decir, por una parte se puede visualizar la teoría de M&P como una simplificación exclusiva del sistema nervioso, mientras que por otra parte se le puede ver como haciéndose cargo de unir dos fenómenos separados en una unidad teórica. Creo que a partir de lo que se desprende de la última parte del artículo, entonces adheriré a la interpretación de Piccinini de lo que constituye el F base tomado por M&P.

Sin embargo, especificado F ¿cómo se da lugar al SO? El SO son ca y cs, siendo una ca clave la *presencia del umbral que permite la trasmisión del potencial*. Este umbral es el que dará lugar a su ley del “todo-o-nada” y se conjugará con una cs tal como que los *estados de activación de una neurona son equivalentes a un valor veritativo*. En general, puede decirse que los principio 1 al 5 funcionan como cs, mientras que el umbral y la estructura muy simplificada de las neuronas funcionan como ca.

El caso del CE es el interesante, ya que a lo largo de esta exposición de la teoría de M&P, la neurona y su red más trivial (específicamente, una AND) se ha presentado mediante tres formulismos diferentes: la primera, la original, la cual es poco intuitiva y de difícil introducción. La segunda, la “estándar”, bastante más fácil de introducir, ya que se basa en funciones booleanas, aunque pierde algún detalle al no contar las funciones booleanas con la noción de “umbral”. La tercera, y que está especificada en el cuadro resumen anterior, ha sido generalizada a partir de diferentes presentaciones de “neuronas conexionistas genéricas”. La pregunta más obvia que surge es ¿cuál de estas formulaciones es la real? La original es la de M&P pareciera ser sin duda. No obstante, ello no significa que sea la “real”, ya que cada una, a pesar de ciertamente cargar con otros CO detrás que no son parte de la intención original de M&P, permiten *completar la estructura del IC*, generándose como único “problema” el tener que dar matices respecto a cómo denominar tal IC, el cual, por ejemplo, podría ser algo así como “neurona MP conexionista genérica” si es que se elige la tercera formulación. Lo interesante aquí es que la existencia de variadas formulaciones permite inferir algo que va más allá del objetivo de completar el CE, a saber, *la existencia de IC derivados de la neurona y red de M&P*.

El CO depende de todos los conjuntos componentes más “abajo” de éste, es decir, por ejemplo, depende del SO y del CE. Dado que, por definición, existe al haber cierto CE – o al menos es posible reconstruirlos parcialmente como se ha hecho hasta aquí con otros conjuntos componentes –, entonces se reconstruirá a partir de los compromisos ontológicos que permitan constituir un CO necesario para el funcionamiento del modelo. Sin embargo, inclusive en un artículo tan acotado como éste es posible encontrar ideas que no calzan en ningún otro conjunto componente excepto en aquel que he denominado TF y que es el subconjunto inútil del CO. Por ejemplo, M&P dicen: “With determination of the net, the unknowable object of knowledge, the “thing in itself,” ceases to be unknowable” (McCulloch & Pitts, 1943, 131). Siguiendo a Piccinini (2004, 201), pareciera ser que M&P sugieren que si un sujeto puede conocer la estructura y la actividad pasada de su propia red, entonces éste podría conocer las “cosas en sí”. Este tópico kantiano tácito, que no es extraño encontrar inserto en el desarrollo del artículo dada la fuerte formación filosófica de ambos autores, *no agrega nada* a la explicación del fenómeno mental mediante neuronas de M&P, por tanto transformándose en el TF que no forma parte de la estructura del IC.

¿Y qué sucede con el CC y (más importante aún) con el SON que completan el IC? *Hasta aquí simplemente hay un IC abstracto incompleto* – el cual, conjeturo, es común dentro de la fase de nacimiento de los IC –, ya que no se ha individualizado una función más específica, en general, se ha hablado de un SO a grandes rasgos, pero pareciera obvio que no se está aplicando a nada. El modo de transformarlo en concreto, y más exactamente en puntual (puesto que se está trabajando sobre una parte estrechísima de tiempo) es buscando un SON. Luego, habría que volver a remitirse al artículo para encontrar algún ejemplo que funcione como tal, como podría ser el caso de *la ilusión del calor*, presentado por los propios M&P quienes construyen una red apropiada. La ilusión es bastante intuitiva: un objeto frío tocando la piel normalmente causa una sensación de frío, pero si es mantenido por un corto tiempo y luego quitado, entonces causará una sensación de calor. Al diseñar su red, M&P razonan del modo siguiente. Ellos parten del conocido hecho fisiológico que hay diferentes tipos de receptores afectados por el calor y el frío, y asumen que hay neuronas cuya actividad implicará una sensación de calor. Luego, asignan una neurona a cada función: recepción del calor, recepción del frío, sensación de calor y sensación de frío. Finalmente, ellos observan que la ilusión del calor corresponde a la siguiente relación entre tres neuronas: la neurona de la sensación de calor dispara, o bien en respuesta al receptor de calor, o bien a la breve actividad del receptor de frío. Considérese la siguiente red (**FIGURA 18**):

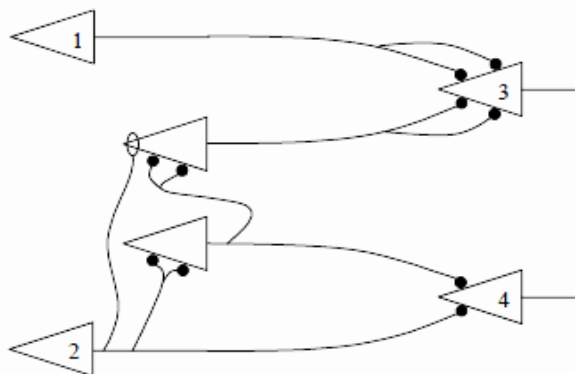


FIGURA 18: la neurona 3 de sensación de calor dispara sys recibe dos entradas, representadas por las líneas que terminan en su soma. Esto ocurrirá cuando, o bien la neurona 1, receptora de calor, dispara o si la neurona 2, de recepción de frío, dispara una sola vez y, luego, inmediatamente se detiene. Cuando la neurona 2 dispara 2 veces seguidas, la neurona intermedia sin número excita la neurona 4 en vez de la 3 generando la sensación de frío. (Figura extraída de Piccinini, 2004)

En este caso la misma ilusión descrita en términos no específicos funciona como SON, los elementos de la red serían el SO y el CC – utilizando una salida “bechteliana-mecanicista” – será el funcionamiento mismo de la red que debería dar a lugar al SON esperado. Si bien M&P señalan que, bajo otros supuestos sobre el comportamiento de los receptores del calor y del frío, la misma ilusión podría ser explicada por diferentes redes, esto, sin embargo, no implica que el IC como estaría siendo completado aquí sea una interpretación que no se ajusta a la definición que he desarrollado en la subsección (2.3.2.1.), sino que meramente significa que si se llegase a conectar de manera más específica ciertas funciones con ciertos redes, y a la vez tales redes fuesen específicamente rastreadas en la neurofisiología, entonces lo único que se estará haciendo en tal caso sería expandiendo el CC.

2.4.2. El perceptrón de Rosenblatt y la neurona de M&P

El trabajo original de Frank Rosenblatt apareció en un informe técnico el año 1957 financiado por la oficina de investigación naval de EEUU. Los desarrollos teóricos se hicieron públicos a través del artículo publicado en *Psychological Review*, llamado *The perceptron: a probabilistic model for information storage and organization in the brain* (1958). Tal artículo condensa no solamente varios modelos de perceptrón, sino que también puede ser considerado como preliminar a artículos posteriores de

Rosenblatt donde se llegaría a la formulación definitiva del perceptrón. No obstante, la formulación del perceptrón está bastante alejada de la manera en que es introducido de manera estándar en manuales de redes neuronales (Anderson, 2007; Rojas, 1996). Es por esto que del mentado artículo de Rosenblatt (1958) extraeré algunas nociones teóricas generales respecto a cómo el perceptrón se vincula con el problema más amplio de la explicación psicológica, además de algunas pautas respecto a su arquitectura. La formulación o CE no la tomaré directamente de Rosenblatt, sino que será a partir de manuales contemporáneos que permitan conectarla y contrastarla lo que se ha desarrollado en la sección anterior, a saber, la neurona de M&P.

El artículo de Rosenblatt comienza con las tres preguntas fundamentales que, desde su perspectiva, debiesen ser respondidas para entender la habilidad de los organismos superiores para el reconocimiento perceptual, la generalización, la recuperación de memorias y el pensamiento. Estas serían (Rosenblatt, 1958, 92):

1. ¿Cómo es la información del mundo físico detectada por los sistemas biológicos?
2. ¿En qué forma tal información es almacenada o recordada?
3. ¿Cómo tal información almacenada influencia tareas de reconocimiento y conducta?

La primera pregunta es más bien empírica y será respondida a medida que mejore la fisiología. Empero, para Rosenblatt, su teoría se encarga de lidiar con la segunda y tercera cuestión. La segunda pregunta es contestada mediante dos posturas opuestas: primero, aquella posición que sugiere que la información es codificada en representaciones o imágenes, mediante alguna clase de mapeo 1 a 1 entre el estímulo sensorial y el patrón almacenado. La segunda postura considera que el sistema nervioso central actúa como una red de interruptores intrincada, donde la retención toma la forma de una nueva conexión o vía entre los centros de actividad. El rasgo distintivo de esta segunda posición teórica es que nunca hay un mapeo simple de un estímulo a la memoria, a partir del cual algún código permita su posterior reconstrucción. Cualquiera sea la información que es retenida debe, de alguna manera indeterminada, ser almacenada como una *preferencia por una respuesta particular*, es decir: “the information is contained in connections or associations rather than topographic representations” (op cit., 93).

Respectivamente, estas dos posiciones que difieren sobre el método de retención de la información, también diferirán respecto a la tercera cuestión antes apuntada, a saber, cómo la información almacenada influencia la actividad actual. El primer grupo, los teóricos de las memorias codificadas, están forzados a conceder que el reconocimiento de cualquier estímulo involucra algún modo de comparación sistemática entre el contenido de lo que está almacenado respecto de los patrones de entrada, de modo de determinar una apropiada respuesta. El segundo grupo considera que, puesto que la información almacenada toma la forma de nuevas conexiones, se sigue que un nuevo estímulo hará uso de tales nuevas conexiones, las cuales han sido creadas automáticamente activando la respuesta apropiada sin requerir ningún proceso separado de reconocimiento o identificación (op. cit.).

La teoría que presentará Rosenblatt va en línea con la segunda posición, una empirista o “conexionista”, y su contenido corresponde a un sistema nervioso hipotético, o máquina, denominado “perceptrón”. El perceptrón está diseñado para ilustrar algunas propiedades fundamentales de los sistemas inteligentes

en general, sin abocarse en profundidad en los específicos y frecuentemente desconocidos detalles biológicos de cada particular organismo. La razón de desarrollar esto es su disconformidad con propuestas teóricas previas, las cuales considera inadecuadas por su lenguaje simbólico lógico y uso de álgebra Booleana. A este inconformismo se agrega que las teorías que están directamente preocupadas por el sistema nervioso biológico y su actividad en el medio ambiente natural son menos exactas en sus formulaciones y muy poco rigurosas en sus análisis, luego tales contribuciones quizás puedan ser consideradas como sugerencias más que como sistemas teóricos definitivos (op cit., 93-4).

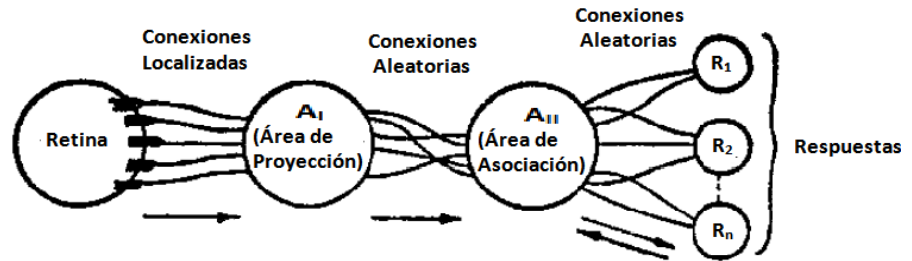
Las teorías sobre las cuales la teoría del perceptrón se basa se resumen en cinco puntos (op. cit., 94-5):

1. Las conexiones físicas del sistema nervioso las cuales están involucradas en el aprendizaje y el reconocimiento no son idénticas entre los organismos. La construcción de las redes más importantes es fuertemente aleatoria, sujeta a un número mínimo de restricciones genéticas.
2. El sistema original de células conectadas es capaz de cierta cantidad de plasticidad
3. Mediante exposición a una muestra grande de estímulos, que sean más “similares”, se tenderán a formar vías a los mismos conjuntos de células de respuesta. Aquellas, las cuales sean marcadamente “disimilares” tenderán a desarrollar conexiones a diferentes conjuntos de células de respuesta.
4. La aplicación de refuerzo positivo y/o negativo puede facilitar o inhibir cualquiera sea la formación de conexiones que esté llevándose a cabo.
5. La “similitud” en tales sistemas estará representada en algún nivel del sistema nervioso por una tendencia de estímulos similares a activar el mismo conjunto de células. La similitud no es un atributo necesario de una clase particular formal o geométrica de estímulo, sino que depende de la organización física del sistema que percibe, una organización que evoluciona a través de la interacción con un ambiente dado.

Hasta ahí, Rosenblatt se dedica a la fundamentación exclusivamente teórica. A continuación describe la arquitectura de su modelo del siguiente modo (op.cit., 95-6):

1. El estímulo incide en una retina de unidades sensoriales (puntos-S), los cuales se asume que respondan sobre la base del todo-o-nada en algunos modelos, o con un pulso de amplitud o frecuencia proporcional a la intensidad del estímulo en otros modelos.
2. Los impulsos son transmitido a un conjunto de células asociativas (unidades-A) en un “área de proyección” (A_i). Esta área puede ser omitida en algunos modelos donde la retina está conectada directamente con el área asociativa (A_{ii}). Las conexiones desde alguna unidad S a otra A pueden ser excitatorias o inhibitorias y, si la suma algebraica de la intensidad de impulsos es igual o mayor al umbral ϑ de cierta unidad A, entonces tal unidad dispara sobre la base de la ley todo-o-nada.
3. Entre el área de proyección y la de asociación (A_{ii}), las conexiones se asumirán como aleatorias. Más allá de la distribución de conexión, las unidades A_{ii} son idénticas con las unidades A_i , y responden bajo condiciones similares.
4. Las “respuestas”, R_1, R_2, \dots, R_n son células (o conjuntos de ellas) las cuales responden de manera similar a las unidades A. Cada respuesta tiene, normalmente, un gran número de puntos de origen localizados aleatoriamente en el conjunto A_{ii} . El conjunto de A-unidades transmitiendo impulsos a una respuesta particular se denominará el *conjunto-fuente* de tal respuesta.

La figura que presenta Rosenblatt (op. cit.) resume los anteriores puntos esenciales de su arquitectura (**FIGURA 19** a la derecha). Las flechas indican la dirección de la transmisión a través de la red. Nótese que las conexiones hasta A_{II} son hacia adelante y no hay realimentación. Cuando se llega al último conjunto de conexiones, entre A_{II} y R , existirá transmisión con retroalimentación. La regla que gobierna las conexiones puede seguir una de las siguientes alternativas:



(a) Cada respuesta tiene una respuesta excitatoria al complemento de su propio conjunto-fuente.

(b) Cada respuesta tiene una conexión inhibitoria de retroalimentación de su propio conjunto-fuente, es decir, tiende a prohibir la actividad en cualquier asociación celular que no transmita a ella.

A partir de aquí introduce diferentes redes arbitrariamente nombradas (α , β , γ) y con diferentes versiones de reglas de aprendizaje – en las células del sistema alfa se incrementa el valor de sus actividades de manera constante, en el beta, el conjunto-fuente tiene una ganancia constante en proporción a su actividad, finalmente en el gamma la ganancia se da respecto a las células inactivas del conjunto-fuente (op.cit., 97-8). Los detalles a continuación del artículo analizan tales sistemas desde una perspectiva estadística, es decir, con una retina de infinitos puntos extrayéndose ecuaciones que modelen tal comportamiento (op.cit., 98-99), por lo tanto difieren notablemente de las versiones contemporáneas que exponen el perceptrón desde una notación que he denominado “conexionista genérica” que utiliza escalares, vectores, matrices y operaciones sobre ellas (Demuth et al., 2014, cap. 2; Anderson, 2007, 218-225).

Esta discordancia entre la formulación original y la contemporánea me lleva a una observación capital: puesto que la idea central de esta sección no es desarrollar en detalle el perceptrón, sino que *contrastar el perceptrón y la neurona de M&P*, entonces lo que se necesita es trasladar el modelo del perceptrón a una estructura tipo-IC, es decir, especificar (entre otros) el CE de tal IC. Sin embargo, sería necesario cierto recorrido de reconstrucción histórica para llegar hasta las versiones contemporáneas que exponen el perceptrón con, por ejemplo, un sesgo – “bias” que dentro de la literatura contemporánea permite transformar el umbral en una de las ponderaciones dentro de la función sumatoria –, u otros elementos expresivos los cuales en el original de Rosenblatt no se encuentran¹¹⁸. En autores inmediatamente posteriores, por ejemplo Block (1962), la formulación del perceptrón (Block, 1962, 143) es más cercana a la de manuales contemporáneos y es en particular en tal artículo de H. D. Block donde se prueba el teorema de convergencia del perceptrón y no en el propio Rosenblatt (1958)¹¹⁹. Por consiguiente, y tratando de no expandirme en una investigación lateral más compleja, *asumiré que hay cierto CE común al perceptrón y neurona de M&P*.

¹¹⁸ En cualquier caso, la noción de umbral transformada a cierto sesgo como parámetro ajustable que se representa como cierta activación a_0 o peso $w[0]$ ya está presente en el ADALINE dos años después (Widrow & Hoff, 1960). Véase más adelante cómo se pueden describir las salidas del perceptrón de tal modo.

¹¹⁹ Este teorema es el que permite demostrar que, en caso de que los patrones de entrada sean linealmente separables, entonces existe una solución posible en la cual el perceptrón se asentará mediante entrenamiento

El requerimiento de cierto CE común al perceptrón y a la neurona de M&P no es meramente una restricción arbitraria a la que estoy acudiendo para mostrar la competencia entre ambos modelos mediante la tesis de los IC, sino que es lo que, al fin y al cabo, *es una primera heurística para presumir preliminarmente conmensurabilidad* (cuyas definiciones más atingentes en este caso están en Def.8.4., Def.8.5. y Def.8.6.) a la que estoy recurriendo en este caso específico. La existencia de CE comunes, obviamente, no es suficiente para asegurarse que existan equivalencias de F y SO, las cuales son muchísimo más básicas que la equivalencia entre CE. No obstante, en este caso específico el que se compartan CO – por ejemplo, al usarse de inspiración el funcionamiento neuronal y la “ley de todo-o-nada”, tanto en Rosenblatt (1958), como en McCulloch & Pitts (1943) – fácilmente individualizables es un buen indicio, el cual permite deducir que *es a partir del CE que será más fácilmente identificables los elementos de los conjuntos que se buscan sigan sean equivalentes según Def.8.5.*

Entrando en materia, en el cuadro resumen de la neurona de M&P y la descripción de su evolución de formulaciones – y el simplísimo ejemplo de cómo AND, en una neurona M&P puede representar la respuesta de un gato ante el estímulo de algo “que se mueve” y “es pequeño” (dentro de la **FIGURA 17**) – se describe de “manera conexionista genérica” el umbral como:

$$\vartheta = \sum_{i=1}^n WN_i(t-1)$$

Dentro de las formulaciones “conexionistas genéricas” del perceptrón, éste es descrito como una arquitectura construida a partir de neuronas de M&P – denominadas también “unidades de lógica de umbral” (ULU) dada sus activaciones en escalón o simplemente también “neuronas lógicas” – las cuales, de modo de no ser descritas apelando a un umbral ϑ con cierto valor, se transforma tal umbral en un *peso extra*, tal como $w[0]$ (o $w[n+1]$, con n la dimensión de la red). Se describirán sus salidas, entonces, como (Anderson, 2007, 219-220):

$$\text{salida ULU} = -1 \text{ si } \sum w[i]x[i] > 0.$$

$$\text{salida ULU} = +1 \text{ si } \sum w[i]x[i] \leq 0.$$

Es fácil ver que en “ $x[i]$ ” es equivalente a los “estados de activación” de la neurona M&P, es decir, los diferentes $N_i(t)$ dentro de la sumatoria, sin embargo “ $w[i]$ ” es un elemento que no existía en la función para calcular el umbral en M&P. Esta incongruencia, o no equivalencia, entre ambas ha sido advertida al lector en la sección anterior en una nota donde asevero que es mediante la “ W ” igual a 1 que es más fácil visualizar cómo se produce la continuidad y “superación” de la teoría de M&P por parte del perceptrón de Rosenblatt. En efecto, Rosenblatt introduce la noción de “peso” de una manera bastante más indirecta de lo cual uno puede *a priori* imaginar, pero tal noción será la *diferencia formal fundamental* que permite decir afirmar que el IC de Rosenblatt *expande* al de M&P (Rojas, 1996, 56; Scott, 2002, 237; Bermúdez, 2014, 220-1). El modo de introducir la existencia de “pesos” es, siempre y cuando mi lectura sea la correcta, la siguiente:

It will be assumed that the impulses delivered by each A-unit can be characterized by a value, V , which may be an amplitude, frequency, latency, or probability of completing transmission. If an A-unit has a high value, then all of its output impulses are considered to be more effective, more potent, or more likely to arrive at their endbulbs than impulses from an A-unit with a

lower value. The value of an A-unit is considered to be a fairly stable characteristic, probably depending on the metabolic condition of the cell and the cell membrane, but it is not absolutely constant. It is assumed that, in general, periods of activity tend to increase a cell's value, while the value may decay (in some models) with inactivity (Rosenblatt, 1958, 97).

En muchos otros sentidos la neurona y las redes de M&P frente a los perceptrones de Rosenblatt son equivalentes: ambos IC abstractos incompletos postulan que sus elementos pueden conformar diferentes arquitecturas, que en tales arquitecturas la información puede transmitirse en dos direcciones – el caso de las “redes con círculos” en M&P –, ambos IC recurren a una versión discreta del sistema nervioso que se trasluce en el uso de nodos y conexiones, ambos IC utilizan la ley del “todo-o-nada” o ambos IC dividen la neurona en dos “cajas” (un “integrador” y una “no linealidad”). Son estos diferentes elementos los que, indistintamente que puedan ser parte de los CE, CO y SO de cada IC, permiten definir que hay conmensurabilidad a través de equivalencias cuando son constreñidos a cierto SON.

La existencia de pesos implica diferencias anexas, por ejemplo, la existencia de reglas o algoritmos de aprendizaje en el perceptrón frente a la neurona lógica. Dado que ambos son de inspiración neuronal, entonces desde un punto de vista biológico es claro que las redes que pueden ser construidas con las neurona de M&P no son muy relevantes para todo tipo de función cognitiva, al ser demasiado cercanas a las compuertas lógicas, debiéndose la red *especificarse completamente antes de ser usada. No existen parámetros libres los cuales puedan ser ajustados a diferentes problemas.* En el caso de una red lógica la función del aprendizaje, solamente, podría ser implementada mediante la modificación de patrones de conexión de la red y los umbrales de las unidades, pero esto es necesariamente más complejo que sólo mediante un ajuste de parámetros numéricos (Rojas, 1996, 55).

Pues bien, la existencia de elementos en un IC que expanden ciertos conjuntos de otro IC anterior es, como he apuntado antes – sección (2.3.2.2.) – *“un criterio de decisión que permite dirimir entre dos IC conmensurables en competencia”*, pero, restringiéndose al orden en el cual debe hacerse el análisis desde la tesis de los IC, antes de llegar a tal punto se debe, primero, trasladarse dos modelos a teorías a estructuras-tipo IC y, segundo, individuarse la equivalencia mostrándose dónde están los elementos que se agregan al IC que se trata de superar. El segundo paso ya ha sido asegurado trabajando exclusivamente sobre IC abstractos, sin embargo el primer paso cuya idea central está condensada en la Def.8.4. (cito: *“La conmensurabilidad entre dos IC se da con dos condiciones necesarias: ambos deben ser concretos puntuales y ambos deben compartir SON.”*) no ha sido detallado un poco más, aunque sea exclusivamente mediante indicaciones.

El modo en que, como se vio en la sección (2.4.1.), un IC abstracto incompleto pase a ser concreto puntual es mediante la adhesión de un SON o algún conjunto componente que funcione como tal, en este caso el perceptrón en su exposición no ha ejemplificado su aplicación a una función. Sin embargo, no bastaría especificar cualquier función, sino que debe existir también una que sea equivalente en el lado de la neurona de M&P de modo de proceder a cumplir ambas condiciones de la Def.8.4. simultáneamente. La existencia de pesos y algoritmos de aprendizaje por sí solos no implican una ventaja de un IC frente a otro sin un SON que sea común y que precise tal ventaja, ya que tales elementos aún se mueven en el nivel de los IC generales.

Ahora bien, ya que para la neurona M&P se dio como caso el SON = “ilusión del calor”, algo similar debiese implementarse en el caso del perceptrón, pero hacerlo sería algo engorroso y no tan evidente como la estrategia siguiente: la manera más sencilla de elegir un SON común a las neuronas lógicas y las unidades del perceptrón es apelando a la implementación de conectivas; si se acepta que, por ejemplo, el sencillísimo caso del gato y el ratón cuenta como un modelo cognitivo que utiliza la neurona M&P como descriptiva y predictiva-explicativa, entonces lo que faltaría sería ver si el perceptrón es capaz de implementar conectivas lógicas, lo cual, en efecto *puede hacerse* (Dawson, 2005, cap. 11).

Las unidades del perceptrón, al ser representadas como ULU con pesos – no se olvide que todo esto es desde el marco contemporáneo utilizando tal CE como guía –, entonces podrán implementar todas aquellas conectivas que sean linealmente separables (Dawson, 2005, 83-4), luego serán iguales en generalidad que una neurona lógica (Rojas, 1996, 38-9), pero con una salvedad que ya ha sido apuntada: *las unidades del perceptrón no necesitan estar cableadas para responder como una conectiva antes de ser usadas, sino que pueden ser entrenadas para que respondan como una*. En términos del caso del “gato” que es un simple AND, una unidad de perceptrón que implemente AND no requiere ser configurada previamente como una unidad de M&P. Por lo tanto, siempre y cuando las conectivas se permitan como SON, entonces es posible visualizar como se tiene un SON equivalente y cómo a partir de éste el desempeño de un IC es “mejor” a otro respecto a tal SON.

En síntesis, es posible ser exhaustivo y mostrar las diferentes equivalencias para *evaluar el grado de conmensurabilidad local* (Def.8.6.), por ejemplo escalando desde el F que da lugar al SOE neurona M&P y el F que da lugar al SOE perceptrón de Rosenblatt, pasando por el SO propiamente tal de cada uno hasta llegar a cómo los CO tienen elementos comunes que pueden ser tratados bajo el mismo CE¹²⁰. Sin embargo, el punto que desarrollo en esta sección es sencillamente mostrar cómo desde las equivalencias en la Def.8.5. es posible encontrar una conmensurabilidad local y cómo de ésta funciona también para mostrar cuándo la competencia es susceptible de ser dirimida a favor de uno u otro IC.

Esta última cuestión es sutil y es menester señalarla; por una parte, la tesis de los IC es de trasfondo pluralista, por tanto no se adjudica ninguna pretensión de establecer criterios de decisión respecto a cuál IC es “mejor” o “peor” respecto a otro. No obstante, por otra parte, desde una pura búsqueda descriptiva de cómo tratar con el concepto general de “competencia” y el más específico de “conmensurabilidad” se llega a un criterio que permite visualizar bajo cuáles condiciones – y por

¹²⁰ Una advertencia: el CO del perceptrón consiste en nociones que no son compartidas por el CO de la neurona lógica. Rosenblatt es explícito en desechar la notación booleana y las formulaciones lógicas, por lo tanto su CO no contendrá proposiciones, ergo es posible, por ejemplo, preguntarse hasta qué punto rescata la intencionalidad el perceptrón frente a la neurona M&P, ya que la “intencionalidad” o el “carácter semiótico” de la neurona lógica venía dada por el supuesto que las neuronas al disparar, o no hacerlo, podían ser tomadas como valores veritativos verdadero o falso, mientras que en el perceptrón no existe algo semejante algo como “verdadero” o “falso” en su construcción. *Tal carácter semiótico es parte del CO de la neurona de M&P, ya que sin tan presuposición el uso mismo del formalismo original basado en la lógica no tiene sostén* (obviamente, es debatible hasta qué punto el uso del formalismo lógico por parte de M&P es suficiente para garantizar que la noción de “intencionalidad” es suficientemente adscrita a un modelo de lo mental, pero ello es discusión aparte). En síntesis, sin necesidad de adentrarse en qué es específicamente lo que es implicado ontológicamente por el perceptrón, *es una tarea bastante complicada ver puntos comunes entre ambos CO desde las formulaciones originales*, ya que tales puntos comunes solamente aparecen en una análisis que utiliza CE tipo “representaciones conexionistas genéricas”.

“condiciones” me refiero a la misma estructura tipo-IC – un IC “expande” a otro. Además, permite mostrar una forma *sobre la cual*, en una jugada filosófica ulterior, se puede pasar a decir que es *tal criterio el que dirime entre dos IC*. Dicho en otros términos, y apoyándome en lo que se ha tratado dentro de esta sección, el que las neuronas lógicas deban estar prefijadas de antemano para mostrar tal o cual comportamiento no es por sí sola una objeción a partir de la cual se derive el que sean “peores” para representar el aprendizaje respecto al perceptrón, sino que *se transforma en una objeción cuando nos circunscribimos a una idea o concepto de aprendizaje no innatista o empirista*, por ejemplo.

En el fondo, el que un modelo cualquiera nos sea desde nuestra subjetividad “más” útil que otro, no está determinado por el mundo aisladamente, sino que la *justificación* de tal utilidad ya viene en parte sesgada por una serie de TF de los cuales muchas veces los científicos son medianamente conscientes. La tesis de los IC, construida sobre el trasfondo pluralista está diseñada para ser “agnóstica” respecto a la necesidad de que un IC esté fundamentalmente ligado a un TF (i.e. son separables CO y TF – Def.4.2.) y los IC como artefactos o herramientas filosóficas están creados como una red de elementos ajustables que es arbitrariamente definida por cierta *hermenéutica mínima* que debe aplicarse cuando se habla sobre la actividad científica. En particular, esto último se puede observar, claramente, cuando *ex profeso* he utilizado el CE “conexionista genérico” para, de manera rápida, llegar a aislar a los “pesos” como la noción que distingue esencialmente a neurona lógica y perceptrón. Luego, la aplicación de la tesis de los IC a ciencias cognitivas no seguirá una lógica estricta para el cumplimiento de sus definiciones, sino que se deberá ir ajustando caso a caso, y sus conclusiones no deben considerarse como definitivas respecto a cierta relación interteórica, sino que meramente como iluminando cierta parcela bastante particular de alguna, mostrando cómo y bajo cuáles condiciones, una vez encontrada la “dimensión” que permite conmensurabilidad, un IC posee ciertas ventajas relativas respecto a otro.

2.5. La relación simbolismo-conexionismo a la luz de la tesis de los IC

La presente sección es la presentación de la aplicación de lo desarrollado antes, la tesis de los IC, a la primera parte de la tesis y se divide en seis apartados que siguen la secuencia argumentativa siguiente:

– i – *Pluralidad en las ciencias cognitivas*; La existencia de pluralidad desde la concepción de las ciencias cognitivas tiene una consecuencia directa en el análisis de la evolución de sus diferentes “paradigmas”, ya que la tesis de los IC es incompatible con una mirada “arriba-abajo” de las ciencias, entonces desde una mirada “abajo-arriba” se deriva que la evolución de las ciencias cognitivas es un entramado muchísimo más complejo de lo que análisis filosóficos anteriores se proponen. De esto se sigue dudar absolutamente de toda categoría anterior utilizada desde filosofía a ciencias cognitivas. Por consiguiente, se reconsiderarán las nociones de *simbolismo* y *conexionismo*.

– ii – *Respecto a una posible definición de “simbolismo”*; Simbolismo y cognitivismo *pueden y deben ser diferenciados*. Este deber surge precisamente de la necesidad de preservar CE y CO separados de los TF que fueron agregados posteriormente por los modeladores simbólicos paradigmáticos como son Newell y Simon. Luego de hacer esta separación sería preciso abocarse a definir “simbolismo”, pero existirán problemas de lograr una definición sin una adecuada investigación que, en el marco de la presente tesis, alargaría notoriamente su extensión. Por ende, se limitará la definición de “simbolismo” a la de “simbolismo cognitivista” defendida por Newell y Simon maduros, o por Fodor, Pylyshyn y McLaughlin.

–iii – *Respecto a una posible definición de “conexionismo”*; Conexionismo y subsimbolismo no necesitan ser diferenciados, ya que sus propios expositores hicieron las diferencias correspondientes. Esto, en teoría, facilitaría el trabajo de llegar a una definición de lo que es el “conexionismo”. No obstante, análogamente a lo que sucede con “simbolismo” no es un trabajo que pueda ser llevado a cabo fácilmente desde una perspectiva que lleve hasta sus últimas consecuencias la tesis de los IC. Por ende, se optará por definir de manera muy acotada al conexionismo, es decir, como “conexionismo subsimbólico” cuyo representante más obvio será Paul Smolensky.

– iv – *¿Competencia o colaboración?*; Restringidas las nociones a comparar mediante IC a “simbolismo cognitivista” y “conexionismo subsimbolista”, entonces se procederá a buscar qué clase de conmensurabilidad es posible achacar a tal relación “inter-teórica”. La conclusión a la que se llegará es que no satisfacen ninguna forma de conmensurabilidad, a menos que se agreguen SON, en cuyo caso es posible encontrar una conmensurabilidad ambivalente: a veces mínima, a veces local. Dada esta conclusión, entonces la relación entre ambos no es ni de colaboración ni de competencia, y cuando se trata de SON acotados, entonces la colaboración podría lograrse mediante cualquier pluralismo epistemológico adscrito por ambos bandos de investigación.

– v – *Posibles objeciones a la conclusión anterior y sus respuestas*; La forma más rápida de desechar la conclusión central a la cual se llegó en el apartado anterior es desestimar el entramado de supuestos que permitió la construcción de la tesis de los IC. No obstante, considero que para desestimar tal entramado sería necesario resucitar estrategias epistemológicas cuyas falencias ya han sido denunciadas por el desarrollo teórico filosófico del pasado. En vista de ello, consideraré exclusivamente algunas posibles objeciones que se puedan levantar desde la suposición básica de la validez de la tesis de los IC.

– vi – *Re-comprensión y re-organización de la polémica simbolismo-conexionismo*; En la sección (1.3.1.) se han delineado una serie de problemas que, preveía, podrían ser solucionados desde la filosofía de las ciencias. Dado que el punto básico – reflexión número (.18) – o el “nudo gordiano” de la tesis es la noción de *competencia* y la tesis de los IC se hace cargo de caracterizar con más detalle tal noción, como he sucintamente desarrollado en la sección (2.4.1) y (2.4.2.), entonces este apartado dentro de la sección final de la tesis se encargará más o menos en detalle de las reflexiones desde la (.01) a la (.17), agrupadas en torno a ciertos tópicos y desde el vocabulario que he desarrollado aquí.

– (2.5.i.) – Pluralidad en las ciencias cognitivas

La tesis de los IC está construida asumiendo un entramado básico de tópicos histórico-filosóficos que he intentado desarrollar lo más posible a lo largo de las secciones (2.1.), (2.2.), (2.3.) y (2.4). Muchos otros supuestos quedan en el tintero y serán, en conjunto, una de las limitaciones que posee la tesis de los IC. Sin embargo, existe un hecho que fuertemente constreñirá el análisis de la polémica simbolismo-conexionismo, a saber, la *pluralidad del desarrollo de las ciencias cognitivas*, la cual no es similar a la pluralidad trivial que acaece en el conjunto total de las ciencias, sino que se traduce en una actitud pluralista en el modelamiento científico, la cual no se condice con un dominio del pluralismo como postura epistemológica de los investigadores. Es en esta pluralidad que surgen el “simbolismo” y el “conexionismo” – ambos aun sin una definición más específica, sino que aquí ligado a sus autores más representativos en la escena de filosofía. Una reconstrucción histórica más o menos basada en la clase de esquematización utilizada en la sección (2.3.2.3.) se presenta en la siguiente página (**FIGURA 20**);

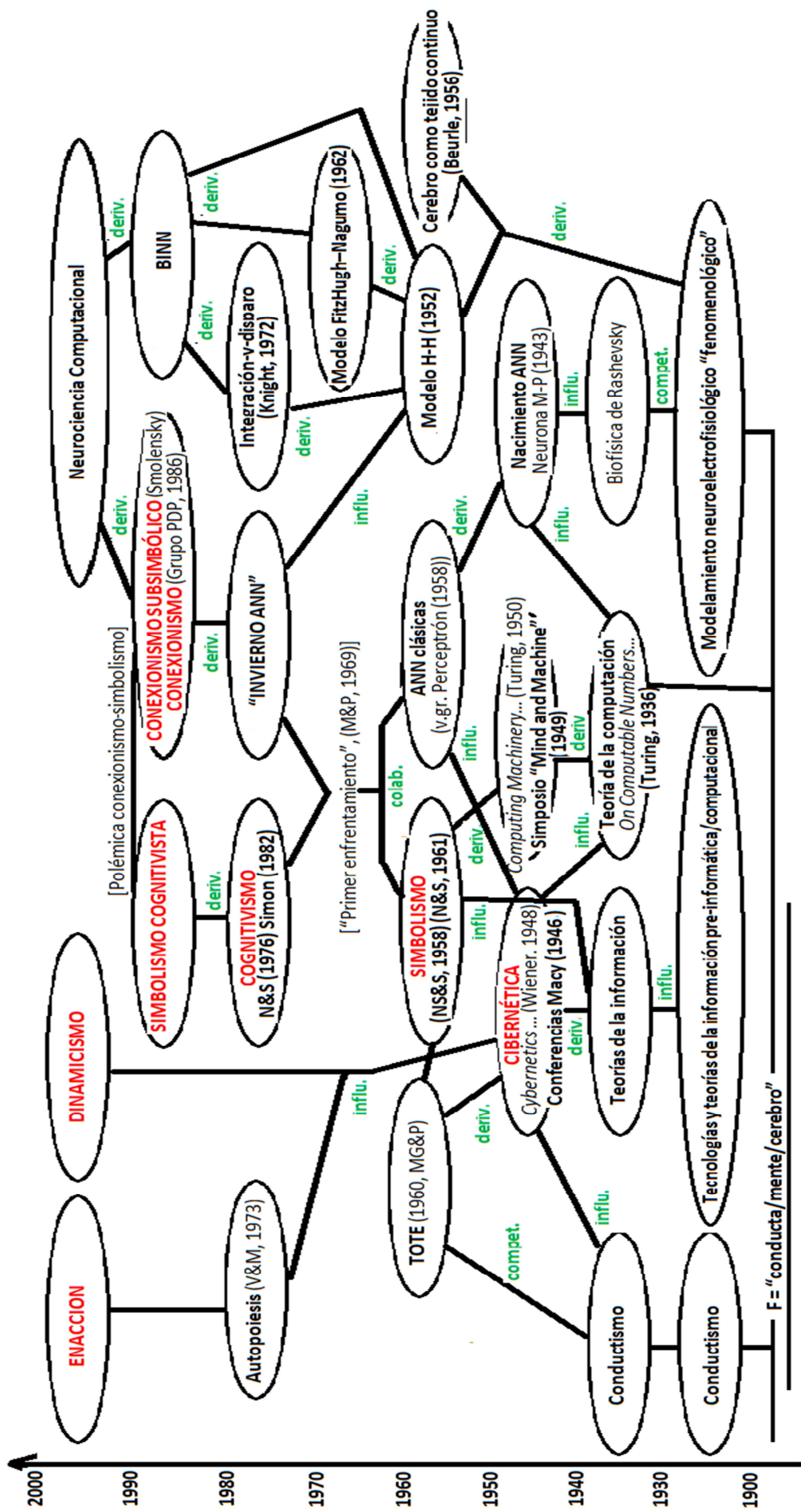


FIGURA 20; En este esquema reducido de la evolución de IC en ciencias cognitivas existen 4 tipos de relaciones "inter-teóricas" etiquetadas en verde: competencia, colaboración, derivación e influencia. La relación de derivación es más fuerte que la de influencia, siendo susceptible de explicarse mediante IC como "nacimiento derivado", mientras que la segunda la entenderé como una relación de derivación no probada que puede eventualmente transformarse en una de derivación.

Los diferentes "paradigmas" son etiquetados en rojo, no obstante, nótese, aparecen como una confluencia de teorías, simposios, modelos, etc., por lo tanto cada uno de los globos está completado con un elemento que a priori funciona indeterminadamente, o bien siendo un IC abstracto, o bien siendo un IC concreto. El acento aquí, contrariamente a los mapas polares (Varela et al./ Dennett), está puesto en la multiplicidad de relaciones sin existencia de hegemonía "paradigmática" obvia.

Existen muchísimas relaciones que quedan en el tintero de las cuales soy consciente, pero que por cuestiones de la dificultad de organizarlos claramente he dejado sin representar; por ejemplo las relaciones desde el conductismo a las ANN en su "invierno" o de Beurle y las teorías de la información.

En la **FIGURA 20** utilizo la relación de influencia para mostrar la posible relación de la cibernética con el enactivismo y el dinamicismo, relación que solo he encontrado – por ahora – anecdóticamente (Port & Van Gelder, 1995; Stewart et al., 2010; Novikov, 2016), así como también la del conductismo con la cibernética (Rosenblueth et al., 1943) y la del trabajo de Nicolás Rashevsky y M&P (Abraham, 2004). Existen otras influencias que requieren aun mejor confirmación, como la generada desde la teoría de la computación a la cibernética, y la del modelo Hodgkin-Huxley en la miríada de investigaciones que, se supone, fue el “invierno de las ANN”, donde muchos investigadores como Shun’ichi Amari, Stephen Grossberg o Paul Werbos asentaron las bases teóricas del de las ANN – y por extensión del conexionismo –, las cuales se desarrollarán explosivamente en la década de los 80’ (Anderson & Rosenfeld, 1998). Vale la pena registrar aquello, ya que la existencia de tal “invierno” es parte común de las versiones históricas estándar del conexionismo, las cuales saltan directamente de la crítica de Minsky y Papert (1969) al conexionismo de investigadores del grupo PDP (Medler, 1998).

La competencia es escasamente representada en este esquema. Solamente tengo certeza de TOTE respecto del conductismo, al menos considerando el elemento del “arco reflejo” perteneciente a tal escuela. La competencia de Rashevsky y modelamiento neuroelectrofisiológico me consta solo indirectamente. Siguiendo la tesis de los IC, en general no se puede tener certeza de competencia sin primero tener certeza de conmensurabilidad, y esta última sólo se da cuando, primero, se traslada una teoría o modelo a una estructura tipo IC y, segundo, se buscan relaciones de equivalencia.

La relación de derivación es la más común. Es interesante el caso del modelo de Hodgkin y Huxley, ya que es un ejemplo, no sólo de éxito científico arrollador, sino de cómo fue prolífico en generar modelos simplificados útiles para otros fines. En particular, el caso de la neurona de integración-y-disparo es esencial para comprender las redes neuronales biológicamente inspiradas (sigla en inglés “BINN”) o redes de pulsos (“spiking” NN) (Maass, 1997; Bishop & Maass, 1998; Scott, 2002; Anderson, 2007), las cuales junto con las ANN conforman el contenido principal de la neurociencia computacional contemporánea. La historia del modelamiento del cerebro como tejido continuo no me he atrevido a relacionarlo a la neurociencia computacional, aun cuando creo que con un poco más de exploración teórica sería fácil encontrar nexos.

La colaboración se dio a principios de los 60’ entre los grupos teóricos que me conciernen en esta tesis, es decir, entre simbolismo y los modelos “ANN clásicos” y también está representada en la **FIGURA 20**. Dentro de este periodo clásico se desarrolla no sólo el perceptrón (1958), sino que también Pandemonium (Seldfridge, 1958) y ADALINE (Widrow & Hoff, 1960), y la mejor muestra de colaboración es el volumen compilatorio de artículos *Computers and Thought* (1963), dentro del cual coexisten tanto el enfoque simbólico modelando las tareas cognitivas superiores, como redes neurales artificiales modelando percepción. Esta coexistencia no problemática es ignorada si uno no es revisionista al historiar el conexionismo (Berkeley, 1996), y el quiebre recién se da, marcadamente, con la obra *Perceptrons* (1969), la cual, según Dreyfus & Dreyfus (1988), sería representativa del “primer round” entre simbolismo y conexionismo. El lector notará ya, explícitamente, la separación de “simbolismo” de “cognitivism”, así como también la de “conexionismo” y “subsimbolismo”. A lo largo de toda la tesis no he entrecomillado los términos simbolismo y conexionismo simplemente por una cuestión estilística. Sin embargo, es imperioso ahora notar que haré en esta sección final borrón y cuenta nueva, de manera de

lograr hacer un análisis transparente que presente qué es lo que estoy entendiendo por cada término. La justificación de tal “borrón y cuenta nueva” se desarrollará a continuación y lo que se entenderá por cada término estará en (2.5.ii.) y (2.5.iii.).

Pues bien, mirando este esquema reducido de la evolución de IC en ciencias cognitivas, es posible observar diferentes “paradigmas”, los cuales en un nivel macroscópico pueden tratarse como tales, pero en un nivel inferior no tiene mucho sentido hacerlo, como he presentado mi caso con TOTE – modelo híbrido entre cibernética y simbolismo –, SAARCS, DINASTY o, más contemporáneamente, con SAL, ya que a nivel de la práctica existe un continuo más que límites rígidos. No considero en sí mismo el hablar de “paradigmas” en ciencias cognitivas como algo negativo o limitado, de hecho es bastante útil y pedagógico en ciertos contextos. No obstante, esta forma de conceptualizar se transforma en un lastre cuando uno entra al terreno de la *competencia* y la *comensurabilidad*, ya que esta última debiese ser investigada en una estructura con dos pasos lógicos de análisis: primero, *preguntarse cómo dos teorías se conectan cada una por su lado con un fenómeno* y, segundo, *cómo tales conexiones son – o debiese ser – el quid* de la investigación de la filosofía de las ciencias, *especialmente* en ciencias cognitivas.

Esta naturaleza “múltiple” de las ciencias cognitivas radica en nacer como una “súper-disciplina”¹²¹ con un F que *no está bien delimitado*, como he apuntado en el esquema reducido, yendo desde el cerebro, pasa por la conducta y llega hasta “lo mental” o cognitivo. Y, cuyos “grupos de teorías y técnicas” que dan origen a los elementos de los diferentes conjuntos componentes de los variados IC que se mueven en el intervalo temporal desde principios del siglo XX, pasando por 1930 y llegando mediados de los 40’, *tampoco estarían bien delimitados*¹²². Ergo, la “metáfora del computador” aparecería como un artefacto filosófico que ha nacido *a posteriori*, cuando se ha tratado de dar un sentido unificado a las ciencias cognitivas desde el intervalo temporal preponderantemente ubicado desde principios de los 50’ a principios de los 60’. Empero, solamente tiene sentido tal “artefacto unificador” si ya desde antes se

¹²¹ El tópico individual de si las ciencias cognitivas se considera una interdisciplina o transdisciplina no es algo que sea relevante para este punto. El término “súper-disciplina” lo he tomado de Dale (2008, 156) y en sí mismo no implica nada más que una visión cándida que establece que las ciencias cognitivas serían la suma simple de teorías que hablan de lo “mental” desde diferentes disciplinas, sin ninguna adenda que implique apelar a compromisos metodológicos o de otra clase (como hace Von Eckart, 1995). Mi definición de “súper-disciplina”, al no implicar nada específico, entonces es de por sí ya compatible con el pluralismo y, además, se condice con las *condiciones históricas originales de las ciencias cognitivas*; “In other words, in the beginnings cognitive science had an explicit negative theoretical commitment against behaviorism, yet did not have an equally explicit positive commitment to any specific theoretical framework” (Franchi & Bianchini, 2011, xvii).

¹²² Expandirme más sobre este punto requeriría un trabajo increíblemente arduo. Para sostener esto simplemente apuntaré un par de situaciones que se pasan por alto en la filosofía de las ciencias cognitivas: primero, que cierta noción de información ya existía en tecnologías decimonónicas como el telégrafo (Gleick, 2011, cap. 5) o el naciente teléfono – ambas tecnologías no computacionales – y que su primer tratamiento formal se da en la física estadística estando desacoplada de su carácter computacional (Shannon & Weaver, 1949, 3), carácter adquirido posteriormente con las tecnologías de la información de mediados del siglo XX. Segundo, la neurociencia computacional está al final de una línea cuyas raíces son increíblemente antiguas, con científicos como Galvani y Volta en el siglo XVIII (Finger, 2000, 119-125), pasando por investigadores de fines del siglo XIX y primera mitad del siglo XX como Nernst, Lapicque, Blair, Rashevsky y Hill, quienes serían los antecedentes al modelamiento cuantitativo cuyo exponente más famoso es Hodgkin-Huxley (Faraci, 2012). En efecto, de manera muy burda y simplificada, estos dos hechos han sido referidos en el esquema reducido como *Tecnologías y teorías de la información pre-informacional/computacional* y *Modelamiento neuroelectrofisiológico “fenomenológico”*.

tiene la preconcepción o sesgo monista explicativo de las ciencias cognitivas como una extensión de lo que ha venido a denominarse como “paradigma” “simbólico”.

Este fenómeno de pluralidad no podría ser observado si faltan una de las siguientes condiciones: primero, un trasfondo pluralista filosófico de las ciencias y una herramienta que sea esencialmente analítica respecto al tratamiento de teorías y modelos – como es el caso de la tesis de los IC. Un crítico podría aducir que si el punto se reduce a “monismo versus pluralismo”, entonces el usar el pluralismo no está justificado más que el monismo como sesgo para la historia de las ciencias cognitivas. No estoy de acuerdo con tal punto, ya que el problema de fondo de la “metáfora del computador” como “artefacto unificador” no es meramente el monismo, sino que lo es, por ejemplo, cuando se trata de historiar un IC “alternativo” al simbolismo, ya que para la historia de tal IC alternativo *no hay una explicación de cómo algo que en cierto punto histórico no era ciencias cognitivas, a partir de cierto punto, por el mero hecho de entrar en discusión con cierta ortodoxia*¹²³ *se transforma automáticamente en “ciencias cognitivas”.*

Mi argumento entonces es sencillísimo: la manera más simple de evitar el problema de explicar cómo en algún punto algo que no se considera parte de ciencias cognitivas pasa a serlo es considerar que nunca no ha sido ciencias cognitivas o, en términos no negativos, *que siempre lo ha sido desde su concepción.* Esto reduce los grados de arbitrariedad a exclusivamente elegir un punto histórico del desarrollo de ciencia cognitivas como relevante (mi propuesta es mover tal punto a las conferencias Macy, ya que la cibernética permite conectar muchísimos más modelos que las conferencias llevadas a cabo en 1956), mientras que desde un historia de las ciencias cognitivas no pluralista los grados de arbitrariedad son mayores, ya que no sólo se elige un punto cronológico como punto de nacimiento, sino que se restringen los elementos de conjunto componentes al tomarse ciertos IC más restrictivos como “fundadores”.

El esquema reducido puede ser tratado como base para desarrollos teóricos múltiples. Por ejemplo, cada etiqueta en rojo puede ser desarrollada de manera autónoma respecto a sus orígenes. Otra manera de usarlo sería, siendo más pulcro, utilizarlo como guía para rastrear la ubicación relativa de un modelo o teoría – un IC – respecto de otros. Sin embargo, el esquema reducido, más allá de estos usos, lo considero por sí solo como base para declarar y defender las siguientes cuestiones:

- Primero, y lo más obvio, que el desarrollo de las ciencias cognitivas ha sido pluralista, es decir, sus modelos se han desarrollado normalmente bajo una actitud investigativa pluralista, aunque no necesariamente bajo una epistemología pluralista. Por tanto, expandir el esquema anterior, más que mostrar cómo la historia es monista, generaría lo contrario, es decir, una mayor pluralidad.
- Segundo, dado que la tendencia general en el modo en el cual se investiga y modela la cognición es pluralista, entonces las excepciones son modelos-hitos (v.gr. GPS), los cuales pueden servir como “definitorios” respecto a cierto “paradigma”, pero solamente como una herramienta simplificada para analizar las relaciones entre “diferentes paradigmas” sin representar una realidad. Una visión pluralista de la historia del modelamiento cognitivo puede aceptar ciertos hitos como significativos para elaborar una noción mejor desarrollada de simbolismo, conexionismo, etc., pero por sí sola no apoya la versión simplificada de escuelas bien separadas hasta sus bases. *La norma es lo híbrido.*

¹²³ De hecho, el mismo aseverar que existe “ortodoxia” lo considero no muy justificado, sino es apelando al mismo monismo que la pone como centro de la historia.

- Tercero, la realidad es que las ciencias cognitivas han enfrentado el gran problema que nunca ningún “paradigma” (etiquetados en rojo dentro del esquema) ha pasado la fase intermedia de asentamiento-madurez exitosamente – al menos dentro del periodo que estoy tomando –, lo que significa que no se puede apelar a la “naturalidad” de ningún concepto – digresión 1 en la parte primera de la tesis. Esto tiene una consecuencia directa en la polémica conexionismo-simbolismo: explica de manera sencilla porqué la fase de la disputa entre Smolensky y Fodor fue una *adenda innecesaria* a la relación conexionismo-simbolismo.
- Cuarto, y finalmente, de lo anterior se deriva que es necesario reconsiderar las definiciones de “simbolismo” y “conexionismo”, las cuales a lo largo del desarrollo de la presente tesis han sido utilizadas en el sentido “arriba-abajo” que caracteriza la filosofía de las ciencias cognitivas, pero que no han sido reconsideradas desde la tesis de los IC.

– (2.5.ii.) – Respecto a una posible definición de “simbolismo”

Los autores que tratan de definir tal término lo hacen, desgraciadamente, sobre la base de supuestos no totalmente legítimos desde la perspectiva de los IC. Me afirmaré sobre ciertos pasajes dentro de la literatura de ciencias cognitivas para ir desarrollando la definición de simbolismo en la cual desembocaré. Considérense los siguientes párrafos:

“Classical cognitivism takes as its starting point the concepts of representation and computation. Very roughly, models of the mind are likened to a von Neumann architecture in such a way that cognitive processing boils down to the computational manipulation of representational inner states. That the brain is a piece of biological hardware, and the mind is the software running on top, means that it can be modeled as a Turing machine. Cognition consists in the rule-governed manipulation of symbols that Newell and Simon’s (1972) *Physical Symbol System* epitomized” (Calvo & Gomila, 2008, 2)

“1. Cognition essentially involves representations and computations. Representations are, in general, symbolic structures, and computations are, in general, rules (such as rules of logic) for manipulating those symbolic structures.

2. A good model for understanding mind-brain functions is the computer – that is, a machine based on the same logical foundations as a Turing machine and on the von Neumann architecture for a digital computer. Such machines are ideally suited for the manipulation of symbols according to rules (...).

The motivating vision here is that cognition is to be modeled largely on language and logical reasoning; having a thought is, functionally speaking, having a sentence in the head, and thinking is, functionally speaking, doing logic, or at least running on procedures very like logic.” (Churchland & Sejnowski, 2006 [1989], 158)

“I will refer to the traditional approach to cognitive modeling as the *symbolic paradigm*, intending to emphasize that in this approach cognitive descriptions are built of entities that are *symbols* both in the semantic sense of referring to external objects and in the syntactic sense of being operated upon by ‘symbol manipulation’.” (Smolensky, 1987b, 1)

Si uno cruza el contenido de los primeros dos párrafos antes citados con la referencia de Paul Smolensky que ha sido ya citada en la primera parte de la presente tesis, entonces la conclusión obvia a la que uno debiese llegar es afirmar que “cognitivismo” y “simbolismo” tienen exactamente el mismo contenido conceptual. De hecho, hay acuerdo general en tratar como sinónimos ambos términos – algo similar ocurre con un tercer término, “computacionalismo”, sin embargo creo que analizar los matices que indican semejanzas y diferencias respecto al simbolismo es un faena a acometer en otro momento¹²⁴ – al punto que sugerir, como haré a continuación, que ambos términos tendrían *elementos no compartidos* es prácticamente una herejía, que además de ir contra las definiciones estandarizadas en ciencias cognitivas, sería una tarea bastante ociosa ¿para qué dedicarse a distinguir entre ambos si, por ejemplo, toda la gama de autores anti-representacionistas – o los mismos del grupo neurocomputacional – avanzan tranquilamente en sus investigaciones habiendo ya desechado ambas nociones por igual?, ¿cuál es la utilidad práctica de hacer una distinción?

En cuanto labor filosófica es útil por sí misma cualquiera sea la clarificación conceptual. Sin embargo, en tanto estoy defendiendo una postura epistemológica como el pluralismo, resulta importante defender que el “simbolismo” es algo separado del “cognitivismo”, ya que separarlos permite extraer enseñanzas desde la historia del modelamiento cognitivo para la misma práctica en ciencias cognitivas. Un ejemplo que, especulo, puede ilustrar tal utilidad práctica potencial, estaría en que en modelos cognitivo-sociales que involucran la interacción de gran cantidad de agentes, los enfoques estrictamente neural-inspirados se vuelven más costosos computacionalmente. Esto me parece evidente intuitivamente.

No obstante, es mejor no conjeturar consecuencias sobre lo no llevado a cabo aun y pasar a cometer la herejía de diferenciar entre simbolismo y cognitivismo. El crítico ante esta intención heterodoxa lo primero que observará será “Es obvio que cognitivismo y simbolismo son sinónimos. No es necesario acudir a las interpretaciones de neurocomputacionistas u autores alternativos para darse cuenta de esto, basta ir a las mismas fuentes para determinar que son sinónimos”. Esta observación es correcta desde la perspectiva de cierta literatura de científicos cognitivos clásicos, los cuales explícitamente en tales textos tratan de dar un *fundamento teórico común a las Ciencias Cognitivas*. Tal meta es explícita en artículos como el de Pylyshyn (1980).

Pero la observación del crítico es parcialmente correcta, ya que tales autores clásicos fundaron sus estudios circunscritos a cierto SON y CC específicos más antiguos. Por ejemplo, si uno va a modelos clásicos como el GPS (Newell et al., 1958; Newell & Simon, 1961; Newell & Simon, 1963) lo que observará es que el SON corresponde a la tarea de explicar-prededir fenómenos/tareas bien definidos, tales como manipulación lógica, resolución de puzzles, jugar ajedrez y escribir programas (Newell & Simon, 1961, 2014-6). Respecto al CC, además de basarse en el programa computacional para la simulación del proceso cognitivo, utiliza protocolos – informes introspectivos – de los sujetos para validar su propio modelo. Estos dos puntos ya por sí solos muestran que enfoques alternativos y neurocomputacionales cimentan sus críticas en aspectos “superficiales” de los trabajos clásicos, es decir a nivel del CO y el TF, siendo que, desde la perspectiva de los IC, lo que validaría la existencia de

¹²⁴ Un buen punto de partida para ir dilucidando tal concepto podría ser la visión global que entrega Piccinini (2012). O si se quisiera hacer el trabajo desde sus bases se puede utilizar las mismas discusiones en las conferencias Macy de “digital versus analógico” que ha sido señalada en la sección (2.2.)

comensurabilidad y, por extensión, de competencia es lo que he definido en Def.8.4., Def.8.5. y Def.8.6., más particularmente en la existencia de equivalencia como he ejemplificado en la sección (2.4.2.). Sin estas condiciones, entonces la existencia de competencia no está bien justificada.

Si la observación que hago se considera legítima, entonces ella constituye el soporte de qué es lo que diferencia a cognitivismo de simbolismo; el cognitivismo surgiría a partir de estos primeros modelos simbólicos circunscritos a cierto SON específico, *sería un trasfondo general que presupone que esta clase de modelos es aplicable a la totalidad del F*, es decir, a todas las funciones cognitivas en conjunto¹²⁵. Así pues, esta clase de cognitivismo es lo que daría lugar, por ejemplo, a arquitecturas simbólicas clásicas como SOAR. Puesto de manera sencilla, la diferencia entre simbolismo y cognitivismo radica en que el simbolismo es una modelamiento cognitivo bien constreñido y anterior a la especulación sobre la posibilidad de extender tal formulación al resto de la cognición.

Luego, lo lógico sería preguntarse, siguiendo el esquema de la tesis de los IC, si el simbolismo posee un CE específico, o más bien simplemente posee cierto CO que es capturado mediante la noción de “símbolo” – cierto objeto discreto que sería la unidad mínima con significado cognitivo – y “regla” – que sería aquella entidad que permite el propio proceso cognitivo. La respuesta más inmediata es esta; individualizar un CE es una tarea casi imposible, ya que, por ejemplo, las formulaciones del GPS y una red semántica como la de Quillian son evidentemente diferentes, aun cuando poseen un CO común que permite, típicamente, dar por sentado que ambos pertenecen al conjunto de los modelos simbólicos¹²⁶.

Si se echa mano de la separabilidad entre TF y CO, entonces considero que *la relación entre cognitivismo y simbolismo está bien definida como una relación entre TF y CO respectivamente, siendo tal CO restringido dependiendo el caso concreto de modelo que se esté tratando* – es decir, dependiendo el SO. No existen recetas para llegar a decir que tal o cual definición corresponde de manera unívoca al cognitivismo, sin embargo se puede acudir a la literatura propia cognitivista para recopilar ideas que sirvan para dar contenido al TF cognitivista. Por ejemplo, un caso de idea constituyente del cognitivismo es la *hipótesis de símbolos físicos*, la cual supone dos cosas: que un sistema de símbolos físicos es suficiente por sí solo para producir inteligencia y que es necesario tener un sistema tal para exhibir inteligencia (Newell & Simon, 1976, 118), definiéndose tal sistema como “(...) a machine that produces through time an evolving collection of symbol structures.” (op. cit., 116), siendo tales estructuras de símbolos patrones de entidades denominadas “símbolos”, las cuales se combinan mediante procesos

¹²⁵ El simbolismo empieza a transformarse en “simbolismo cognitivista” a medida que extiende sobre la totalidad de la cognición ciertas propiedades que derivan, justamente, de que el propio simbolismo estaba circunscrito a cierto SON. Por ejemplo, rasgos como el procesamiento serial, memoria a corto plazo pequeña o una memoria a largo plazo con rápida recuperación, pero lento almacenamiento, serían en parte evidenciadas por el comportamiento de resolución de problemas mismo. En palabras de Newell y Simon: “No problem-solving behavior has been observed in the laboratory that seems interpretable in terms of simultaneous rapid search of disjoint parts of the solver's problem space. On the contrary, the solver always appears to search sequentially, adding small successive accretions to his store of information about the problem and its solution.” (Newell & Simon, 1971, 149).

¹²⁶ La red semántica de Quillian es un caso límite (Quillian, 1988 [1968]). En este párrafo, por mor del argumento, lo he colocado en el lado simbólico, como también habrían hecho los conexionistas del grupo PDP. Sin embargo, si uno parte de la noción de “conexionismo” que utilizan F&P, entonces las redes semánticas serían conexionistas. Esta ambigüedad de las redes semánticas no debe perderse de vista.

que las crean, modifican, reproducen y destruyen. Otro caso, puede ser la especulación de Herbert Simon de poder extender la explicación basada en sistemas de procesamiento de información a la función cognitiva de resolución de problemas a ambientes de tareas no estructurados (Simon, 1978, 286-7). Sintetizando, quizás una de las frases que mejor representa la metafísica del cognitivismo sea la de Newell y Simon en 1972 "(...) the programmed computer and human problem solver are both species belonging to the genus IPS [Information Processing System]" (Martínez-Freire, 2011, 82).

Sin embargo ante la diversa gama de ideas que pueden recopiladas para formar el TF cognitivista, la pregunta será entonces ¿cuántos cognitivismos hay? Y la respuesta es, todos aquellos que arbitrariamente se quisiera construir. Cruzando este análisis con lo hecho en la primera parte de la tesis, parece claro tras un examen no profundo que el cognitivismo de Newell y Simon *tiene puntos de sutil diferencia* con el de Fodor, a quien se ha examinado en compañía de Pylyshyn y McLaughlin, de lo que se sigue que utilizar el apelativo de "cognitivismo" para agruparlos a todos parece una salida poco prolija. Para muestra un botón; Newell y Simon dirán respecto a la conexión entre su propuesta simbólica y la neurofisiología lo siguiente:

"Explanation in psychology will not rest indefinitely at the information-processing level. But the explanations that we can provide at that level will narrow the search of the neurophysiologist, for they will tell him a great deal about the properties of the structures and processes he is seeking. (...) We cannot claim to see in today's literature any firm bridges between the components of the central nervous system as it is described by neurophysiologists and the components of the information-processing system we have been discussing here. But bridges there must be, and we need not pause in expanding and improving our knowledge at the information-processing level while we wait for them to be built." (Newell & Simon, 1971, 157-8)

Esta explícita espera por una conexión con la neurofisiología es, considero, una muestra de una autonomía menor que la defendida por Fodor de la explicación cognitiva respecto del "nivel de implementación". Hay en común una preponderancia de lo "arriba-abajo", sin embargo se concede la necesidad de conexión con el nivel inferior. Por otra parte, el uso de los niveles de análisis de Marr no es algo que pueda verse en el contexto de Newell y Simon, más bien parece claro que son Newell y Simon los que inspiran a Marr, siendo las ideas de Marr contemporáneas a las de Fodor. En general, una manera fácil de caracterizar el cognitivismo fodoriano es que agrega un componente ontológico extra que no se encuentra en Newell y Simon, a saber, la existencia de un lenguaje del pensamiento, el cual, en cuanto puede constreñirse a cierto SON, podría llegar a funcionar como un IC concreto puntual. No obstante, LOT no es justamente un ejemplo de IC concreto puntual, ya que también posee el rasgo de ser aplicable a todo el espectro de F, equivalentemente a lo que hace el cognitivismo de Newell y Simon cuando se "exporta" desde un SON a la totalidad del F mediante una generación de elementos extras.

Dado que construir una definición de cognitivismo y simbolismo de manera acabada requiere, según la propia prescripción de la tesis de los IC, y el esquema del apartado (2.5.i.) que esboza cómo el simbolismo se levanta sobre una pluralidad no bien estudiada de otros enfoques (lo que obliga, por ejemplo, a la investigación integral de contextos de descubrimiento y justificación, la construcción de una secuencia de IC concretos puntuales desde los cuales se extraiga un IC abstracto generalizado u otras restricciones inherentes a las definiciones que he desarrollado en la sección (2.3.2.) y sus

subsecciones respectivas), entonces tomaré un atajo utilizando el mismo contexto de la polémica entre simbolismo y conexionismo desarrollada en la primera parte de la tesis, es decir, muy sintéticamente me atenderé a lo siguiente:

- Simbolismo y cognitivismo no son sinónimos ya que serán definidos, respectivamente, como *el CO y el TF de un modelo simbólico cualquiera*. La pista clave que me permite hacer esta diferencia es la lectura directa de autores clásicos y constatar cómo sus esfuerzos se restringen a SON delimitados, mientras que el cognitivismo surge ulteriormente como un proyecto sobre la posibilidad de extender, no sólo la forma de modelamiento, sino que ontologías y/o métodos.
- Dada la separabilidad CO y TF (Def.4.2.), entonces “cognitivista” será quien *adscriba fundamentalmente* a cierto TF específico sobre la cognición y no meramente *adscriba heurísticamente*¹²⁷. El que los modeladores estrella simbolistas sean cognitivistas, *no es suficiente desde la tesis de los IC para decir que todo modelamiento simbolista es cognitivista*.
- El cognitivismo de Newell y Simon difiere del de Fodor, por tanto, arbitrariamente – y por motivos puramente prácticos dentro de esta misma tesis –, utilizaré el cognitivismo fodoriano como la definición de cognitivismo. Tal cognitivismo puede ser entendido esencialmente como lo caracterizó Smolensky (1987a) donde se “gradúa” posturas desde el racionalismo al eliminativismo.
- Debe existir un CO para tal cognitivismo – o, lo que es equivalente decir, algún simbolismo para tal cognitivismo –, luego, circunscribiéndome a lo desarrollado antes, tomaré los rasgos arquitecturales defendidos por F&P como la “teoría” a trasladar a una estructura tipo-IC, de lo que se obtiene, creo que con meridiana certeza, que *no hay motivos para suponer que tal IC será concreto puntual*. LOT en particular, que es nombrado en el transcurso de la crítica al conexionismo, es abiertamente filosofía de la psicología, por tanto puede considerarse un caso de IC abstracto. Empero, no puede aducirse que sea una generalización de modelos concretos, sino que es un desarrollo “extra” de las formulaciones originales de investigadores que sí desarrollaron IC concretos – Newell y Simon –, por lo tanto considero que lo adecuado sería decir que *el simbolismo cognitivista es un caso de IC abstracto incompleto* – con Fodor como representante paradigmático. Esta definición será retomado en (2.5.iv.).

– (2.5.iii.) – Respecto a una posible definición de “conexionismo”

La definición de conexionismo es relativamente más sencilla que la de “simbolismo”, ya que no hay mezcla entre el TF y el CO, por lo tanto no necesito el largo rodeo que precedió en el apartado anterior a la definición de simbolismo como distinto de cognitivismo. Esta situación “privilegiada” (desde la perspectiva de los IC) puede extraerse de pasajes de la exposición del grupo PDP cuando se dice que:

¹²⁷ Es preciso reconocer que he jugado con mis reglas desde el comienzo de la tesis, aun cuando no lo hice explícito; la separabilidad está presente desde el comienzo del presente trabajo cuando apunté por vez primera la expresión “simbolismo cognitivista” en la sección (1.2.). El cognitivismo es separable del simbolismo en cuanto muchas de las tesis que lo caracterizan (por ejemplo, la afirmación fuertemente metafísica de que los sistemas de procesamiento de información y solucionadores de problemas humanos pertenecen al mismo género) no son estrictamente necesarias para entender el funcionamiento de un modelo o teoría simbólica analizado según la óptica de los IC. Empero, esta separabilidad no puede ser observada, si antes no existe un marco que conlleve algo como la Def.4.2.

“We do not see ourselves capable as yet to produce the super-model which would connect all of our areas of exploration together. Rather, we feel that the PDP framework which we are developing forms a kind of *metatheory* from which specific models can be generated for specific applications. The success of the particular models reflects indirectly on the metatheory, but we feel that the proper approach is to study detailed models of detailed applications while at the same time keeping one eye on the bigger picture.” (Rumelhart & McClelland, 1986, 145).

O sea, y utilizando lo desarrollado en la primera parte de la tesis, el marco general de ocho elementos que fueron descritos al comienzo de la sección (1.1.1.) constituirían tal “metateoría” conexionista que sería agnóstica respecto a cuestiones como el innatismo o el empirismo (Rumelhart & McClelland, 1986, 139-142), es decir se define por un lado cierto CO y los TF que libremente pueden ser anexados o no. Los integrantes mismos grupo PDP no solamente divergían en los modelos formales que utilizaban para explicar-predecir, sino que también tenían diversos intereses respecto al futuro de la investigación PDP, pudiendo evidenciarse esto en cómo actualmente las investigaciones de, por ejemplo, Sejnowski versus la de McClelland (o versus la de Hinton que se ha dedicado intensivamente a *machine learning*) se han mantenido en niveles de detalle de modelamiento neuronal totalmente dispares.

Ahora bien, desde el esquema antes expuesto se supone una relación que iría desde la neurona de M&P y el perceptrón al conexionismo, empero una investigación cabal debiese involucrar muchísimos más elementos. Por ejemplo, la inclusión de cómo ADALINE (parte de las ANN clásicas) que usó por primera vez la regla delta, influyó en el grupo PDP en la formulación exacta de la regla delta generalizada (Rumelhart & McClelland, 1986, 319-322). *O cómo se relacionan los modelos PDP con otros modelos neuronal-inspirados que estaban avanzando paralelamente*¹²⁸ ¿se trataba de competencia o colaboración? Dado que una investigación exhaustiva requeriría muchísimos más elementos teóricos de los que dispongo, entonces una investigación centrada en un desarrollo desde los IC anteriores al conexionismo que permitan definirlo me es imposible pragmáticamente.

Por otra parte, una estrategia alternativa para definir el conexionismo puede ser dentro del marco amplio de las ANN. He ligado más cercanamente lo que he denominado “grupo de las ANN” al conexionismo, haciendo tal conexión tácitamente al utilizar manuales de ANN para extraer lo que podrían ser las “formulaciones conexionistas genéricas” de la neurona M&P y el perceptrón, empero tampoco considero adecuado afirmar que existe un IC claro que englobe a la totalidad de las ANN y al conexionismo. Por ejemplo, las redes neuronales se consideran como definidas por el paralelismo, dado que la computación de sus elementos constitutivos es en gran medida independiente una de otra, sin embargo existen varios modelos simples que escapan de tal regla, tales como una red alimentada hacia adelante con sólo dos unidades más un sesgo implementando una regresión lineal simple (Sarle, 1997).

¹²⁸ Extenderme en este punto sería a estas alturas bastante inconveniente. Sin embargo, en la introducción a la sección (2.4.1.) comenté que la diferencia *esencial* entre la neurona de M&P y otros modelos se debe a que es un modelo *no estrictamente eléctrico*. Los modelos que estoy pensando van en línea a modelos como los de Lapicque (1907) y Hodgkin-Huxley (1952). Pues bien, dicho en otras palabras, mi observación crítica es que, aun cuando compendios paralelos respecto al tópico amplio del modelamiento nervioso y cerebral toman en cuenta los volúmenes PDP (MacGregor, 1987), los volúmenes conexionistas prestan nula atención a su relación con otros enfoques.

Luego, dado que el CE puede ser establecido arbitrariamente – como he hecho en las secciones (2.4.1.) y (2.4.2.) –, pero muchos de los CO variarán dependiendo la ANN – como es el caso del paralelismo que podría haber sido considerado como parte de la ontología básica de ellas –, entonces es una mala idea utilizar el marco desde las ANN, o lo que es equivalente cualquier manual de redes neuronales contemporáneo, para individualizar el conexionismo, ya que, en términos sencillos, *del uso de un CE común no se seguiría impecablemente un CO común*. El uso de un CE extemporáneo para conmensurar entre perceptrón y neurona M-P se justifica por la propia investigación que constata elementos ontológicos comunes, pero si se quisiera hacer algo equivalente entre ANN y conexionismo, entonces habría que hacer el mismo estudio para verificar cómo ciertos elementos ontológicos son comunes mientras se explica por otro lado cómo otros se considerarían excepciones. Ello sería un trabajo posible de llevar a cabo, pero involucra una investigación paralela bastante pesada. Subsumir conexionismo a ANN es un individualización tipo “arriba hacia abajo” que choca con la tesis de los IC de inspiración “abajo hacia arriba”, es decir, buscando cómo aparece un SO y cómo sobre tal hay otros conjuntos componentes. En conclusión, utilizar una estrategia contraria a la “esencia” de la tesis de los IC debe estar de bien justificada por una investigación exhaustiva.

Siendo entonces claro que tampoco es una salida utilizar el marco de las ANN, entonces nuevamente recurriré a la estrategia pragmática utilizada en el apartado anterior con el simbolismo y cognitivismo, es decir, acudiré a lo que se hizo en la primera parte de la tesis y asumiré lo siguiente:

- Conexionismo y subsimbolismo no son sinónimos ya que serán definidos, respectivamente, como *el CO y el TF de un modelo conexionista cualquiera*. La lectura directa del grupo PDP contrastada con la postura de Smolensky permite por sí sola hacer la distinción.
- Análogamente a lo que se dijo sobre simbolismo y cognitivismo, la separabilidad entre CO y TF, permite distinguir entre “subsimbolista” y mero “conexionista”. Esto es fácil de ver, dado que desde el inicio del grupo PDP existió libertad en cómo cada integrante utilizaba el marco general conexionista o CO conexionista.
- Arbitrariamente utilizaré el TF subsimbólico de Smolensky, a pesar que existan otros TF notorios que tienen como CO al conexionismo, como es el caso de los Churchland, cuyo TF podría ser denominado “eliminacionismo”. La referencia básica es el propio Smolensky (1987a).
- Debe existir un CO para tal subsimbolismo, y tal CO puede considerarse el marco general conexionista desarrollado en la primera parte de la tesis. Dado que tal marco está separado de SON, entonces, si se toma al conexionismo subsimbolista como un IC, y se trata de trasladar a una estructura tipo-IC, *no hay motivos para suponer que tal IC será concreto puntual*. Ergo, el *conexionismo subsimbolista* se tratará de un caso de *IC abstracto-generalizado* – con Smolensky como representante paradigmático. Esta definición se retomará en la siguiente sección.

– (2.5.iv.) – ¿Competencia o colaboración?

La manera tradicional de enfrentar la pregunta respecto a la relación interteórica simbolismo-conexionismo es partir suponiendo que, puesto que parten del *realismo representacional* y utilizan el *vocabulario de los niveles de análisis de Marr*, entonces tales ámbitos *eo ipso* permiten decir que las

“teorías” del simbolismo y el conexionismo entran en competencia, en particular en su búsqueda común de dar cuenta de una *arquitectura cognitiva*.

No obstante, ¿es ello suficiente? Un examen desde la misma polémica muestra que tal perspectiva tradicional es demasiado optimista. Luego, tomando la diferencia hecha en los apartados anteriores de simbolismo/cognitivismo y conexionismo/subsimbolismo, si uno se remite a lo que explícitamente (y muchas veces tácitamente) dicen conexionistas y simbolistas cognitivistas, sus ideas de qué se entiende por “arquitectura” se diferencian de manera crítica, no pudiéndose encontrar equivalencia alguna.

Por ejemplo, si se utiliza el vocabulario de los niveles de análisis de Marr como marco común sobre lo cual asentar un mínimo análisis conceptual, entonces dentro de la presentación del grupo PDP hay fuerte ambigüedad respecto a si se está trabajando a nivel de implementación o algorítmico, lo cual deja en un status incierto qué se entiende por “arquitectura”. Al principio de su obra, por ejemplo, se dice que “In our view, people are smarter than today's computers because the brain employs a basic computational architecture that is more suited to deal with a central aspect of the natural information processing tasks that people are so good at.” (Rumelhart & McClelland, 1986, 3), lo que puede ser tomado como relacionando arquitectura con el nivel de *implementación*, dado que apelan al *hardware*. No obstante, posteriormente se dirá que “We are investigating an architecture in which cooperative computation and parallelism is natural. (...) The real point is that we seek algorithms that are as parallel as possible. We believe that such algorithms are going to be closer in form to the algorithms which could be employed by the hardware of the brain” (op. cit., 113) pudiendo ser interpretado que la “arquitectura” es algo que se extiende a nivel algorítmico.

Por otra parte, desde Smolensky será relativamente más simple extraer una noción de “arquitectura cognitiva” al delinear un TF subsimbolista que dará luces respecto a cómo tratar el CO conexionista. Para Smolensky los “levels of cognition should not be thought of by analogy to levels of computer systems” (Smolensky, 1987d, 100), por lo tanto utilizar el marco de los niveles de análisis de Marr no sería una jugada lícita. Luego, lo que se entiende por “arquitectura cognitiva” sería equivalente a la especificación de cualquier modelo conexionista, caracterizado por moverse en cierto nivel subconceptual y cómo su funcionamiento permitiría la *emergencia* de “lo cognitivo” (Smolensky, 1988a).

Teniendo esto a la vista, entonces lo que falta es qué entienden los simbólicos cognitivistas. F&P dirán:

Put differently, the architecture of the cognitive system consists of the set of basic operations, resources, functions, principles, etc. (generally the sorts of properties that would be described in a “user’s manual” for that architecture if it were available on a computer), whose domain and range are the *representational states* of the organism” (Fodor & Pylyshyn, 1988, 10).

Es decir, primero, es claro que hay una dependencia del uso de los niveles de análisis de Marr – i.e. la analogía computacional lo permea todo en el simbolismo cognitivista – y, segundo, no hay una “emergencia de lo cognitivo” similar a la conexionista, ya que los estados representacionales de cierto organismo son, tanto dominio, como rango de las operaciones que se hacen sobre éstos. Luego, la noción de “arquitectura cognitiva” cognitivista será aquella organización que cumpla tales rasgos básicos de semejanza con un sistema computacional.

Uno puede extenderse un poco más respecto al tema, y mostrar que la “forma definitiva” que *deben tomar* cada una de las arquitecturas cognitivas por cada lado son totalmente diferentes; para Smolensky – y para los conexionistas del grupo PDP – la arquitectura cognitiva *debe parecerse al cerebro*. Sin embargo, para F&P tal cuestión no es necesaria, y asume que la arquitectura *debe parecerse a un computador*. Los argumentos para esto último no dejan de ser persuasivos, ya que F&P apuntan al hecho obvio que las teorías de nivel superior no necesitan parecerse en su formulación al nivel inferior que sobre el cual se implementa el fenómeno tratan de explicar, solamente deben parecerse en el caso que se muevan en el mismo nivel (Fodor & Pylyshyn, 1988, 10).

Puede ser que este análisis esté equivocado y que, aunque no sea iguales las nociones de “arquitectura cognitiva”, basta como común denominador para la conmensurabilidad el que sean conexionistas y simbolistas *realistas representacionales* o que usen el mismo vocabulario de niveles de análisis. Lo segundo puede ser desechado de inmediato, ya que, como se señaló poco antes, el uso de tal vocabulario por parte del grupo PDP *es confuso*, y por parte de Smolensky *es nulo*. En efecto, uno puede extenderse en tal tema y mostrar cómo autores neurocomputacionales contemporáneos consideran que el aporte de Marr puede verse como avanzando por el lado neurocomputacional, sucesor claro del conexionismo, y no del lado simbólico. Por ejemplo, Terrence Sejnowski dirá sobre la obra clásica de Marr (2010 [1982]):

A remarkable feature of Marr's book is the degree to which biological considerations enter on almost every page in inspiring computational analysis, in choosing between algorithms and in providing the ultimate measure of success. Computational explanations for our visual and mental abilities eventually may be found, and seeking such explanations is essential-this was Marr's message. *However, he was far from abandoning biological and psychological data in reaching this goal.* (Sejnowski, 1991, 300)

La cuestión del realismo representacional vale la pena considerarlo con algo de detención: las representaciones existen para conexionismo y simbolismo, ambos bandos lo declaran sin pudor, pero dentro de sus IC ocupan diferentes posiciones. Por una parte, para el simbolista la representación es parte del SO, mientras que para el conexionista es parte del SON, o dicho de manera más tradicional a la filosofía de las ciencias – pero menos útil desde mi perspectiva – para el simbolista cognitivista es parte del *explanans*, mientras que para el conexionista subsimbólico es parte del *explanandum*. Ello ya debería bastar para generar duda de la posibilidad de equivalencia, pero creo que la cuestión decisiva está en considerar el formato mismo de las representaciones; por una parte el CE para representaciones simbólicas se limita a considerarlos discretamente dentro de algún lenguaje informático, mientras que el CE conexionista es matricial y vectorial. De modo extemporal puede aducirse que tal diferencia es superficial y sugerirse el uso de una estrategia como la hecha en (2.4.2.), sin embargo tal uso se justificaba por la existencia *previa* de un CO equivalente y, más profundamente (o “abajo” cerca del F), por *características abstraídas y supuestas equivalentes*. En este caso, dado que estoy analizando una polémica bastante circunscrita temporalmente, entonces *me limito a constatar que no hay elementos relevantes para una equivalencia entre SO entre conexionismo y simbolismo*.

Creo que lo anterior es suficiente para mostrar que el análisis tradicional debiese ser desandado y se precisa buscar un medio conceptual alternativo para dar cuenta de la relación entre simbolismo y conexionismo. Pues bien, las nociones de “simbolismo” y “conexionismo” han sido apuntadas al final de los apartados (2.5.ii.) y (2.5.iii.), respectivamente, y se definen como una conjunción de TF y CO; “simbolismo cognitivista” vs. “conexionismo subsimbolista”. Aplicando las definiciones que componen la tesis de los IC, entonces si cada uno es trasladado a una estructura tipo-IC, más allá de los problemas sobre la posibilidad de dar con un CE unívoco, ambos serán IC abstractos por el hecho de carecer de SON, siendo el primero un IC abstracto incompleto y el segundo un IC abstracto generalizado (Def.7.3. y 7.3.1.). Luego, si se les trata de conmensurar, entonces, siguiendo la Def.8.4. no existiría modo de hacerlo, ya que no cumplen los requisitos mínimos de ser IC concretos puntuales, de lo cual menos se puede esperar que presumiblemente puedan compartir un SON. Ambos IC están primordialmente constituidos por conjuntos componentes demasiado “altos”, contruidos sobre SO, CC y SON fragmentarios, lo cual impide aplicar la Def.8.5. sobre cómo lograr conmensurabilidad local.

A lo anterior se agrega algo más: “simbolismo cognitivista” y “conexionismo subsimbolista” no pueden ser tratados como “teorías”, sino que a sí mismos se consideran como englobando modelos y teorías, por tanto siendo “meta-teóricos” ambos por igual. Ello es importante, ya que entonces la relación interteórica entre tales IC abstractos constituiría una relación “inter-teórica” – nótese las comillas – o *inter-metateórica*. Esto puede ser aplicado inclusive sin la adhesión de los TF respectivos.

El anterior análisis parece demasiado simplista y pareciera que falta algo, pero no ha faltado nada realmente. Se sigue totalmente si se acepta el desarrollo que he llevado a cabo hasta acá. Pueden agregarse o quitarse elementos que no han sido tratados en profundidad, como por ejemplo, el funcionalismo – postura filosófica de la mente – en mayor o menor grado, el empirismo versus el innatismo, la relación con la “teoría” de la psicología popular, etc. no obstante, tales elementos no hacen mella a la conclusión anterior, a saber, que *simbolismo cognitivista y conexionismo subsimbolista no pueden ser conmensurados localmente, por tanto no entran en competencia directa*.

¿Existe otra clase de competencia que no sea basada en una conmensurabilidad local – como fue ejemplificada con el perceptrón y la neurona de M&P? Dentro de las definiciones de la tesis de los IC he dejado abierta la posibilidad de conmensurabilidad abstracta o pseudo-conmensurabilidad (Def.8.8.) previendo una situación como la presente donde “chocan” dos pseudo IC. El grado de conmensurabilidad abstracta depende dos condiciones: que se compartan los conjuntos componentes no existentes (o que se compartan los existentes que es idéntico) y que se compartan los F.

La primera condición se cumple, es decir, tanto simbolismo cognitivista como conexionismo subsimbolista carecen de SON, de SO y de CE definidos, y con respecto a la segunda se produce un fenómeno complejísimo: un F es simplemente un “dominio empírico” del cual se *extraen* características y sobre el cual se *suponen* características (Def.1.2.), pero la realidad misma es tan compleja que aducir que basta que un F sea *intuitivamente* una “unidad compacta” para que toda pareja o grupo de IC que trata sobre tal F automáticamente trate entonces de lo mismo, parece demasiado cándido. Un mismo F puede descomponerse en muchas facetas, y estar metafísicamente compuesto de muchos otros F, los cuales requerirán un tratamiento sistemático por separado. Luego, la única manera de argumentar a

favor de que dos IC abstractos compartan F es ver cómo, si son “completados” o “saturados” cada uno por separado por un SO y SON, pueden comportarse “como si” existiese competencia directa.

Para lo anterior considérese el caso de F = “lenguaje”, restringiéndose a una contrastación apresurada entre el modelo conexionista de inflexión verbal pretérita (McClelland & Rumelhart, 1986, cap. 18) versus el modelo SPA – *symbolic pattern associator* – (Ling, 1994), teniendo el segundo la meta manifiesta de diseñar un sistema simbólico que superase el desempeño del modelo conexionista para mostrar las ventajas de asumir “rasgos psicológicas realistas”, como por ejemplo la existencia de reglas. En particular SPA se dice que es simbólico porque: primero, la representación del input/output del programa aprendiz es un conjunto de símbolos fonema los cuales son los elementos básicos que gobiernan la inflexión pretérita, en particular se utiliza el sistema de representación de fonemas UNIBET. Segundo, el programa aprendiz opera en tales símbolos directamente, y el conocimiento adquirido se representa en la forma de reglas de producción utilizando estos símbolos fonema también. Tercero, tales reglas de producción al nivel fonológico pueden ser fácilmente generalizables a reglas de primer orden más abstractas que usan categorías de más altos nivel o más abstractas, tales como morfemas y raíces verbales (Ling, 1994, 214).

En contraste, los modelos conexionista operan sobre un sistema representacional distribuido, es decir vectores de rasgos fonéticos. Particularmente, Rumelhart y McClelland utilizan un esquema de codificación que categoriza cada fonema en cuatro dimensiones; la primera dimensión divide los fonemas en tres tipos (consonantes, consonantes continuas y vocales), la segunda subdivide tales tipos, la tercera clasifica los fonemas en tres lugares generales de articulación (frente, media y trasera) y la cuarta subcategoriza consonantes (sonoras vs mudas) y vocales (cortas vs largas) (McClelland & Rumelhart, 1986, 235). Este esquema es una compactación del sistema de *wickelphones* que transforma tales unidades fonemas en *wickelfeatures* (op. cit., 233-4). Luego, si uno compara las características abstraídas de F conformado el SO de cada uno, se topará con que, a pesar de tratar de caracterizar ambos modelos fonemas, son diferentes sistemas de representación, cuya diferencia no es trivial como sería el caso de dos sistemas numéricos de diferente base, pero basados ambos en notación posicional, sino que es bastante relevante al deberse a la presión de factores del resto de estructura del IC y el SO.

Tales “factores del resto de estructura del IC y el SO” son, por ejemplo, las características supuestas del subsistema cognitivo de generación de inflexiones verbales, donde el SPA utiliza reglas de producción, mientras que el modelo conexionista de Rumelhart y McClelland utiliza el procedimiento de convergencia del perceptrón (op. cit., 225-6). Por tanto, y tratando de hasta aquí evitar muchos detalles, cuando se comparan SPA y el modelo PDP conexionista de inflexión es fácil ver que cumplen una comensurabilidad mínima o virtual al, primero, ser trasladables a estructuras tipo IC con un SON bien delimitado y, segundo, al compartir solamente el F, pero no el SO, siendo tal F = “lenguaje/inflexión verbal/pretérito”.

No obstante, de esta situación particular no se sigue necesariamente la imposibilidad de comensurabilidad local (aunque quizás no sobrepasando en gran grado la comensurabilidad mínima) entre IC concretos puntuales conexionista y simbolistas. Si se contrasta el mismo SPA con la ANN estilo-NETtalk de Bullinaria (1994), entonces se observa en este caso el uso común del sistema de

representación de fonemas UNIBET. Luego, existiría SO común y posibilidad de conmensurabilidad no mínima¹²⁹. Aun así, sin certeza de si, o bien hay conmensurabilidad mínima, o bien conmensurabilidad local para todos los casos de comparación de cualquier par dado de “modelo conexionista aplicado vs modelo simbólico aplicado”, entonces no hay razones para suponer que el F es común en IC concretos puntuales más allá de una caracterización intuitiva del SO. De lo anterior se concluye que *para IC abstractos el “F intuitivamente común” tiene un estatuto de existencia bastante dudoso, ya que hay casos, como el mencionado primero (Rumelhart-McClelland vs SPA), donde los ca y cs no son equivalentes entre los respectivos SO, lo que implica solo una noción intuitiva compartida de F.*

Por lo tanto, esta situación de ambivalencia respecto que a veces existe conmensurabilidad local y otras veces solamente virtual entre dos IC concretos puntuales conexionista versus simbólico, es lo que crea la *generalización apresurada* de suponer que, ya que se cumple *a veces* que se comparten F entre el simbolismo cognitivista versus el conexionismo subsimbólico, entonces al ser ambos IC abstractos con los mismos conjuntos componentes inexistentes, cumplen la Def.8.8. siendo así conmensurables abstractamente. Esto sería lo que daría la *apariencia de competencia*, que finalmente eso es lo que es lo que implica la noción de “conmensurabilidad abstracta”. Empero, *el caso es que las condiciones mínimas 8.8.1. y 8.8.2. no se cumplen de manera rigurosa en conjunto, por tanto, recurriendo a la Def.8.9. puede verificarse que simbolismo cognitivista y conexionismo subsimbólico cumplen todas las condiciones (8.9.1., 8.9.2. y 8.9.3.) para ser considerados inconmensurables en todas las formas concebibles. Queda así constatado que ambos IC no compiten y cumplen ambivalentemente la apariencia de competencia.*

La noción de “fenómeno” dejada al arbitrio de la intuición, aunque puede ser una salida seductora por su sencillez, no resiste un análisis profundo. Aun aplicando de manera general la tesis de los IC – la cual en sí misma es aproximativa y muy amplia al punto de tratar de subsumir en sí absolutamente todos los instrumentos empíricos contemporáneos –, es imposible dar sentido a qué se entiende por tal noción, inclusive en casos tan sencillos como el tiempo pretérito, *sino es antes mediante la intervención de algún CE que permita aseverar equivalencia*. Se puede presumir equivalencia, sin embargo la tesis de los IC deja la presunción como mera condición anterior a la posible fecundidad de responder la pregunta, o sea tal presunción no implica una respuesta de ningún tipo. Cuando se traslada un modelo a una estructura IC, una de las preguntas-heurísticas básicas que guía la resolución de tal tarea es cómo el F pasa a ser SO, ergo qué consta como característica abstraída y, acto seguido a tal pregunta, es cómo tal característica se representa (por ende, de inmediato se pregunta por el CE). No hay contenido sin formato, pero para efectos del análisis mediante IC contenido y formato son elementos de conjuntos disjuntos.

Para reforzar lo anterior, y recurriendo a una analogía ya utilizada por los autores vistos (Smolensky y McLaughlin), puede decirse que química y física están en una relación inter-metateórica de competencia, ya que comparten la explicación sobre el mismo fenómeno, un fenómeno que podría ser definido algo así como “el cambio de la materia”. No obstante, en ambos casos lo que se entiende por “cambio” y “materia” varía tan enormemente que no hay intersección posible *grosso modo* de F. Si se pasa a IC concretos puntuales de ambas disciplinas, es decir, se pasa a IC con SON bien delimitado dentro de la

¹²⁹ Es necesario decir que SPA puede ser considerado ya algo “híbrido”, al estar emparentado a las ANN al utilizar el algoritmo C4.5, el cual es una mejora del ID3, perteneciente éste la clase de sistemas con aprendizaje supervisado. Recuérdese que una ANN con retropropagación es un caso de tal clase de sistemas (Ling & Marinov 1993, 248-9)

gran cantidad de subdisciplinas que componen cada disciplina, entonces recién podría generarse posibilidad de conmensurabilidad virtual entre IC concretos puntuales de química y física. En consecuencia, la relación conexionismo subsimbolista y simbolismo cognitivista no es ni la de mecánica cuántica y química, ni de mecánica cuántica y mecánica clásica. *La relación entre ambos es de física y química o viceversa, es decir están hablando de lo mismo sólo de un modo superficial, intersectándose solamente en casos específicos que no permiten generalizar desde la perspectiva de los IC.*

Esta conclusión, al menos para mí, era totalmente insospechada, ya que esperaba que, al construir un sistema conceptual relativamente más preciso que permita entender cómo se dan las relaciones interteóricas, y que fuese actualizado a las nociones menos normativas y más descriptivas de la filosofía de las ciencias contemporánea, podría encontrar dónde exactamente se da la conmensurabilidad entre ambas escuelas. Pero, sacando componentes que son totalmente imprecisos para la práctica propia del modelamiento, tales como los TF, y moviéndose estrictamente en el nivel de las discusiones filosóficas donde poco se entra al detalle de modelos y su conexión con los fenómenos, entonces no queda nada.

Con todo lo extraído, por consiguiente, es posible contestar a la disyuntiva que titula este apartado y en general la tesis que ha sido postulada en la introducción a este trabajo investigativo, a saber si es posible establecer colaboración entre conexionismo y simbolismo, o al contrario, es siempre el caso que hay competencia. Circunscribiéndome al conexionismo subsimbólico y al simbolismo cognitivista no hay ni una cosa ni la otra, ya que las categorías de colaboración y competencia han sido reservadas para casos de IC concretos; hay discusión filosófica, debate en general, pero en términos que son totalmente irrelevantes respecto al F, dado que – sin olvidar que me restrinjo a lo expuesto en la sección (1.1.) y sus subsecciones – lo que se discute son aspectos demasiado “superiores” de la estructura de IC, *olvidando que tales conjuntos componentes “más arriba” tienen sentido única y exclusivamente en conexión estrecha con el contenido de los conjuntos “inferiores”.*

La competencia es posible observarla cuando se delimita SON y se cumple la Def.8.5. es decir, por ejemplo, en el caso de SPA (Ling, 1994) versus la ANN estilo-NETtalk de Bullinaria (1994) o, el caso en el que he tratado de ser más exhaustivo, en el ejemplo dado en (2.4.2.). Pero, curiosamente, también en esta misma dimensión de IC concretos puntuales es posible encontrar colaboración como he ejemplificado a lo largo de la tesis con los modelos híbridos en diferentes grados. Por lo tanto, *no hay competencia ni colaboración a priori*, sino que depende de la relación que vayan estableciendo, contingentemente, entre sí los investigadores. Filosóficamente, uno puede construir argumentos para decir que a priori hay, o bien competencia, o bien colaboración, pero ello será impuesto desde el exterior de la práctica científica, la cual funciona compitiendo o colaborando a partir de, conjeturo, *motivaciones que no necesariamente derivan directamente de F.* Lo que he logrado recabar desde lo histórico de las ciencias cognitivas es que, si nuevamente se flexibilizan y se dejan al arbitrio de la intuición los términos de “simbolismo” y “conexionismo”, entonces al comienzo de su relación a fines de los 50’ y principios de los 60’ es posible establecer que había colaboración, precisamente, porque las funciones cognitivas estaban “bien repartidas”, es decir, a las redes neuronales les tocaba la percepción y al simbolismo tareas de alto nivel. Pero, a partir de fines de los 60’ y el trascurso de los 70’ se rompió la calma y ya no se repartían, sino que “competían”. La tesis de los IC aplicada mostraría que tal competencia no era homogénea siendo, en el peor de los casos, la mínima posible y en el mejor de los casos, siendo local.

– (2.5.v.) – Posibles objeciones a la conclusión anterior y sus respuestas

Una manera eficaz de rechazar la conclusión central de (2.5.iv.), a saber, la *incommensurabilidad – en todos los sentidos – del simbolismo cognitivista y el conexionismo subsimbolista*, es desestimar el entramado de supuestos que permitió la construcción de la tesis de los IC, los cuales han sido delineados desde las digresiones 1 y 2 en la primera parte de la tesis hasta la sección (2.4.). No obstante, considero que para desestimar tal armazón conceptual sería preciso ir en contra de algo sobre lo cual he sido explícito, a saber “dar por hecho que *la filosofía de las ciencias es progresiva*” (sección (2.3.2.)), lo cual conllevaría retomar viejas cuestiones epistemológicas que, desde mi perspectiva, no tienen cabida en el quehacer filosófico contemporáneo.

Luego, supóngase que se acepta la tesis de los IC como medio de análisis de las relaciones interteóricas e inter-metateóricas en general y de la polémica entre simbolismo y conexionismo en particular, es decir supóngase que ha sido aceptado todo lo desarrollado hasta la sección (2.4.2.). Entonces, considero que las siguientes son algunas objeciones previsibles:

Objeción (i): “Desechar el fenómeno de ‘arquitectura cognitiva’ – estrategia utilizada al principio del apartado (2.5.iv.) – como terreno común y definitorio de la competencia entre simbolismo y conexionismo no está bien fundamentado, ya que descansa en suponer que no se puede encontrar una definición neutral de ‘arquitectura cognitiva’ más allá de Fodor y Smolensky. Tal definición neutral existe y a partir de ella es posible decir que hay commensurabilidad y, por extensión, competencia”.

R. Esta objeción tiene una base que apela a décadas de acuerdo implícito entre investigadores de ciencias cognitivas. De otro modo puesto, asume como base que existe una definición neutral que sería algo así como: ‘una arquitectura cognitiva es *la suma ciertos elementos subcognitivos y cómo tales elementos subcognitivos interactúan para generar un proceso cognitivo*’¹³⁰, por tanto lo que distingue a conexionistas y simbolistas es meramente los elementos subcognitivos que se elijan (representaciones discretas versus distribuidas) y la interacción entre éstos (reglas versus operaciones vectoriales).

Creo que, más allá de si es posible lograr una total neutralidad objetivamente, el lenguaje es engañoso y el hecho que algo pueda plantearse como superficialmente dando un “terreno común y definitorio” a la competencia entre dos “paradigmas” no significa que existan tales entidades comunes en el fondo. La definición neutral de arquitectura cognitiva que ofrezco como ejemplo podrá tener cierta utilidad introductoria al campo de las ciencias cognitivas, pero no es informativa para explicar rigurosamente qué sucede en casos concretos de funciones cognitivas explicadas desde modelos diferentes. La tesis de los IC constriñe la existencia de commensurabilidad a los casos de IC concretos puntuales, por lo tanto, si se quisiese pasar desde “definiciones neutrales” a explicar cómo compiten los casos concretos, entonces se llegaría a una versión de competencia exactamente como la que contiene la tesis de los IC, es decir, se pasaría a revisar cómo se abstraen y suponen características desde un F para construir un SO, como tal SO es confirmado, etc.

¹³⁰ Me he inspirado de John Anderson (2007, cap. 1) y Andrew Brooks (2009) para condensar esta noción “neutral” de arquitectura cognitiva.

Objeción (ii): “Supongamos que es correcta la visión pluralista de las ciencias cognitivas en (2.5.i.), aun así es posible conmensurar y mostrar la competencia entre simbolismo y conexionismo, sin agregar las adendas de ‘cognitivista’ y ‘subsimbolista’ respectivamente, mediante el contraste de los modelos-hito que agruparían en torno a sí aquello que podría ser denominado ‘paradigma’ simbólico o conexionista”.

Esto es parcialmente cierto, pero no lleva a la conclusión que desea llegar el crítico. Si se toma, por ejemplo, al GPS como “modelo hito” versus cualquier otro dentro de los presentados en los volúmenes PDP se llegará a que, en el mejor de los casos, existe conmensurabilidad virtual: por una parte el SON del GPS está bastante bien delimitado en la resolución de problemas y utiliza protocolos verbales, metodología que es ajena a lo que es posible encontrar en los volúmenes PDP. Por otra parte, el fenómeno típico en el cual se concentraron los conexionistas es la percepción, siendo el lenguaje la excepción, estando tales cuestiones fuera de las preocupaciones primarias de los modelos-hito entre los 60’ y 70’ para los simbolistas.

Es posible expandir los SON de cada lado y buscar alguno donde confluyan los F. No obstante, dado que conexionismo y simbolismo tienen CO diferentes, ergo tienen formas de construir SO diferentes, entonces lo máximo que concedo utilizando “modelos-hito” es que se puede dar cierta forma de conmensurabilidad mínima o virtual, la cual en sí misma, sin una precisa *formulación de criterios adicionales de decisión*, no nos permite definir qué sucederá más allá de tal tipo de competencia no directa. La tesis de los IC ha sido explícitamente construida sobre un trasfondo pluralista, por consiguiente lo más neutral posible respecto a establecer criterios que puedan ser considerados normativos y/o monistas; esto último ha sido recalcado al final de la sección (2.4.2.) donde advierto que, a pesar que desde la tesis de los IC se muestra cómo un IC “expande” a otro, esto por sí solo no debe ser considerado como evidencia de que el que expande es mejor que el más restringido, en realidad puede ser el caso que, por ejemplo M&P hayan estado conformes con su propio modelo y lo considerasen definitivo para sus propias metas.

La subdeterminación de dos IC con conmensurabilidad virtual y la falta de criterios de decisión adicionales en la tesis de los IC es un problema para el monista explicativo que, sin quizás ser consciente de ello, espera que la competencia entre dos teorías sea un hecho siempre transparente y apegado a cierta “racionalidad” que dirima entre ambas desembocando en alguna teoría definitiva heredera de las teorías “claramente” ganadoras. El pluralismo explicativo como trasfondo es agnóstico al respecto y, lo máximo que puede ser extraído de éste, es que la misma dinámica de los IC sería lo único que funcionaría como tamiz de para dirimir entre aquellos IC que perduran y los que no.

Objeción (iii): “El simbolismo [o el conexionismo] está mal definido, ya que es un IC abstracto que al ir tomando diferentes SON es concreto caso a caso. Si se define de tal manera, entonces simbolismo y conexionismo sí pueden ser conmensurados, de lo que se seguiría que es posible dirimir *directamente* cuál de ambos es mejor.”

R. El análisis mediante el cual afirmo la inconmensurabilidad de ambos se da en un contexto bastante restringido, del conexionismo subsimbolista versus el simbolismo cognitivista, sin embargo considero que se mantiene aun variando la definición de conexionismo y simbolismo. La objeción supone dos pasos: primero, que es posible construir una definición más allá de este contexto restringido que permita

subsumir de manera clara un conjunto de modelos como “simbólicos” versus otro grupo como “conexionista” y, segundo, que tal definición es suficiente para una posterior conmensurabilidad.

Concedo que el primer paso es posible de concretar, sin embargo tal tarea implicaría tomar una postura arbitrariamente fundada y taxativa sobre casos límites como las redes semánticas. Por ejemplo, en caso que se asuma que es la representación local versus la distribuida la única característica que diferencia a “conexionismo” de “simbolismo”¹³¹, implicaría reevaluar hasta qué punto clásicos conexionistas del grupo PDP lo son realmente si utilizan estrategias localistas en su construcción. Una definición de “conexionismo” que deje fuera precisamente algún caso emblemático conexionista me parece una definición innecesariamente restrictiva.

No obstante, el problema se encuentra en el segundo paso, el cual es imposible de llevar a cabo, ya que, si podemos estar seguros de algo, es que simbolismo y conexionismo, como se quisieran definir, tienen CE bastante diferentes, los cuales se manifiestan en sus CO divergentes. Esto significa que es imposible trasladar simbolismo y conexionismo a IC que permitan SO que puedan ser comparados como se hizo con el caso de la neurona de M&P y el perceptrón, y que, cuando es posible encontrar puntos comunes en el SO, lo que da paso a conmensurabilidades locales entre IC concreto conexionista versus simbólico, entonces *tales casos simplemente no son suficientes para generalizar al resto de todos los modelos*.

Objeción (iv): “Simbolismo y conexionismo no pueden conmensurarse más allá de la definición que se les quiere dar, luego es verdadero que no pueden competir directamente. Sin embargo, si se aplica la misma estrategia que se utilizó para conmensurar entre la neurona de M&P y el perceptrón de Rosenblatt, es decir, se busca un CE desde el cual se homogenicen los SO de modo de determinar cuál es mejor respecto a cierto SON, entonces pueden ser conmensurados a partir de un marco que los *trascienda*.”

R. Esto es correcto y depende crucialmente que se encuentre un CE que los homogenice, pero – conjeturo – tal estrategia, más que mostrar cómo se da la competencia, revelaría justamente lo que estoy concluyendo aquí, es decir, mostraría más evidentemente el cómo es que no están ni en competencia ni en cooperación. En cualquier caso, esta última objeción supone una cantidad de *nuevos supuestos* sobre los cuales existe la posibilidad de que se demuestre conmensurabilidad, pero ello significa que entonces tal objeción aún falta ser desarrollada hasta sus últimas consecuencias como para que sea realmente un obstáculo del que deba preocuparme.

– (2.5.vi.) – Re-comprensión y re-organización de la polémica simbolismo-conexionismo

La tesis de los IC permite extender y detallar el campo sobre el cual se desenvuelven simbolismo cognitivista y conexionismo subsimbolista. De este cambio de perspectiva, el cual toma en cuenta la complejidad del F, se sigue que la relectura de la polémica, o sea, su re-comprensión y re-organización, es una reacción necesaria a una filosofía de las ciencias monista que ha dominado el análisis de las ciencias cognitivas. Es por ello que todos los detalles más famosos de la polémica pueden ser tratados

¹³¹ Siendo honesto, esta salida es la típica en ciencias cognitivas, quizás no con el énfasis de “local” versus “distribuido”, sino que cuando se exponen introductoramente las nociones de “representación discreta” versus “representación continua” (Dietrich, 2007, 13-4)

asépticamente (o al menos lo más posible¹³²) desde una perspectiva pluralista. A continuación retomaré, a la luz del análisis mediante IC, brevemente lo que dejé como reflexiones en la sección (1.3.1.), consistiendo tales reflexiones algunas de las aristas conceptuales más destacadas de la polémica simbolista-conexionista.

- (.01) (.09) (.18) Estas reflexiones se sintetizaban, conceptualmente, en cortar el nudo gordiano de la “competencia” (como se apuntó en (1.3.4.)) y, posteriormente, en dar nociones mejor entendidas de lo que se concebía por “conexionismo” y “simbolismo”. Luego, dentro del apartado (2.5.iv.) ya se ha dado respuesta a todas. En particular, por ejemplo, la noción de “modelos clásicos” que utiliza F&P puede mostrarse que es un IC abstracto al igual que la “metateoría” conexionista. Si ambos son, respectivamente, equivalentes al “simbolismo cognitivista” y “conexionismo subsimbolista” es una cuestión que puede ser adoptada, pero que no varía el análisis hecho antes, dado que los “rasgos arquitecturales” de F&P no cambian el hecho de que se tratan de IC abstractos. Ahora bien, dado que se ha llegado a que conexionismo subsimbolista y simbolismo cognitivista ni compiten ni colaboran, entonces se responde a (.13) que preguntaba sobre cuál salida era la adecuada. Lo que sucede cuando se circunscriben tales IC abstractos a un SON ya ha sido desarrollado sucintamente al final de (2.5.iv.)
- (.02) (.03) (.04) Se sintetizan en dos cuestiones: primero, si es implementacionista o no el conexionismo y, segundo, cuál es la relación del conexionismo con otras estrategias “bottom-up”. La primera cuestión tiene sentido para los conexionistas que adoptan el vocabulario de niveles de análisis de Marr, entre los cuales Smolensky no está como se apuntó en (2.5.iv.). En caso que un conexionista adoptase los niveles de análisis de Marr dentro de su investigación, entonces tampoco es muy claro cómo, automáticamente, pasa a ser implementacionista el trabajo de modelamiento cognitivo de tal conexionista, dado que tal conexionista puede declarar con tranquilidad que el uso de tales niveles es meramente heurístico (Def.4.2.) siendo éstos *no autónomos de los datos del nivel implementacional*. El implementacionalismo como postura tiene sentido si se acepta la autonomía que defienden F&P del nivel computacional. Respecto a la segunda cuestión, otras estrategias explicativas (otros IC) “bottom-up” han sido delineadas dentro del esquema en el apartado (2.5.i.), sin embargo la relación con el conexionismo más allá de los aspectos históricos que me parecen evidentes, es difícil de decir con mayor precisión sin entrar a estudiar concienzudamente tales otras estrategias teóricas. Ello queda pendiente, aunque lo que ha aportado este trabajo es un medio para llevar a cabo tal estudio de contraste.
- (.05) Se resume en cómo diferenciar el uso técnico de las ANN de su uso como instrumento para estudio científico de la cognición. La forma de diferenciar entre ambos *no se aclara mediante análisis por IC*. Un SON es “cualquier cosa sobre lo cual se evalúa el éxito del resto de conjunto de

¹³² Esta suerte de “agnosticismo” respecto al tratamiento de ciertos tópicos bastante filosóficos – considérese la relación entre psicología popular y conexionismo, o innatismo y conexionismo, etc. – que contiene la tesis de los IC, debo confesar, ha sido investigado en su posibilidad *ex profeso* inspirado en la opinión de Terrence Horgan sobre simbolismo cognitivista y conexionismo no subsimbolista: “Agnosticism about the various competing approaches is one epistemically reasonable stance, given the current evidential situation.” (Horgan, 1997, 28). Ambas posturas son consideradas por Horgan como “just-so story”, es decir “fábulas” sobre cómo funciona la mente, dado que su soporte en evidencia es bastante escaso, y cuando se trata de expandir y profundizar en el todo de las funciones cognitivas tienden a fallar.

componentes de IC” (Def.5.1.), por lo tanto una aplicación técnica, por ejemplo la predicción de cambios en el mercado de acciones mediante una ANN, puede constar con todos los elementos para estructurarse como IC. La justificación de la utilización de ciertas ANN por sobre otras en ciencias cognitivas, así como la cuestión del diferenciar el uso técnico del uso “científico”, está sujeta *criterios adicionales de decisión*, los cuales pueden pasar a ser una extensión de las definiciones dadas entre (2.3.2.1.) y (2.3.2.3.). No obstante, hay que advertir que, a la postre, tal adenda de criterios puede ir contra parte del entramado que me permitió elaborar la tesis de los IC, por ejemplo al tener que ser normativo-monistas, en vez de descriptivos/normativos-pluralistas.

- (.06) Pregunta por la atingencia del debate innatismo/empirismo y, en general, la legitimidad de la argumentación filosófica apriorística. La Def.4.2. permite separar aquello relevante para la explicación de aquello no relevante, e innatismo versus empirismo como posturas filosóficas no son precisas para la utilización de un IC, sino que meramente pueden llegar a servir como inspiración y no interfieren, finalmente, en el desempeño del sistema explicativo-predictivo construido. Por ejemplo, parte del TF de modelos simbólicos clásicos tenía que ver con cómo Newell y Simon veían al ser humano en un género común con los sistemas de procesamiento de información, pero ello no es necesario asumirlo para evaluar el grado de explicación-predicción que tiene un IC como el GPS. No hay un CE privilegiado para la explicación-predicción defienden los propios simbolistas clásicos (Newell & Simon, 1961, 2012-3), no es privilegiado ni el programa ni las leyes matemáticas como formato de explicación, por lo tanto lo único que queda para evaluar son criterios adicionales que no han sido tocados por la presente tesis, ya que tales criterios adicionales tendrán inmediatamente un carácter normativo. Si se lleva hasta sus últimas consecuencias lo anterior (en general toda la tesis de los IC), entonces las discusiones filosóficas debiesen ser “aterrizadas” al nivel de los conjuntos componentes más cercanos al F para volver a ganar su legitimidad.
- (.07) (.08) (.14) Se preguntaban respecto a cómo, por ejemplo, “representación” u otras nociones terminaban en sectores contrapuestos en el *explanans* versus el *explanandum*, respectivamente en simbolismo y conexionismo, siendo que partían de las mismas observaciones de la conducta, además de cómo superar la “imagen simplificada” de la competencia. Esta reflexión está mal guiada desde el comienzo, ya que supone que las características abstraídas y supuestas son comunes a ambos bandos, lo cual es incorrecto. El fenómeno de quedar nociones como “representación” y otras en sectores de la explicación separados, precisamente se debe a que los “SO típicos” se construyen de maneras distintas. Por otra parte, se agregan divergencias en los conjuntos componentes superiores. Todo esto conlleva que las nociones que parecen a primera vista similares – como es el caso de usar los “mismos” niveles de análisis de Marr – no lo sean si se trasladan modelos de cada lado a estructuras tipo IC. La “imagen simplificada” es superada precisamente cuando se dan definiciones bien delimitadas de “conexionismo” y “simbolismo”, las cuales dependen de un análisis mediante IC, el cual obliga a lo que se ha hecho a lo largo de la tesis, es decir, a una reconstrucción histórica que ha tratado de ser más o menos exhaustiva.
- (.10)(.16) Se sintetizan en el problema para encajar los modelos híbridos y la “imagen simplificada” de la relación conexionismo-simbolismo. Pues bien, los modelos híbridos son un problema exclusivamente en caso que previamente se suponga la existencia de rasgos unívocos para individualizar conexionismo y simbolismo, es decir, en tanto se mantenga la imagen simplificada. Sin tal imagen, entonces la tesis de los IC, en conjunción con la pluralidad de las ciencias cognitivas,

transforman lo híbrido en la norma en el modelamiento cognitivo, dejando a los modelos más “puros” (conexionistas o simbolistas) como casos interesantes para su estudio histórico y como útiles para su investigación con fines pedagógicos, pero en ningún caso como representantes de cómo se da la *construcción real* de modelos para la explicación de lo cognitivo. Tal fenómeno real sería lo que trataría de ser capturado mediante análisis con IC.

- (.12) Apuntaba a retomar problema mente-cuerpo. Este problema en particular no puede ser tratado por la tesis de los IC, dado que pertenecen a subdisciplinas separadas (aunque con lazos en común); filosofía de la mente versus filosofía de las ciencias. Esto se evidencia en que el concepto de “reduccionismo”, que puede ser re-definido a la luz del análisis mediante IC, solamente se mueve en el nivel entre teorías y modelos trasladados a estructuras tipo IC. La cuestión de si el concepto “mente” es o no saturado por un modelo o teoría específico es una discusión que escapa a la filosofía de las ciencias. Quizás puedan agregarse definiciones normativas monistas a la tesis de los IC que combinen estos campos, pero ello queda a futura investigación.
- Desde (.15) citaré parte de la pregunta misma, por su interés para la mayoría de los filósofos que se interesan en la polémica simbolismo-conexionismo: “¿existen “en realidad” relaciones de constitutividad como las descritas por Fodor, así como “sistematicidad”, “composicionalidad” o “productividad” de la cognición?”.

Uno de los conjuntos componentes de los IC es su CO, dentro del cual caben principios generales respecto al SO que permiten llegar al SON. Dentro de tal CO es posible situar cuestiones como, por ejemplo, el “principio de gravitación universal”. Este elemento no es mero TF, sin embargo depende de cuestiones que están “bajo” éste, tales como las mismas idealizaciones de los objetos reales y sus movimientos – su mecánica – a nivel del SO. Por lo tanto, cuando se pone en duda tal CO lo que se hace, muy *grosso modo*, es presentar un SO diferente sobre el mismo F. Sin embargo, puesto que tal CO mencionado está asentado sobre una idealización que subsume un hecho tan básico como que todas las cosas a nuestro alrededor caen, entonces poner en duda tal CO es, simultáneamente, tener que hacerse cargo de tal *hecho evidente a cualquier espectador*.

Considérese la “sistematicidad”. Por un lado, primariamente, es un rasgo de la cognición y, por otro lado, secundariamente, es del lenguaje, según lo que se desprende de F&P. Dentro de tal contexto, entonces la sistematicidad del pensamiento es un elemento del CO del IC abstracto “simbolismo cognitivista”. Ignorar, criticar o no aceptar tal elemento debiese ser justificado desde un SO diferente, pero sobre el SO simbolista cognitivista sólo hay algunos rasgos muy ambiguos y fragmentarios, ya que se trata de un IC abstracto y, al contrario que como ocurre en física que se construye sobre elementos observables, *cuando se trata de psicología humana un fenómeno cognitivo pasa a ser evidente indirectamente cuando forma un patrón conductual robusto mensurable por cualquier espectador*. Un muy buen caso es el presentado por Miller (1956), el cual permitió asentar ciertas bases distintivas de lo que vendría a ser el límite de nuestro desempeño en tareas de memoria a corto plazo, al punto que hoy en día *define la capacidad límite de tal subsistema*. Sobre esto último se puede estar a favor o en contra del SO construido para lidiar con los datos experimentales, por ejemplo, respecto a cómo tal SO está en un formato o CE de vocabulario informacional, pero no se puede simplemente ignorar este verdadero “cuello de botella” que existe en el procesamiento del estímulo en el ser humano.

Dentro de la crítica cognitivista se apela a que el lector acepte que la sistematicidad del pensamiento es el F existente tras el *supuesto* patrón robusto de sistematicidad lingüística. Sin embargo, como he apuntado en la primera parte – sección (1.2.2.)(B2) – existe gran cantidad de literatura crítica que, más allá de poner en entredicho el TF, CO o CE que adscribe y defiende de modo monista el simbolismo cognitivista, levanta dudas acerca de la robustez misma de tal sistematicidad del pensamiento. Y, en honor a la ecuanimidad intelectual, tal clase de crítica es totalmente legítima y bien fundada, ya que ni en F&P (1988), ni en F&M (1990), se pueden hallar rastros de datos empíricos sobre la robustez de la sistematicidad que sirvan como cimiento fundamental para la construcción de una crítica contundente.

La fuerza de la crítica de cognitivista deriva de factores que no están basados en algún hecho evidente, sino que derivan de contingencias propias de la dinámica de los IC, es decir, de una probable y muy relativa hegemonía preexistente del cognitivismo en el momento del florecimiento del conexionismo. Concretamente, la “sistematicidad del pensamiento”, tiene sentido y realidad substantiva dentro del pensamiento fodoriano, además de tener valor para aquel investigador que está convencido de antemano que tal sistematicidad es en sí misma un hecho evidente. Pero, más allá de eso – a menos que algún patrón conductual se haya establecido, en cuyo caso la discusión toca tierra y se construye sobre otros antecedentes –, no tiene sentido suponer que tal sistematicidad está “ahí, en la realidad”.

Este análisis particular, que se ha hecho mediante IC, puede extenderse a cada uno de los elementos fodorianos del simbolismo cognitivista y, casi con total certeza, se llegará a una conclusión igual que a la que se ha llegado con la sistematicidad del pensamiento.

- (.17) Buscaba una diferencia “esencial” entre la crítica de Minsky y Papert (1969) y la crítica de cognitivista, además de plantear la pregunta sobre la necesidad de revisar la historia estándar del conexionismo. La diferencia puede deducirse de la misma aplicación de análisis mediante IC, es decir, la primera crítica está basada en un IC concreto cerrado, se hace restringiendo bastante de lo que se habla, siendo *Perceptrons* una exposición de las falencias de un grupo bien constreñido de modelos. La crítica de F&P está basada en un IC abstracto incompleto, al punto que muchos de los requisitos que, por ejemplo, Smolensky intenta ir satisfaciendo en sus réplicas a F&P, se denuncian como no cumplidos por F&M, siendo aquello justificado en cuestiones que corresponden al TF del IC abstracto que es defendido por Fodor y otros cognitivistas, en particular, derivado de posturas sobre la *causalidad* que *debiese* poseer el nivel cognitivo.

La segunda cuestión de (.17) apenas ha sido esbozada en el contexto de la sección (2.5.i.) y demanda un trabajo aparte, dado que lo más básico que requiere para empezar a ser contestado es construir una definición de “historia estándar conexionista”. No obstante, algo de lo que se puede tener certeza es que, por ejemplo, el impacto de la obra *Perceptrons* en la historia de las ANN en general ha sido sobredimensionada.

Conclusiones

3.1. Ventajas y limitaciones e inconsistencias de la tesis de los IC

(i) Ventajas;

- **Mejora de “resolución” frente a otros análisis filosóficos de las ciencias:** La filosofía cuenta fábulas, verosímiles y, a veces, útiles. La filosofía de las ciencias y de las ciencias cognitivas también cuenta fábulas, verosímiles y, quizás intentándolo un poco más, útiles. Pero no dejan de ser cuentos, formas de acercarse a algo que no está siendo tratado de forma directa, sino que muchas veces tratando sobre lo que otros hacen o dicen respecto a algo más, en este caso, respecto a la ciencia cognitiva y el fenómeno de lo mental. Mi fábula no persigue ser meramente verosímil, sino que persigue indicar que lo que se ha hecho hasta ahora se sostiene sobre cuentos que, cuando se busca su base, se derrumban cuando se trata de ir a los detalles más concretos sobre lo que se dice. En la jerga desarrollada hasta aquí, la filosofía de las ciencias cognitivas trata sobre IC concretos disímiles en sus detalles, los cuales, al ser eliminados, dan a lugar a una visión uniforme – IC abstractos de los dos modos – que cumple objetivos que no son realmente logrados, precisamente al perderse “resolución” suficiente para decir cosas significativas.

El debate simbolismo-conexionismo – al menos en la faceta que se desarrolló en la sección (1.1.) – no tiene el sentido que creen que tiene quienes entran en discusión. La tesis de los IC, precisamente, lo que intenta es volver a incorporar la “resolución” perdida en el íterin de la polémica. El conexionismo subsimbolista es un IC abstracto-generalizado, que al cargarse de un TF y CO superficialmente común al simbolismo cognitivista, genera la impresión de competencia respecto a tal IC abstracto-incompleto. Sin embargo, si se sigue las definiciones desarrolladas en el punto 8 dentro de la sección (2.3.2.2.) y, además, se considera válido el desarrollo de la sección (2.5.), conexionismo y simbolismo son inconmensurables, y su conexión se da un nivel que es irrelevante. En otras palabras, la discusión parte suponiendo que ambos bandos se están sosteniendo sobre el mismo F, pero no tenemos razones adecuadamente convincentes para hacerlo, a menos que *supusiéramos sin cuestionamiento* que el que ambos lados declaren en común, por ejemplo, el “realismo representacional” o “el uso de niveles algorítmicos” es suficiente para que se estén refiriendo al mismo F, SO o SON.

- **Presenta una imagen de cómo se separa formalismo de compromisos ontológicos:** La construcción de un SO directamente entra en ciertos compromisos ontológicos, los cuales, dependiendo la postura del investigador se transforman en privilegios ontológicos, contra los cuales uno puede establecer argumentos como los delineados en la sección (1.3.3.) y en (2.2.). La tesis de los IC de manera explícita diferencia entre TF y CO, por lo tanto *permite que dentro del análisis de una teoría o modelo exista el margen claro para diferenciar entre una “defensa epistemológica” versus una “defensa ontológica”*. Dado el pluralismo explicativo abrazado, entonces no hay a priori ninguna razón para suponer que ciertos compromisos ontológicos son mejores que otros, a menos que tales compromisos se vayan evaluando en el contexto de la dinámica de los IC, dentro de la cual existe la permanente posibilidad de eliminación o disgregación/subsunción por presiones que devienen desde el mismo estudio de cierto F y las relaciones con otros IC.

Si se aplica análisis mediante IC, entonces la separabilidad de TF del CO permite desechar el cognitivismo y no directamente las herramientas formales usadas en el modelamiento por el simbolismo, pero esta estrategia puede usarse también, por ejemplo, separando el formalismo dinamicista de la hipótesis de la cognición como sistema complejo o del uso de formalismo conexionista sin la necesidad de ser eliminativista. El “agnosticismo” es sólo un problema para el monista que busca un asiento metafísico definitivo, pero para la investigación en ciencias tal búsqueda de “lo definitivo” la torna dogmática y, a la larga, no científica.

- **Muestra de qué modo se integran la historicidad y conceptos científicos:** La tesis de los IC ha asimilado el historicismo hasta sus últimas consecuencias, al punto que asumo que la ciencia y su historia no son procesos separados – nótese ello en el paso desde las definiciones contenidas en la sección (2.3.2.1.) a aquellas contenidas en la sección (2.3.2.2.) –, sino que confluyen para dar a lugar a lo que posteriormente es interpretado y transmitido a los legos como un actividad que trasciende sus propias constricciones históricas, es decir, una ciencia como totalmente ajena de supuestos que deriven no directamente del mundo. Esta visión ingenua ha sido recién a mitad del siglo XX puesta en duda, pero ello trajo consigo cierto pesimismo respecto a la tarea del filósofo de las ciencias, quien ya no tendría su propio objeto de estudio, el cual se definía como labor no naturalizada. Creo que la situación debe verse optimistamente, es decir, hoy más que nunca es necesario abocarse al estudio directo de las prácticas científicas de modo de ir actualizándose respecto a los “tamices” que se están usando en las comunidades disciplinarias y que determinan que los IC puedan sobrepasar más allá de la fase de asentamiento. “Lo científico”, aquel abstracto que esperaban, tanto científicos, como filósofos de la ciencia encontrar mediante el análisis de sillón, simple y llanamente no existe más que en la imaginación de éstos y lo que es menester hacer es desarrollar herramientas conceptuales – como, quizás, he mostrado con la tesis de los IC – que funcionen como instrumentos perfectibles y multifacéticos para enfrentar la permanente interacción entre ideas, formulaciones y métodos.

Lo que puede ser extraído a partir de un IC – una teoría o modelo, etc. – no necesariamente se condecirá con un problema “real” o en cierto F, ya que el IC es solamente representación y nunca investigación directa de la realidad – de hecho no hay en sí misma algo así como “investigación directa de la realidad”, lo más que se acerca a tal cosa sería nuestro contacto directo con el mundo sin mediación teórica –, por lo tanto la solidez de lo que es extraído de un IC depende de que el propio IC ya haya atravesado las fases que le permitieron llegar a cierta madurez, la cual se define por la *naturalización exitosa de sus conceptos*. Esta historicidad ínsita en la noción de los IC permite ligar su validez a los efectos de los vaivenes de la competencia/colaboración continua, y es precisamente esta historicidad la que nos faculta, con cierta prudencia, asegurar que, por ejemplo, el mundo de la cotidianeidad que se desenvuelve ante nuestros ojos es “mejor capturado” por el mundo que nos es relatado por nuestras mejores teorías.

- **Permite transparentar y organizar debates:** el pluralismo explicativo desarrollado en la sección (2.2.) fue exclusivamente diseñado como un trasfondo útil para la posterior sistematización de definiciones que componen la noción de IC (sección (2.3.) y sus subsecciones). Por lo tanto, no implica la misma tesis de los IC un compromiso con una postura epistemológica de las ciencias pluralista definida, sino que es compatible con cualquier pluralismo de la gama presentada al final de (2.2.), e inclusive si se tiene una postura monista, tal postura puede ser aislada de los conjuntos

componentes más “abajo” de los IC, es decir, aquellos que lidian más directamente con cómo se abstrae y suponen características del F. Por consiguiente, el pluralismo en sí mismo permite no desestimar a priori ninguna estrategia explicativa, inclusive aquellas defendidas de manera monista, sino que, en principio, permitiría individuar (si es que existe) aquello que daría “privilegio” a cierto IC por sobre otro. El caso presentado de “expansión” de la neurona de M&P mediante el perceptrón es discutible para un innatista extremo, pero, justamente, *la ventaja de analizar mediante IC es que permite explicitar cuáles mapeos de elementos están siendo desestimados y cuáles están siendo tomados como centrales en cierto búsqueda de conmensurabilidad.*

(ii) Limitaciones;

- **Mejora la “resolución” de análisis anteriores, pero no lo suficiente:** Los análisis basados en visiones simplificadas como con el uso de “paradigmas” cruza muchos análisis filosóficos de las ciencias cognitivas. Si se utiliza la forma de análisis graficada que he bosquejado en la subsección (2.3.2.3.), el resultado sería algo así como lo que es el “esquema reducido de evolución de IC en ciencias cognitivas” en la sección (2.5.i.). No obstante, esta versión es realmente solamente un resultado muy preliminar de lo que involucraría llevar a cabo un proyecto historiador de las ciencias cognitivas que pusiera el énfasis en la permanente hibridación o, lo que es equivalente, la pluralidad de facto de las ciencias cognitivas. Me he permitido presentar este bosquejo incompleto en aras de dar forma a la intuición que subyace en el desarrollo de esta tesis, a saber, que “simbolismo cognitivista” y “conexionismo subsimbolista” son posiciones que aparecen tras un proceso un refinamiento y estrechamiento fuertísimo del foco investigativo, el cual es olvidado cuando se pasa a la discusión desde elementos superiores de los IC, pero que es esencial para el establecimiento de las equivalencias entre IC. Empero, la tesis de los IC parte con un planteamiento y un trasfondo que pertenece a la filosofía de las ciencias en general, siendo por ello mismo de “peor resolución” que un medio conceptual basado, estrictamente, en filosofías de las ciencias especiales. El análisis mediante IC está creado para ir transparentando preliminarmente una relación inter-teórica muy *grosso modo*, por lo tanto tiene un enfoque *mesoscópico* – ni *macroscópico* al no apelar a “paradigmas”, ni *microscópico* al no estructurarse desde alguna filosofía de ciencia particular.
- **No hay suficiente prolijidad en la explicitación y desarrollo de los principios filosóficos de construcción:** Los marcos lingüísticos de Carnap y el realismo interno de Putnam son un par de ejemplos de las influencias no declaradas que estructuración intelectual de la tesis de los IC. Ambas propuestas filosóficas permiten lidiar con, por ejemplo, la “existencia” de los objetos teóricos. Esta cuestión filosófica puede ser retomada dentro de la tesis de los IC, a partir de los cuales sería posible distinguir entre dos tipos de pregunta: una primera del tipo “¿existen las sillas?” versus una segunda del tipo “¿existe la gravedad?”. A nivel filosófico se tiende a dar por sentado que son la misma pregunta, en tanto ambas son “metafísicas”, pero de ambas solamente la segunda pregunta podría ser encajada dentro de un IC (recuérdese las definiciones 0. en la sección (2.3.2.1.)). Es concebible un mundo futuro donde para la ciencia la “gravedad” conceptualmente pase al desuso. Otro mundo posible sería aquel donde nunca se hubiese llegado a una teoría exitosa de la gravedad y se hubiese llegado (Kepler) a una exitosa de la *vis motrix*. En tal caso, sería parte del lenguaje común decir que las cosas caen por efecto de la *vis motrix* y no de la gravedad. Pero, con el primer

tipo de pregunta, no sucede lo mismo, ya que con el ejemplo de “silla”, parece intuitivamente menos probable que pase al desuso un término que hace referencia a un objeto que tiene una historia cultural muchísimo más antigua y amplia, y que trasciende un análisis desde la ontología de nuestras mejores teorías. Luego, la cotidianeidad la he dejado afuera, justamente para no cometer errores como los carnapianos, pero en general esta clase de reflexiones han sido purgadas de lo presentado en el presente trabajo. Existen grandes agujeros expositivos también para llegar a las digresiones 1 y 2 en la primera parte de la tesis, las cuales, a pesar de tener una conexión directa con algunos puntos centrales de la polémica conexionismo-simbolismo, se sostendrían muchísimo mejor sobre una investigación que tenga en cuenta filosofía de las ciencias formales y su relación con los hechos que son sistematizados en las ciencias empíricas.

- **No hay suficiente cantidad de aplicaciones:** mi medio conceptual es altamente especulativo, en tanto que ha sido inducido a partir de conocimientos superficiales, los cuales pueden ser resumidos en: algunos hitos iniciales en la historia de las ciencias cognitivas, conocimiento general de filosofía de las ciencias, conocimiento general de algunos hitos que son lugares comunes científicos y algunos intervalos del desarrollo de las ciencias cognitivas, en particular, el intervalo de la discusión simbolismo-conexionismo. Esta superficialidad es fácilmente visible al tomar un caso particular para la aplicación de la tesis de los IC – neurona M&P vs perceptrón – que fuera bastante evidente para ejemplificar un caso de conmensurabilidad local. Por tanto, no sería honesto rechazar la posibilidad de mejora y refutación, siempre que esta última esté basada en casos donde las aplicaciones de la tesis de los IC llevan a situaciones que no se condicen con lo que es el “F” real en este caso, es decir, la historia de las ciencias y la investigación científica en curso.

El medio conceptual desarrollado aquí pone bastante énfasis en la visualización de la relación interteórica abstracta, pero, a pesar de ser parte de la filosofía de las ciencias en general, ello no significa que lo sea en el sentido que se entiende usualmente el esfuerzo generalista de esta clase de filosofía. El sentido especial de lo hecho aquí lo da el énfasis en la historia de las ciencias que debe subyacer sobre la filosofía de las ciencias, aun cuando ésta tenga aspiraciones normativas. *Lo normativo debe ser compatible con la experiencia, con lo a posteriori, a pesar que sea esencialmente a priori. Luego, el trozo investigativo de filosofía de las ciencias desarrollado en la presente tesis, no está construido “sobre” la historia de la ciencia, en el sentido de trascenderla, sino que está construida “sobre la historia” en cuanto trata de ser compatible con ella.*

Esta observación llevadas hasta sus últimas consecuencias significa que lo hecho aquí está sometido a contrastación de la propia historia científica y es totalmente falsable por quien quiera darse el tiempo de buscar errores de adecuación entre la tesis de los IC y la historia, por ejemplo, en que la tesis de los IC plantee constructos que no tendrían contraparte en la actividad del modelamiento científico. Aun cuando reconozco que el componente hermenéutico es fuerte – por ejemplo, al acudir a contextos de descubrimiento, los cuales estarán difícilmente libres de los juicios tácitos de los historiadores de la ciencia que se encargan de recopilarlos y exponerlos –, esto creo que no es impedimento para “ver” intuitivamente cuando la tesis simplemente está siendo usada forzosamente y cuando se ajusta naturalmente a los hechos.

- **Demasiado flexible:** Dado que el trabajo hecho aquí, acabo de conceder, es susceptible de ser invalidado por al arbitrio del historiador experto – y por supuesto del filósofo de las ciencias –, entonces es posible que la propia tesis de los IC sea readaptada con un trasfondo totalmente

diferente, por ejemplo tomando como válidas no todas las premisas que constituyen el núcleo de la tesis de los IC. Por ejemplo, la Def.0.3. que supone la separabilidad de los conjuntos componentes de los IC, y de la cual deriva su estudio basado en la convencionalidad (“0.6. La individuación mediante fijación de conjuntos componentes en un IC siempre será convencional.”) puede no tomarse en serio, lo cual traería una transformación sustancial a la tesis de los IC, los cuales podrían seguir teniendo una estructura como la estipulada aquí, pero que sin la Def.0.3. se seguiría cierta esencialidad a cada IC, la cual un historiador de las ciencias podría llegar a justificar satisfactoriamente de sobra en ciertos casos concretos de IC.

(iii) Inconsistencias

- **“Patrones de virtud epistemológica” versus pluralismo:** La característica quinta del pluralismo explicativo definido en la sección (2.2.), lo distingue del anarquismo epistemológico, el cual es tácitamente asimilado al pluralismo “todo vale” descrito en la misma sección, en dos cuestiones: primero, aún existe posibilidad de apelar ciertos “patrones de virtud epistemológica” entre dos teorías con exactamente iguales poderes predictivos. Segundo, el pluralismo explicativo no es conservador de todo modelo o teoría simplemente para mantener la diversidad, pero tampoco adhiere rigurosamente a las integraciones. El problema particular de la característica quinta es cómo dar sentido a tales “patrones de virtud”, los cuales son tratados dentro de la segunda parte de la tesis también como “criterios para dirimir entre teorías”, compatibilizándolo con un pluralismo que, intuitivamente, no sería fácil de asimilar a tales criterios que derivarían de algún modo u otro al monismo, el cual, por ejemplo, sería alguna monista respecto a los elementos “imprescindibles” que debiesen estar en algún o algunos conjunto componentes del “IC deseable”. La inconsistencia, luego, es bastante obvia.

Una salida fácil sería restringir el rasgo quinto del pluralismo explicativo a lo segundo antes apuntado, es decir, a no ser conservador, pero esta salida fácil es claramente incompatible con la práctica científica, la cual muchas veces se guía por criterios de decisión entre teorías en competencia que no responden a presiones que devienen desde algún tipo de monismo, sino que vienen guiadas por lo pragmático, por ejemplo, permitir que un IC sea compatible con ciertos otros preexistentes o seleccionar aquel con ontología más restringida. Por tanto, es deseable mantener que un buen medio conceptual que describa relaciones “inter-teóricas” posea espacio para tales “patrones de virtud epistemológica”. Por ahora no encuentro una salida cómoda, aunque en particular este problema de dos IC que tengan exactamente el mismo poder predictivo me parece a simple vista una cuestión que no da comúnmente en las teorías o modelos, ya que la norma es que existan sutiles diferencias desde los SO hacia arriba, lo cual implica que no entran en una *competencia directa perfecta*.

- **Actitud pluralista, “actitud pluralista investigativa” y los antecedentes del pluralismo:** En el contexto de la sección (2.2.) se definió que “actitud pluralista” es una “*forma de actuar dentro de la investigación*”, que no necesariamente reconoce una epistemología pluralista. Posteriormente se denominará también como “actitud pluralista investigativa”. No obstante, el modo en el cual llegué a tal “actitud pluralista” fue bastante indirecto: a partir de la descripción de Bokulich que versaba sobre la filosofía de las ciencias de Heisenberg y mostrando cómo su filosofía ya mostraba cierto

pluralismo para compatibilizar la mecánica clásica y la cuántica, pero lo hacía antes del estudio conceptual del mismo pluralismo en el contexto de la filosofía de las ciencias. Por lo tanto, más que una actitud pluralista, según como la he definido, la de Heisenberg sería un antecedente de la diferente gama de pluralismos epistemológicos y metafísicos. Esto es importante, ya que significa que el modo mediante el cual llegué a establecer la existencia de actitud pluralista investigativa, a pesar de lo declarado en la sección (2.2.), no responde a una lectura de Bokulich (2008), sino que la he extraído intuitivamente inspirándome desde los mismos casos de ciencias cognitivas. Por tanto, la inconsistencia aquí fue declarar una *rationale* que no correspondía a la verdadera.

3.2. Perspectivas de desarrollo e investigación futura

El tratamiento que he dado desde la tesis de los IC a la polémica simbolismo-conexionismo concluye que, si se restringen las definiciones a nivel de definir el primero como “simbolismo cognitivista” y el segundo como “conexionismo subsimbolismo”, entonces tal relación inter-metateórica no cae en una categoría ni de colaboración ni de competencia. Esta conclusión se aplica de manera ambivalente en los casos de competencia que se trate de IC concretos puntuales. Sin embargo, aun cuando he llegado a completar el objetivo principal de la investigación enunciado en la introducción a este trabajo, quedan dos cuestiones sobre las cuales se requiere también cierta indagación intensiva: primero, la investigación sobre posibilidad de cierto “terreno común” en ciencias cognitivas que permita que como disciplina se comporte de modo similar a como sucede con otras ciencias y, segundo, volver a acometer la relación competitiva entre IC concretos conexionistas y simbólicos, pero armado con algún conjunto de principios de decisión o “patrones de virtud epistémica”. Brevemente, apuntaré sobre cada una de estas cuestiones algunas sugerencias.

- **La cuestión de “un terreno común” para ciencias cognitivas:** el pluralismo explicativo puede ir ligado o no una interpretación constreñida ontológicamente de los fenómenos, por ejemplo, exclusivamente cuando se trata de una teoría o modelo particular. Sin embargo, más allá de esta excepción permisible, es ante todo un enfoque epistemológico sobre el conjunto de teorías y modelos dentro de cierto ámbito, así como también es una posición de apertura a la posibilidad de cambiar la estructura de las propias teorías. Luego, el pluralismo puede dar lugar a una salida sencilla para redefinir que cuenta como “terreno común” (TC), es decir, como redefinido como más allá de un consenso sobre generalizaciones o “hechos básicos”, sino que como *una actitud pluralista en la práctica misma de la construcción de IC que versen sobre lo cognitivo.*

Sin embargo, esta salida ya ha sido tomada en la práctica introductoria a ciencias cognitivas. Por ejemplo, si toma algún manual contemporáneo a ciencias cognitivas (v. gr. Anderson, 2014 o Bermudez, 2014) se puede observar que no existe preeminencia de algún “paradigma” sobre otro, más bien una introducción a nociones básicas que pueden ser utilizadas como se estime conveniente. Esto también se puede visualizar inclusive al interior de “paradigmas”, los cuales desde la visión del lego parecen bastante uniformes como la misma neurociencia computacional, pero que no lo son realmente. Por tanto, esta salida de TC, ya ha sido tomada en la práctica y sus efectos ya son notorios en la investigación, por ejemplo, representada por la búsqueda de arquitecturas cognitivas con una integración mayor de funciones cognitivas. Es por esto que, más que buscar un TC desde el “interior” de ciencias cognitivas, lo que debiese lograrse en una investigación que

retome tal cuestión, es la pregunta *sobre si existe cierta preeminencia de ciertos IC sobre cierto trozo del F cognitivo en tanto se asuman como deseables ciertos criterios de decisión entre teorías*. Esto último conecta este tópico con el siguiente que quedó sin desarrollar y que será expuesto en el siguiente punto sobre los “criterios de decisión entre IC en competencia”.

Ahora bien, conjeturo que desde filosofía de la mente es posible desarrollar algún aporte que “actualice” la idea del funcionalismo de los 60’, el cual sirve de trasfondo filosófico unificador del cognitivismo, pero que si no es restringido a su versión computacionalista, puede, inclusive, subsumir a todos los “paradigmas” en ciencias cognitivas *sirviendo al modo de un TC*. Esto puede desarrollarse un poco más si se distingue entre dos “grandes tipos” de funcionalismo: el primer funcionalismo poseía una idea de lo mental que partía del computador. Hay un segundo tipo de funcionalismo que se basa en la conectividad neuronal y las propiedades eléctricas de las neuronas, ya que, por ejemplo, una neurona hecha de grafeno u otros materiales que pudiese replicar los detalles de conectividad eléctrica funcionaría como una. No obstante, así como el funcionalismo tipo-computador es intuitivamente inadecuado, el funcionalismo tipo-cerebro-eléctrico no es menos inadecuado.

Empero, una salida que preserve caprichosamente el “misterio” que posee lo mental me parece un extremo que tampoco se debe adoptar. Ergo, es necesario aceptar como legítimo suponer que *en algún momento nuestras réplicas cognitivas artificiales, si son cargada con suficiente fidelidad material al fenómeno “real”, debiese ser considerando como individuos legítimos de la clase de sistemas cognitivos reales*. El mismo caso del cerebro puede servir: ¿Qué sucede si se replican otras propiedades como la capacidad de cambio de los potenciales ante un estímulo químico – neurotransmisores y otros? Es más ¿Qué sucede si logramos completar la teoría de la neurona “definitivamente”? Porque, me parece claro, en algún momento llegarán a ser indistinguibles dos neuronas desde el nivel funcional. De hecho, nuestras mejores teorías finalmente a lo que aspiran es a aislar poderes causales en la materia, al punto que hemos logrado, por ejemplo, sintetizar artificialmente diamantes, los cuales son indistinguibles de los naturales desde la perspectiva de nuestras teorías químicas de la materia (aunque sí distinguibles desde los niveles físicos). Es el caso que “lo mental” se da en diferentes configuraciones anatómicas y con diferencias bioquímicas, por lo tanto nos queda, o bien una teorización *bottom-up* estricta que conducirá necesariamente a una cantidad potencialmente jamás instrumentable de “tipos de mentes”, o bien una teorización integral que incluya lo *top-down*. Dado que la segunda estrategia parece más razonable por cuestiones prácticas, entonces alguna forma de funcionalismo es la adecuada.

Mi idea es simple: tal funcionalismo adecuado al caso de la cognición no es ni uno tipo-computador ni tipo-cerebro-eléctrico, por lo tanto debe ser algún tipo de funcionalismo que se apoye de manera definitiva de nuestras mejores teorías sobre la bioquímica y ciencia de los materiales; tal funcionalismo podría denominarse como *restringido materialmente*. Luego, si se está de acuerdo con tal clase de funcionalismo como TC, entonces una consecuencia inmediata sería que *los procesos cognitivos quedarían mejor delimitados en tanto se les supone una base material definitiva*. El funcionalismo restringido empíricamente se ajustaría mejor a la colaboración, ya que sería *por definición dependiente de un mayor rango de IC que los anteriores funcionalismos*.

- **Criterios de decisión entre IC en competencia:** Sobre esta cuestión algo he reflexionado, pero la justificación y aplicación no ha sido desarrollada, por lo tanto lo que a continuación expondré serán sugerencias que permitirían mejorar la tesis de los IC a costa de perder cierto “agnosticismo” y, probablemente, a costa de estar totalmente errado:

Def.9. Un IC es mejor que otro si, cumpliéndose conmensurabilidad local, posee elementos constitutivos que no posee su contraparte y tiene mayor poder predictivo-explicativo.

Esta definición explícitamente renuncia al “agnosticismo” que poseían las ocho definiciones, y sus consecuencias respectivas, que se desarrollan en la sección (2.3) de la segunda parte de la tesis y sus subsecciones. Ergo, apelando a esta definición el remate de la sección (2.4.2.) cambiaría ostensiblemente declarando que, por ejemplo, el perceptrón es derechamente *mejor* que la neurona de M&P, siendo no válida la contraargumentación desde el innatismo. La única defensa a la que podría recurrir un innatista sería decir que la neurona de M&P no busca poder predictivo-explicativo respecto al fenómeno cognitivo del aprendizaje, lo cual sería un escape bastante excéntrico, ya que tal función es central a la cognición.

Def.10.1. Un IC es “profundo” o “no especulativo” en grados descendentes al cumplirse alguna de las siguientes condiciones:

10.1.1. Si existe al menos un ca que sea materialmente idéntica entre el SO y el F.

10.1.2. Si existe al menos un ca en SO que sea materialmente “rastreadable” en F.

Una defensa de la validez de la Def.10.1. podría ser la siguiente: considérese una clase de modelo muy simple que es usado usualmente para responder cuestiones técnicas más que científicas, los modelos a escala. Supongamos que tenemos dos modelos a escala de una silla de madera, descontextualizados de su propio IC, donde uno es un modelo que es del mismo material que la silla original y otro que es de un material diferente. Dado que ambos, por definición, están a escala de la silla original, entonces no hay manera de dirimir entre ellos por cuál es mejor dado que están descontextualizados del IC donde son parte. En cambio, es posible, intuitivamente, apelar a que aquel modelo que es del mismo material que la silla original es, bajo ciertas condiciones contextuales, mejor modelo que el otro modelo que no es del mismo material. Por ejemplo, si se quisiera evaluar la rapidez con la cual se quemaría la silla original, el modelo a escala del mismo material es perfecto. En cambio, si se quisiera evaluar exclusivamente la funcionalidad de la silla original, entonces ambos modelos a escala son igualmente útiles. Pareciera ser intuitivo que, dejando fijo el parámetro de la forma, es la materia misma la cual nos permite dirimir entre cuál de dos modelos a escala tiene potencialmente mayor cantidad de contextos de utilidad.

Por otra parte, las condiciones 10.1.1. y 10.1.2. permiten distinguir, respectivamente, entre los IC que están primariamente contruidos materialmente de los cuales están contruidos formalmente. El problema que preveo central a esta cuestión es el significado de “rastreadable”. El sentido es fácilmente visible en IC contruidos formalmente como el perceptrón, pero no es así el caso con uno simbólico, ¿cómo “rastrear”, por ejemplo, un “esquema”? Lo anterior, ¿transforma de inmediato los IC simbólicos en “especulativos” (Cfr. Def.10.2. a continuación)? No me atrevería a llegar tan lejos, ya que hacerlo sería negar que el modelamiento simbólico tiene su propio SON que los avala caso a caso. Algo que está ínsito

en las ocho definiciones centrales es que *la evaluación del éxito de un IC depende exclusivamente de que el SON sea logrado*.

Def.10.2. Un IC es “especulativo” si no cumple con la Def.10.1.

Si algo existe como hecho innegable dentro de la ciencia es que mucha de ella, al menos muchas teorías bastante afamadas, parte guiada por una idea metafórica o un pensamiento analógico (Nersessian & Magnani, 2002) sobre los fenómenos, donde se especula que el fenómeno de interés tendría un funcionamiento equivalente al funcionamiento de otro cualquiera ya bien conocido – nótese que este uso de las metáforas es diferente del uso pedagógico. Esta búsqueda de “isomorfismos” se iría refinando hasta llegar con una formalización basada en características abstraídas “rastreables” (10.1.2.), la cual transformaría este IC especulativo – clase del cual los IC “metafóricos” serían su caso más representativo – en uno no especulativo, es decir, le otorgaría cierto grado de profundidad.

Por consiguiente, si se uniesen las ideas presentadas dentro de la Def.10.1. y Def.10.2., entonces se podría generar un escala que iría desde el menos profundo – “metafórico/analógico” – pasaría por otros intermedios – matemático/formal – y terminaría en aquellos de máxima profundidad donde el fenómeno mismo es aislado en ciertos de sus componentes materiales – “nivel material específico”.

Def.11. Un IC es “hijo” de otro IC si y sólo si a través de análisis mediante IC se muestra cómo, cumpliéndose conmensurabilidad local, extiende la estructura de un IC anterior de modo de abarcar mejor explicativa-predictivamente cierto F.

Def.11.1. Un IC es más “fértil” relativamente a otro dependiendo de la cantidad de IC-hijos que posee.

La fertilidad definida mediante Def.11. y Def.11.1. es fácilmente visible como un buen criterio para evaluar entre dos IC. Una idea base como esta definición ya estaba contenida en una nota de la segunda parte de la tesis cuando nombré el “*potencial evolutivo de la teoría*”, el cual podría ser investigado a partir de la noción de fertilidad.

Bibliografía

- Abdi, H. (1994). A neural network primer. *Journal of Biological Systems*, 2(03), 247-281.
- Abraham, T. H. (2002). (Physio)logical circuits: The intellectual origins of the McCulloch–Pitts neural networks. *Journal of the History of the Behavioral Sciences*, 38(1), 3-25.
- Abraham, T. H. (2003). From theory to data: Representing neurons in the 1940s. *Biology and Philosophy*, 18(3), 415-426.
- Abraham, T. H. (2004). Nicolas Rashevsky's mathematical biophysics. *Journal of the History of Biology*, 37(2), 333-385.
- Aizawa, K. (1997). Exhibiting versus Explaining Systematicity: A Reply to Hadley and Hayward 1. *Minds and Machines*, 7(1), 39-55.
- Alfonseca, M; De la Cruz, M; Ortega, A, Pulido, E. (2006). *Compiladores e intérpretes: teoría y práctica*. Pearson Prentice Hall, Madrid.
- Anderson, B. (2014). *Computational neuroscience and cognitive modelling: a student's introduction to methods and procedures*. Sage.
- Anderson, J. (2007). *Redes Neurales*. Alfaomega Grupo Editor, S.A., México.
- Anderson, J. A. & Rosenfeld, E. (1998). *Talking Nets: An Oral History Of Neural Networks*. MIT Press
- Anderson, J. A., & Rosenfeld, E. *Neurocomputing: Foundations of research* (1988). Cambridge, MA: MIT Press.
- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?*. Oxford University Press.
- Andler, D. (1992). From Paleo-to Neo-Connectionism. En Vijver, G. (1992). *New perspectives on cybernetics: self-organization, autonomy and connectionism*. Springer Science & Business Media.
- Barnden, J. A. (1992). Connectionism, generalization, and propositional attitudes: A catalogue of challenging issues. En Dinsmore, J. (1992). *The Symbolic and Connectionist Paradigms: Closing the Gap*. Lawrence Erlbaum Associates Publishers.
- Bechtel, W. & Richardson, R. (2010) *Discovering Complexity - Decomposition and Localization as Strategies in Scientific Research*. Cambridge, MA: MIT Press
- Bechtel, W. (1985). Contemporary connectionism: are the new parallel distributed processing models cognitive or associationist?. *Behaviorism*, 13(1), 53-62.
- Bechtel, W. (1988). Connectionism and rules and representation systems: Are they compatible?. *Philosophical Psychology*, 1(1), 5-16.

- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421-441.
- Bechtel, W., & Abrahamsen, A. A. (1990). Beyond the exclusively propositional era. *Synthese*, 82(2), 223-253.
- Bechtel, W., Abrahamsen, A., & Graham, G. (1998). The life of cognitive science. En Bechtel, W. & Graham, G. (1998). *A companion to cognitive science*. Oxford: Blackwell.
- Beer, R. D., Chiel, H. J., & Sterling, L. S. (1989). Heterogeneous neural networks for adaptive behavior in dynamic environments. In *Advances in neural information processing systems* (pp. 577-585).
- Beer, R. D., Chiel, H. J., Quinn, R. D., Espenschied, K. S., & Larsson, P. (1992). A distributed neural network architecture for hexapod robot locomotion. *Neural Computation*, 4(3), 356-365.
- Bennett, M. R. (1999). The early history of the synapse: from Plato to Sherrington. *Brain research bulletin*, 50(2), 95-118.
- Bermudez, J. L. (2005). *Philosophy of Psychology: A Contemporary Introduction*. Routledge.
- Bermúdez, J. L. (2014). *Cognitive science: An introduction to the science of the mind*. Cambridge University Press.
- Bezrucho, B. P., & Smirnov, D. A. (2010). *Extracting knowledge from time series: An introduction to nonlinear empirical modeling*. Springer Science & Business Media.
- Boden, M. A. (2006). *Mind as Machine: A History of Cognitive Science*. Oxford University Press.
- Bokulich, A. (2008). *Reexamining the quantum-classical relation*. Cambridge University Press.
- Bridgman, P (1991 [1927]) The operational character of scientific concepts. En Boyd, R., Gasper, P., & Trout, J. D. (1991). *The philosophy of science*. MIT Press.
- Brooks, A (2009) The possibility architecture cognitive. En Dedrick, D. & Trick, L. (eds). (2009) *Computation, Cognition, and Pylyshyn*. Cambridge, MA: MIT Press.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, 47(1), 139-159.
- Brooks, R. A. (1999). *Cambrian intelligence: the early history of the new AI*. MIT Press.
- Bullinaria, J. A. (1994). *Learning the past tense of English verbs: Connectionism fights back*. Manuscrito no publicado.
- Callister, W. D., & Rethwisch, D. G. (2014). *Materials science and engineering: an introduction* (9th ed.). New York: Wiley.

- Calvo Garzon, F. (2003). Nonclassical connectionism should enter the decathlon. *Behavioral and Brain Sciences*, 26(05), 603-604.
- Calvo, P., & Gomila, T. (Eds.). (2008). *Handbook of cognitive science: An embodied approach*. Elsevier.
- Casadevall, A. & Fang, F. C. (2008) Descriptive Science. *Infection & Immunity*, p. 3835–383
- Catterall, W. A., Raman, I. M., Robinson, H. P., Sejnowski, T. J., & Paulsen, O. (2012). The Hodgkin-Huxley heritage: from channels to circuits. *The Journal of Neuroscience*, 32(41), 14064-14073.
- Chalmers, D. J. (1993). Why Fodor and Pylyshyn were wrong: The simplest refutation. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pp. 340-347.
- Chemero, A. (2009). *Radical embodied cognitive science*. MIT Press.
- Chemero, A., & Silberstein, M. (2008). After the Philosophy of Mind: Replacing Scholasticism with Science. *Philosophy of Science*, 75(1), 1-27.
- Chomsky, N. (1956). Three models for the description of language. *Information Theory, IRE Transactions on*, 2(3), 113-124.
- Chomsky, N. (1967). *Recent contributions to the theory of innate ideas* (pp. 31-40). Springer Netherlands.
- Churchland P.S. & Sejnowski, T. (2006 [1989]). Neural Representation and Neural Computation. En Bermúdez, J. L. (Ed.). (2006). *Philosophy of psychology: contemporary readings*. Routledge.
- Churchland, P. M. (1984). *Matter and consciousness: A contemporary introduction to the philosophy of mind*. Cambridge, MA: MIT Press.
- Churchland, P. S. (1989). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT press.
- Cioffi-Revilla, C. (2014). *Introduction to Computational Social Science: Principles and Applications*. Springer Science & Business Media.
- Clark, A. (1998). *Being there: Putting brain, body, and world together again*. MIT press.
- Clark, A., & Toribio, J. (1994). Doing without representing?. *Synthese*, 101(3), 401-431.
- Cleeremans, A. (2014). Connecting conscious and unconscious processing. *Cognitive science*, 38(6), 1286-1315.
- Clemence, G. M. (1947). The relativity effect in planetary motions. *Reviews of Modern Physics*, 19(4), 361.
- Cohen, I. B., & Smith, G. E. (2002). *The Cambridge Companion to Newton*. Cambridge University Press.

- Cowan, J. D. (1998). Entrevista con J.A. Anderson y E. Rosenfeld. En Anderson, J. A. & Rosenfeld, E. (1998). *Talking Nets: An Oral History Of Neural Networks*. MIT Press.
- Cruse, H., & Schilling, M. (2015). Mental states as emergent properties. From walking to consciousness. *Open Mind*.
- Cruse, H., Kindermann, T., Schumm, M., Dean, J., & Schmitz, J. (1998). Walknet – a biologically inspired network to control six-legged walking. *Neural networks*, 11(7), 1435-1447.
- Cummins R. (1975) Functional Analysis. *Journal of Philosophy* 72:741-64
- Dale, R. (2008). The possibility of a pluralist cognitive science. *Journal of Experimental and Theoretical Artificial Intelligence*, 20(3), 155-179.
- Dawson, M. R. (2005). *Connectionism: A hands-on approach*. John Wiley & Sons.
- Dawson, M. R., Medler, D. A., & Berkeley, I. S. (1997). PDP networks can provide models that are not mere implementations of classical theories. *Philosophical Psychology*, 10(1), 25-40.
- Degenaar, J., & Myin, E. (2014). Representation-hunger reconsidered. *Synthese*, 191(15), 3639-3648.
- Demuth, H. B., Beale, M. H., De Jess, O., & Hagan, M. T. (2014). *Neural network design*.
- Dennett (1998 [1986]) The logical geography of computational approaches: A view from the East Pole. En Dennett (1998) *Brainchildren*. MIT Press.
- Dietrich, E. (2007) Representation. En Gabbay, D. M., Woods, J., & Thagard, P. (2007). *Philosophy of Psychology and Cognitive Science: A Volume of the Handbook of the Philosophy of Science Series*. Elsevier.
- Dolan, C. P., & Smolensky, P. (1989). Tensor product production system: a modular architecture and representation. *Connection Science*, 1(1), 53-68.
- Douglas, H. E. (2009). Reintroducing prediction to explanation. *Philosophy of Science*, 76(4), 444-463.
- Dreyfus, H. L., & Dreyfus, S. E. (1988). Making a mind versus modelling the brain: artificial intelligence back at a branchpoint. *Daedalus*, 117(1), 185-197.
- Dupuy, J. P. (2009). *On the origins of cognitive science: the mechanization of the mind*. MIT Press.
- East, I. (1990). *Computer architecture and organization*. Pitman Publishing.
- Edelman, S. (2008). On the nature of minds, or: truth and consequences. *Journal of Experimental & Theoretical Artificial Intelligence*, 20(3), 181-196.
- Ellis, B. (1956) On the Relation of Explanation to Description. *Mind*, Vol. 65, No. 260, pp. 498-506

- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71-99.
- Eysenck, M. W., & Keane, M. T. (2010). *Cognitive psychology: a students handbook*. Psychology Press.
- Faraci, F. (2012) *The 60th anniversary of the Hodgkin-Huxley model. A critical assessment from a historical and modellers viewpoint*. Tesis, Leiden University
- Feyerabend, P. (1975). *Against Method: Outline of an Anarchism Theory of Knowledge*. Verso Books.
- Finger, S. (2000). *Minds behind the brain: A history of the pioneers and their discoveries*. Oxford University Press.
- Flanagan, O. (2006). Varieties of naturalism. *Oxford handbook of religion and science*, 430-452.
- Fodor, J. (1997) Connectionism and the problem of systematicity (continued)- why Smolensky's solution still doesn't work. *Cognition*, 62(1), 109-119.
- Fodor, J. A. (1974). Special sciences (or: the disunity of science as a working hypothesis). *Synthese*, 28(2), 97-115.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.
- Fodor, J. A. (1989). Why there still has to be a language of thought. In *Computers, Brains and Minds* (pp. 23-46). Springer Netherlands.
- Fodor, J., & McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, 35(2), 183-204.
- Fodor, J., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2), 3-71.
- Forster, M., & Sidel, E. (1994). Connectionism and the fate of folk psychology: a reply to Ramsey, Stich and Garon. *Philosophical Psychology*, 7(4), 437.
- Franchi, S., & Bianchini, F. (Eds.). (2011). *The Search for a Theory of Cognition: Early Mechanisms and New Ideas*. Rodopi.
- Gardner, H. (1985). *The Mind's New Science. A History of the Cognitive Revolution*. Basic Books.
- Gentner, D. (2002). Analogy in scientific discovery: The case of Johannes Kepler. En Magnani, L. & Nersessian, N. (2002). *Model-based reasoning: science, technology, values*. Springer Science & Business Media.
- Gerard, R. W. (1951). Some of the problems concerning digital notions in the central nervous system. En Von Foerster, H. (1951) *Cybernetics: Circular causal and feedback mechanisms in biological and social systems. Transactions of the Seventh Conference*. New York, Macy Foundation.

- Gerstner, W. (1998). Spiking neurons. En Maass, W., & Bishop, C. M. (1998). *Pulsed neural networks*. MIT Press.
- Gibson, R. F. (2004) *The Cambridge Companion to Quine*. Cambridge University Press
- Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological bulletin*, 119(1), 23-26.
- Gleick, J. (2011). *The Information: A History, A Theory, A Flood*. Pantheon Books.
- Godfrey-Smith, P. (1993) Functions consensus without unity *Pacific Philosophical Quarterly* 74: 196-208.
- Gomila, T. (2012). *Verbal Minds: Language and the Architecture of Cognition*. Elsevier.
- Goodwin, G. C., Graebe, S. & Salgado, M. (2000) *Control System Design*. Prentice Hall.
- Guarini, M. (1998). *Rules and representations in the classicism-connectionism debate*. Doctoral dissertation, The University of Western Ontario London.
- Hanson, N. R. (1962). Leverrier: the zenith and nadir of Newtonian mechanics. *Isis*, 359-378.
- Haselager, P., de Groot, A., & van Rappard, H. (2003). Representationalism vs. anti-representationalism: a debate for the sake of appearance. *Philosophical Psychology*, 16(1), 5-24.
- Hava Siegelmann. (1999). *Neural Networks and Analog Computation: Beyond the Turing Limit*. Springer Science & Business Media.
- Haykin, S. (2008) *Neural networks a comprehensive foundation*. Prentice-Hall, New Jersey
- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. En Anderson, J. A., & Rosenfeld, E. (1988) *Neurocomputing: Foundations of research*. Cambridge, MA.
- Hempel, C & Oppenheim P. (1948) Studies in the Logic of Explanation. *Philosophy of Science*, Vol. 15, No. 2., pp. 135-175.
- Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. *Parallel models of associative memory*, 161-187.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4), 500.
- Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2006). Automata theory, languages, and computation. *International Edition*, 24.
- Hutten, E. H. (1954). The role of models in physics. *The British Journal for the Philosophy of Science*, 4(16), 284-301.

- Jilk, D. J., Lebiere, C., O'Reilly, R. C., & Anderson, J. R. (2008a). SAL: An explicitly pluralistic cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 20(3), 197-218.
- Jilk, D. J., Lebiere, C., O'Reilly, R. C., & Anderson, J. R. (2008b). Beyond red states and blue states in cognitive science. *Journal of Experimental and Theoretical Artificial Intelligence*, 20(3), 265-268.
- Johnson, K. (2004). On the systematicity of language and thought. *The Journal of Philosophy*, 101(3), 111-139.
- Keil, F. C. (2010). The feasibility of folk science. *Cognitive science*, 34(5), 826-862.
- Keil, F., & Wilson, R. (Eds.). (1999). *The MIT encyclopedia of cognitive science*. MIT Press.
- Kellert, S. H., Longino, H. E., & Waters, C. K. (Eds.). (2006). *Scientific pluralism [electronic resource]* (Vol. 19). U of Minnesota Press.
- Kleene, S. (1951). *Representation of events in nerve nets and finite automata*. RAND Project Air Force, Santa Monica, CA.
- Knight, B. W. (1972). Dynamics of encoding in a population of neurons. *The Journal of general physiology*, 59(6), 734-766.
- Kolak, D., Hirstein, W., Mandik, P., & Waskan, J. (2006). *Cognitive science: An introduction to mind and brain*. Routledge.
- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions, 2nd enl. ed.* University of Chicago Press.
- Lange, T. (1992). Hybrid Connectionist Models: Temporary Bridges Over the Gap Between the Symbolic and the Subsymbolic. En Dinsmore, J. (1992). *The Symbolic and Connectionist Paradigms: Closing the Gap*. Lawrence Erlbaum Associates Publishers.
- Lapicque, L. (1907) Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen*, 9(1), 620-635.
- Lashley, K. S. (1950). In search of the engram. En Anderson, J. A., & Rosenfeld, E. *Neurocomputing: Foundations of research* (1988) Cambridge, MA.
- Lee, G., Flowers, M., & Dyer, M. G. (1990). Learning distributed representations of conceptual knowledge and their application to script-based story processing. *Connection Science*, 2:4, 313-345.
- Ling, C. X. & Marinov, M. (1993). Answering the connectionist challenge: A symbolic model of learning the past tenses of English verbs. *Cognition*, 49(3), 235-290.
- Ling, C. X. (1994). Learning the past tense of English verbs: The symbolic pattern associator vs. connectionist models. *Journal of Artificial Intelligence Research*.

- López-Muñoz, F., Boya, J., & Alamo, C. (2006). Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramón y Cajal. *Brain research bulletin*, 70(4), 391-405.
- Maass, W. & Bishop, C. M. (1998). *Pulsed Neural Networks*. MIT press.
- Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9), 1659-1671.
- MacGregor, R. J. (1987). *Neural and brain modelling*. Academic Press.
- Magnani, L. & Nersessian, N. (2002). *Model-based reasoning: science, technology, values*. Springer Science & Business Media.
- Marr, D. Ullman, S., & Poggio, T. A (2010 [1982]) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press.
- Marraffa, M., & Paternoster, A. (2012). Functions, levels, and mechanisms: Explanation in cognitive science and its problems. *Theory & Psychology*, 23(1), 22-45.
- Martínez-Freire, P. (2011) Is Embodied Cognition a Good Conceptual Revolution?. En González, W. (Ed.) (2011) *Conceptual Revolutions From Cognitive Science to Medicine*. Netbiblo.
- McClelland, J. L. & Rumelhart, D. E., & PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2 Psychological and Biological Models*. Cambridge, MA. The MIT Press.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- McLaughlin, B. P. (1993a). The connectionism/classicism battle to win souls. *Philosophical Studies*, 71(2), 163-190.
- McLaughlin, B. P. (1993b). Systematicity, conceptual truth, and evolution. *Royal Institute of Philosophy Supplement*, 34, 217-234.
- Medler, D. A. (1998). A brief history of connectionism. *Neural Computing Surveys*, 1, 18-72.
- Mehrotra, K., Mohan, C. K., & Ranka, S. (1997). *Elements of artificial neural networks*. MIT press.
- Menary, R. (2007). *Cognitive integration: Mind and cognition unbounded*. Basingstoke: Palgrave Macmillan.
- Mesarovic, M. D., Waterman, T.H. (Ed.). (1968). *Systems theory and biology: proceedings of the III Systems Symposium at Case Institute of Technology*. Springer Science & Business Media.
- Miller, D. L. (1947). Explanation versus description. *The Philosophical Review*, 56(3), 306-312.

- Miller, G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2), 81.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Holt, Rinehart and Winston, Inc.
- Mindell, D. A. (2002). *Between human and machine: feedback, control, and computing before cybernetics*. JHU Press.
- Minsky, M. & Papert, S. *Perceptrons* (1969). En Anderson, J. A., & Rosenfeld, E. *Neurocomputing: Foundations of research* (1988) Cambridge, MA: MIT Press.
- Minsky, M. (1963). *Neural Nets and Theories of Memory*. Technical Report. Cambridge, MA: MIT
- Mix, K. S., Smith, L. B., & Gasser, M. (Eds.). (2010). *The spatial foundations of cognition and language: thinking through space* (Vol. 4). OUP Oxford.
- Nagel, E. (1961). *The structure of science: Problems in the Logic of Scientific Explanation*. New York, Routledge.
- Nagel, S. (2005). *Embodied, Dynamical Representationalism – Representations in Cognitive Science*. Osnabruck, Germany: PICS.
- Nazarova, M., & Blagovechtchenski, E. (2015). Modern Brain Mapping – What Do We Map Nowadays? *Frontiers in psychiatry*, 6:89
- Nersessian, N. (2008) *Creating Scientific Concepts*. Cambridge, MA: MIT Press
- Newell, A. (1982). *Intellectual issues in the history of artificial intelligence* (No. CMU-CS-82-142). Carnegie-Mellon Univ Pittsburgh PA Dept of Computer Science.
- Newell, A., & Simon, H. A. (1956). The logic theory machine--A complex information processing system. *Information Theory, IRE Transactions on*, 2(3), 61-79.
- Newell, A., & Simon, H. A. (1961). Computer Simulation of Human Thinking A theory of problem solving expressed as a computer program permits simulation of thinking processes. *Science*, 134(3495), 2011-2017.
- Newell, A., & Simon, H. A. (1963). GPS, a program that simulates human thought. En Feldman, J., & Feigenbaum, E. A. (Eds.). (1963). *Computers and Thought*. McGraw-Hill.
- Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3), 113-126.
- Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological review*, 65(3), 151.

- Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological review*, 65(3), 151.
- Newell, A., Shaw, V.C. and Simon, H.A. (1959). Report on a general problem-solving program. Proceedings ICIP. Paris: UNESCO House
- Novikov, D. A. (2016). *Cybernetics: From Past to Future* (Vol. 47). Springer.
- Oaksford, M.; Chater, N. and Stenning, K. (1990). Connectionism, Classical Cognitive Science and Experimental Psychology. *AI & Society*. 4: 73-90.
- Parks, R. W., Levine, D. S., & Long, D. L. (1998). *Fundamentals of neural network modeling: Neuropsychology and cognitive neuroscience*. MIT Press.
- Piccinini, G. (2004). The First computational theory of mind and brain: a close look at McCulloch and Pitts's "logical calculus of ideas immanent in nervous activity". *Synthese*, 141(2), 175-215. Springer.
- Piccinini, G. (2012). Computationalism. En Margolis, E. & Stich, S. (2012) *The Oxford Handbook of Philosophy and Cognitive Science*. Oxford University Press.
- Piccinini, G., & Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive science*, 37(3), 453-488.
- Pinker, S. (1997). *How the mind works*. Penguin Books. NY: Norton.
- Pinker, S. (2003). *La tabla rasa: La negación moderna de la naturaleza humana*. Paidós, Barcelona.
- Pinker, S., & Mehler, J. (1988). *Connections and symbols* (Vol. 28). Mit Press.
- Port, R. F., & Van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. MIT press.
- Putnam, H. (1975 [1967]) The Nature of Mental States. En Putnam (1975) *Mind, Language, and Reality*, Cambridge: Cambridge University Press.
- Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3(01), 111-132.
- Quillian, M. R. (1988 [1968]). Semantic memory. En Collins, A., & Smith, E. E. (1988). *Readings in Cognitive Science, a Perspective From Psychology and Artificial Intelligence*. Morgan Kaufmann Publishers.
- Quine, W. V. O. (1951) Two Dogmas of Empiricism. *The Philosophical Review*, Vol. 60, No. 1, 20-43.
- Quine, W. V. O. (1968 [1960]) *Palabra y Objeto*. Barcelona, España: Ed. Labor.
- Quine, W. V. O. (1969). *Ontological relativity and other essays*. Columbia University Press.

- Quine, W. V. O. (2001) *Acerca del conocimiento científico y otros dogmas*. Barcelona, España:
- Rabossi, E. (1997). *Filosofía de la mente y ciencia cognitiva*. Barcelona; Paidós.
- Ramírez, A. (2005) *La transformación de la epistemología contemporánea: de la unidad a la dispersión*. Santiago de Chile: Editorial Universitaria.
- Ramsey, W., Stich, S., & Garon, J. (1990). Connectionism, Eliminativism and The Future of Folk Psychology. *Philosophical Perspectives*, 4, 499-533.
- Rojas, R. (1996) *Neural Networks-A Systematic Introduction*. Springer-Verlag. New York.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386. En Anderson, J. A., & Rosenfeld, E. *Neurocomputing: Foundations of research* (1988) Cambridge, MA: MIT Press.
- Rosenbloom, P. S., Laird, J. E., Newell, A., & McCarl, R. (1991). A preliminary analysis of the Soar architecture as a basis for general intelligence. *Artificial Intelligence*, 47(1), 289-325.
- Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of science*, 10(1), 18-24.
- Rottschaefer, W. A. (1976) Observation – Theory-Laden, Theory-Neutral or Theory-Free? *The Southern Journal of Philosophy*, 14(4), 499-509.
- Rumelhart, D. E., McClelland, J. L., & PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1 Foundations*. Cambridge, MA: MIT Press.
- Sarle, W.S., ed. (1997), Neural Network FAQ, part 1 of 7: Introduction, periodic posting to the Usenet newsgroup comp.ai.neural-nets, URL: <ftp://ftp.sas.com/pub/neural/FAQ.html>:
- Schickore, J., & Steinle, F. (Eds.). (2006). *Revisiting discovery and justification: Historical and philosophical perspectives on the context distinction* (Vol. 14). Springer Science & Business Media.
- Scott, A. (2002). *Neuroscience: A mathematical primer*. Springer Science & Business Media.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, 3(03), 417-424.
- Sejnowski, T. J. (1991). David Marr: A Pioneer in Computational Neuroscience. En Vaina, L. M., & Poggio, T. (1991). *From the retina to the neocortex: selected papers of David Marr*. Birkhauser Boston.
- Selfridge, O. G. (1958). Pandemonium: a paradigm for learning. En Anderson, J. A., & Rosenfeld, E. *Neurocomputing: Foundations of research* (1988) Cambridge, MA: MIT Press.
- Serban, M. (2014). *Towards explanatory pluralism in cognitive science*. Doctoral dissertation, University of East Anglia.

- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press. Urbana, IL.
- Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*(2), 209-215.
- Simmons, G. F., & Robertson, J. S. (2002). *Ecuaciones diferenciales: con aplicaciones y notas históricas*. Mc Graw Hill, España.
- Simon, H. A. (1978) Information-Processing Theory of Human Problem Solving. En Estes, W. K. (ed.) *Handbook of Learning and Cognitive Processes*. Lawrence Erlbaum Associates.
- Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, *26*(2), 145.
- Slovan, S. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, *119*(1), 3.
- Smith, L. S. (2006). Implementing neural models in silicon. En Zomaya, A. Y. (Ed.). (2006) *Handbook of nature-inspired and innovative computing: integrating classical models with emerging technologies*. Springer Science & Business Media.
- Smolensky, P. (1987a). The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn. *The Southern Journal of Philosophy*, *26*(S1), 137-161.
- Smolensky, P. (1987b). *On the Proper Treatment of Connectionism*. Boulder, CO: University of Colorado, Department of Computer Science & Institute of Cognitive Science.
- Smolensky, P. (1987c). On variable binding and the representation of symbolic structures in connectionist systems (Internal Rep. No. CU-CS-355-87). Boulder, CO: University of Colorado, Department of Computer Science & Institute of Cognitive Science.
- Smolensky, P. (1987d). Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review*, *1*(2), 95-109.
- Smolensky, P. (1988a). *Connectionism, constituency, and the language of thought*. Boulder, CO: University of Colorado, Department of Computer Science & Institute of Cognitive Science.
- Smolensky, P. (1988b). Putting together connectionism—again. *Behavioral and Brain Sciences*, *11*(01), 59-74. Cambridge Univ Press.
- Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, *46*(1), 159-216.

- Smolensky, P. (1995). Reply: Constituent Structure and Explanation in an Integrated Connectionist /Symbolic Cognitive Architecture. En Macdonald, C., & Macdonald, G. (1995). *Connectionism: debates on psychological explanation*. Blackwell.
- Spencer, J. P., Perone, S., & Buss, A. T. (2011). Twenty years and going strong: A dynamic systems revolution in motor and cognitive development. *Child development perspectives*, 5(4), 260-266.
- Stahal, G. (1970). *Al explorar lo infinito*. Santiago de Chile: Ed. Universitaria.
- Steinle, F. (2006). Concept formation and the limits of justification: "Discovering" the two electricities. En Schickore, J., & Steinle, F. (2006). *Revisiting discovery and justification: Historical and philosophical perspectives on the context distinction*. Springer Science & Business Media.
- Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, 3(2), 425-437.
- Stewart, J. R., Gapenne, O., & Di Paolo, E. A. (2010). *Enaction: Toward a new paradigm for cognitive science*. MIT Press.
- Stillings, N. A., Weisler, S. E., Chase, C. H., Feinstein, M. H., Rissland, E. L., & Garfield, J. L. (1995). *Cognitive science: An introduction*. Bradford Book.
- Suppes, P. (1978). The plurality of science. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Vol. 1978, pp. 3-16). University of Chicago Press, Springer, Philosophy of Science Association.
- Tabor, W. (2009). A dynamical systems perspective on the relationship between symbolic and non-symbolic computation. *Cognitive neurodynamics*, 3(4), 415-427.
- Thagard, P. (1988). *Computational Philosophy of Science*. Cambridge, MA: MIT Press.
- Thagard, P. (1992). *Conceptual revolutions*. Princeton University Press.
- Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT press.
- Tilley, R. J. (2013). *Understanding solids: the science of materials* (2nd ed.) John Wiley & Sons.
- Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *J. of Math*, 58(345-363), 5.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.
- Van Bouwel, J. (2014). Pluralists about pluralism? Different versions of explanatory pluralism in psychiatry. In *New directions in the philosophy of science* (pp. 105-119). Springer International Publishing.

Van de Vijver, G. (1992). The Experimental Epistemology of Walter S. McCulloch a Minimalistic Interpretation. En Vijver, G. (1992). *New perspectives on cybernetics: self-organization, autonomy and connectionism*. (Vol. 220). Springer Science & Business Media.

Van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, 14(3), 355-384.

Van Gelder, T. (1995). What might cognition be, if not computation?. *The Journal of Philosophy*, 92(7), 345-381.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and brain sciences*, 21(05), 615-628.

Varela, F. J. (2002). *Conocer: las ciencias cognitivas, tendencias y perspectivas: cartografía de las ideas actuales*. Gedisa.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and human experience*. Cambridge, CA: MIT Press.

Vassallo, N., & Bianchi, C. (2009). *Naturalizing meaning through epistemology: some critical notes* (pp. 311-321). En Suárez, M., Dorato, M., & Rédei, M. (2010) *EPSA Epistemology and Methodology of Science: Launch of the European Philosophy of Science Association*. Springer.

Verlinde, E. (2011). On the origin of gravity and the laws of Newton. *Journal of High Energy Physics*, 2011(4), 1-27.

Von Eckardt, B. (1995). *What is cognitive science?*. MIT press.

Von Foerster, H. (Ed.). (1953). *Cybernetics: circular causal and feedback mechanisms in biological and social systems: transactions of the eighth conference, March 15-15, 1951, New York, NY*. Josiah Macy, Jr. Foundation.

Von Foerster, H. (Ed.). (1955). *Cybernetics, Circular Causal and Feedback Mechanisms in Biological and Social Systems: Transactions of the Tenth Conference April 22, 23, and 24, 1953, Princeton, NJ*. Josiah Macy, Jr. Foundation.

Weisberg, M. (2007). Who is a Modeler?. *The British journal for the philosophy of science*, 58(2), 207-233.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. En Anderson, J. A., & Rosenfeld, E. (1988) *Neurocomputing: Foundations of research Cambridge*. Cambridge, MA: MIT Press.

Wiener, N. (1961). *Cybernetics or Control and Communication in the Animal and the Machine* (Vol. 25). MIT press.

Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Cambridge, MA: MIT Press.

Zednik, C. (2009). The Varieties of Dynamicism. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2298-2303). Austin, TX: Cognitive Science Society.

Zenil, H. (2005) *Encaje de las Redes Neuronales Recurrentes Analógicas en la Jerarquía Aritmética y los lenguajes que reconocen según la complejidad de sus pesos*. Tesis UNAM.

Zubicaray, G. I. (2006). Cognitive neuroimaging: Cognitive science out of the armchair. *Brain and Cognition*, 60(3), 272-281.