



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN PSICOLOGÍA

MODELAMIENTO NEURO-COMPUTACIONAL DEL
APRENDIZAJE DE CLASES DE ESTÍMULOS
EQUIVALENTES.

TESIS

QUE PARA OPTAR POR EL GRADO DE
DOCTOR EN PSICOLOGÍA

PRESENTA

ÁNGEL EUGENIO TOVAR Y ROMO

TUTOR PRINCIPAL:

DR. ALVARO FLORENCIO TORRES CHÁVEZ
FACULTAD DE PSICOLOGÍA UMAM

MIEMBROS DEL COMITÉ TUTOR

DR. FLORENTE LÓPEZ RODRÍGUEZ
FACULTAD DE PSICOLOGÍA UMAM

DR. VÍCTOR RAMÍREZ AMAYA
INSTITUTO DE NEUROBIOLOGÍA UNAM

DRA. IRMA YOLANDA DEL RÍO PORTILLA
FACULTAD DE PSICOLOGÍA UMAM

DR. OSCAR ZAMORA ARÉVALO
FACULTAD DE PSICOLOGÍA UMAM

MÉXICO, D.F. JUNIO 2014



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Agradecimientos

Agradezco muy sinceramente a todas las personas que me acompañaron durante esta etapa.

Particularmente agradezco a mis padres, por su infinito apoyo y motivación para mi desarrollo personal y profesional.

A mis hermanos, por estar siempre pendientes de mí, por su influencia en mi forma de pensar y por ser los mejores amigos con los que he podido crecer.

Al Dr. Alvaro Torres, por tenerme toda la confianza desde el primer día que comenzamos a trabajar juntos. Por tratarme siempre como un compañero y amigo, y de esa forma convertirse en el mejor maestro. Por compartir y enriquecer una particular visión de la investigación en psicología, en la que seguiremos creyendo y trabajando. Y por su incondicional apoyo y motivación.

A todos los académicos de la Facultad de Psicología que me brindaron su apoyo, sus consejos y que han compartido conmigo sus puntos de vista. Particularmente a los doctores: Florente López Rodríguez, Oscar Zamora Arévalo, Irma Yolanda del Río Portilla, Natalia Arias Trejo, Vladimir Orduña Trujillo y Marina Menez Díaz.

A los estudiantes de la Facultad de Psicología, que confiaron en mí y han apoyado la realización de mi trabajo. Particularmente a Adriel Ruiz Rodríguez, Blanca González Machorro, Yaely Estrada García y Daniel Manzanares Chichil.

Al Dr. Gert Westermann, por aceptarme como colaborador en su laboratorio y por enriquecer mi visión del modelamiento computacional en la psicología.

A todos mis amigos. Particularmente agradezco a aquellos que han estado conmigo desde la preparatoria: Rodrigo Alonzo Muñoz, Renée Patiño García, Axel Klein Ríos, Javier Pérez Alburquerque, Juan Leyva Vargas, Karina Pérez Zamora, Alejandra Cruz Valle y Adán Bonilla Morales.

Agradezco también a aquellos que han estado conmigo desde la licenciatura, con quienes aparte de una inigualable amistad he tenido oportunidad de compartir el crecimiento académico: Ana Canseco Alba, Francisco Valencia Ruiz, Luis Ignacio Reyes Aguirre, Leonardo Lara Valderrábano y Sarahí Gallardo Pineda.

Y a aquellos a quienes tuve la oportunidad de conocer durante esta etapa y que sin ser oficialmente mis compañeros de generación se convirtieron en mi mejor compañía: Violeta Félix Romero, Mariana Gaytán Camarillo, Maricruz Vargas Ramírez y Gabriel Valadez Roque.

Finalmente agradezco a la Universidad Nacional Autónoma de México, a la Facultad de Psicología, donde fue realizado la mayor parte de este trabajo, a la Universidad de Lancaster en el Reino Unido, donde realicé una estancia de investigación y al Consejo Nacional de Ciencia y Tecnología (329684/229180), que me dio una beca doctoral y una beca mixta para apoyar mis estudios en México y en el Reino Unido.

INDICE

Resumen	3
Abstract	4
Introducción.....	5
Capítulo 1. Equivalencia de Estímulos.....	8
Estudios de EE con Carácter Aplicado.....	12
Procedimientos Experimentales.....	13
Estudios Neurofisiológicos.....	16
Explicaciones del Fenómeno de EE.....	20
Transferencia Mediada por Estímulos Nodales y Efectos de Nodalidad.....	23
Capítulo 2. Modelos Neurocomputacionales.....	27
Desarrollo del Conexionismo	28
Arquitectura y Funcionamiento de los Modelos Neurocomputacionales	31
Modelos Neurocomputacionales de EE.....	36
Justificación	39
Objetivos.....	40
Capítulo 3. Experimentos, Modelos Neurocomputacionales, Resultados y Discusiones.....	41
Experimento 1	41
Método.....	42
Resultados y Discusión.....	47
Modelo Neurocomputacional 1	50
Simulación 1	53
Resultados y Discusión.....	54
Simulación 2	58
Resultados y Discusión.....	58
Experimento 2.....	59
Método.....	59
Resultados y Discusión.....	61
Modelo Neurocomputacional 2	67
Simulación 1. Aprendizaje de AB y BC.....	76
Resultados y Discusión.....	77
Simulación 2. Sidman y Tailby (1982).....	78
Resultados y Discusión.....	80
Simulación 3. Devany, Hayes y Nelson (1986).....	82
Resultados y Discusión.....	85
Simulación 4. Spencer y Chase (1996).....	88
Resultados y Discusión.....	90
Capítulo 4. Discusión General.....	93
Referencias.....	100
Apéndice A	111
Apéndice B.....	112
Apéndice C.....	113
Apéndice D Artículo publicado. Tovar y Torres (2012).....	114

RESUMEN

Una clase o categoría de estímulos equivalentes puede estar compuesta de estímulos que no necesariamente guardan una similitud perceptual pero son funcionalmente equivalentes. Las relaciones que hay entre los estímulos en estas clases pueden ser directamente aprendidas o derivadas a partir de otras relaciones. Un cuerpo considerable de evidencia empírica ha demostrado que los humanos pueden aprender la equivalencia entre un estímulo A y uno B, y entre el estímulo B y uno C, para después derivar, sin entrenamiento o instrucciones explícitas, la equivalencia entre A y C. En esta tesis se presenta un acercamiento neurocomputacional para explorar el proceso de formación de clases de estímulos equivalentes. Los dos modelos aquí presentados lograron simular la formación de clases en distintas condiciones; se simularon procedimientos de estímulos compuestos e igualación a la muestra, y se simuló la ejecución de personas con desarrollo normal y desarrollo atípico. Los datos generados por el primer modelo permitieron analizar cómo el procedimiento de estímulos compuestos está afectado por algunos sesgos semánticos en el tipo de respuesta, se reportan dos experimentos con humanos que aportan evidencia empírica para éste modelo. El segundo modelo se diseñó con un algoritmo biológicamente plausible y permitió generar descripciones cuantitativas de la estructura de relaciones de una clase de equivalencia como resultado de historias de aprendizaje particulares. Con este modelo se simularon tres estudios clásicos del área de equivalencia de estímulos. A partir de estos modelos se pudieron discutir aspectos teóricos respecto a la formación de clases de equivalencia, generar predicciones para futuros trabajos de investigación y establecer un vínculo más entre el área de equivalencia de estímulos y el área de categorización.

ABSTRACT

Stimulus equivalence classes or categories can be composed of perceptually different but functionally equivalent stimuli. The stimulus relations in these classes can be directly learned or can be derived from previously learned relations. A considerable body of empirical evidence has shown that humans can learn the equivalences between the stimuli A and B, and B and C, for later derive the equivalence between A and C in the absence of particular instructions. In this thesis I show a neurocomputational approach to explore equivalence class formation. Different procedures and situations were modeled; simulations of learning with matching to sample and compound stimuli procedures in normal and atypical populations were performed. Results from the first model showed the existence of semantic biases in the response options used in compound stimuli procedures, two experiments with human participants were run to provide empirical evidence for this model. The second model was designed on the consideration of biological plausibility; it generated quantitative descriptions of stimulus relations and class structure as a result of specific training histories. Three classic studies on equivalence classes were successfully simulated with the second model. Both models allowed the discussion on theoretical assumptions about equivalence class formation, generated predictions for future research, and established a link between the areas of stimulus equivalence and categorization.

INTRODUCCIÓN

La categorización es un proceso psicológico básico. Permite a los organismos adaptarse a su medio, mostrar una conducta organizada y manifestar economía cognitiva (Rosch, 1978). Una categoría es “una clase de estímulos que ocasionan respuestas comunes en un contexto particular” (Zentall, Galizio & Critchfield, 2002, p. 238). El proceso de categorización ha sido estudiado desde numerosas y diversas perspectivas; psicólogos conductuales, cognitivos, neurocientíficos se han interesado en distintos aspectos de este fenómeno (e.g., Cohen & Lefebvre, 2005; Grossman et al., 2002).

Un cuerpo considerable de investigación se ha enfocado en las categorías perceptuales, donde los objetos del ambiente son clasificados de acuerdo a las propiedades físicas que comparten (Ashby & Maddox, 2005). Sin embargo, de acuerdo a la definición de categoría con la que parte esta tesis, no es necesario que los elementos de una categoría posean algún grado de similitud perceptual. La categoría *Perros*, por ejemplo, aparte de estar conformada por distintos ejemplares de perros, puede contener elementos que no mantienen una relación perceptual con un perro, pero en un contexto determinado son funcionalmente equivalentes, como la palabra “perro”, la palabra en inglés “dog”, el sonido de un ladrido, etc. Las condiciones bajo las cuales se establecen este tipo de categorías (que son relativamente insensibles a las similitudes perceptuales, Goldstone & Kersten, 2003), han sido extensamente estudiadas en el análisis experimental de la conducta bajo los nombres de Equivalencia de Estímulos (EE), o formación de clases de equivalencia (Sidman, 1994, Wilkinson & McIlvane, 2001).

Los elementos que conforman una clase de estímulos equivalentes mantienen entre sí relaciones que pudieron ser directamente aprendidas por las consecuencias que siguieron a la conducta de tratar estímulos distintos como iguales, o pudieron ser derivadas a partir de las

relaciones que hay entre otros miembros de la clase. Por ejemplo, si un humano aprende que los perros son peligrosos y despliega un conjunto de conductas en la presencia de un perro, tras aprender la equivalencia entre la imagen de un perro y el sonido de un ladrido, y la equivalencia entre el ladrido y la palabra “perro”, podrá derivar la equivalencia entre el perro y la palabra que lo designa (*si AB y BC entonces AC*) y el repertorio conductal que aprendió en la presencia del perro podrá ser desplegado también ante la palabra.

Se ha argumentado que una serie de repertorios conductuales complejos, como aquellos implicados en la formación de conceptos, el significado, la sintaxis, y en general los relacionados con la conducta simbólica, pueden ser analizados con metodologías de EE (Dickins & Dickins, 2001; Wilkinson & McIlvane, 2001). Esto es porque mucha de la conducta simbólica se puede operacionalizar en términos de relaciones derivadas entre estímulos y transferencia del control de estímulos entre los elementos de una clase o categoría.

Las relaciones estímulo-estímulo que comúnmente se estudian en experimentos de EE pretenden analizar al aprendizaje de asociaciones entre: palabras y los objetos que representan, representaciones visuales arbitrarias y objetos reales, conceptos matemáticos (cantidades) y sus representaciones pictóricas, y asociaciones entre diferentes ejemplares de una categoría, entre otras.

Un avance sustancial en el entendimiento de la categorización proviene del modelamiento computacional (ver Kruschke, 2008, para una revisión de modelos de categorización), sin embargo la mayoría de los modelos se han centrado principalmente en las categorías perceptuales (e.g., Kruschke, 1992; Love, Medin, & Gureckis, 2004), de tal suerte que no resultan adecuados para analizar EE. Por otro lado, los modelos formales de aprendizaje, principalmente en el marco del condicionamiento Pavloviano, que pueden dar cuenta del

establecimiento de asociaciones entre estímulos perceptualmente diferentes, se han enfocado, en su mayoría, en las correspondencias de únicamente dos estímulos (eventos) como resultado de su co-ocurrencia en el ambiente, de la competencia de claves, y la predicción (e.g., Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010; Rescorla & Wagner, 1972; Sutton & Barto, 1981). Existe poco trabajo de modelamiento computacional que permita acercarse al análisis de EE; el área de EE ha sido prolífica, sin embargo se ha desarrollado desligada de los avances de otras perspectivas que estudian procesos similares.

En esta tesis se estudia la formación de clases de estímulos equivalentes con el uso de modelos neurocomputacionales, estos han sido muy usados como una herramienta heurística en las ciencias cognitivas y representan una perspectiva integradora, donde se pueden analizar procesos conductuales para generar hipótesis de corte cognitivo a partir de modelos que permiten considerar los hallazgos de la neurofisiología asociada al proceso. Se pretende generar modelos que describan el proceso de formación de clases de equivalencia y por ende puedan encaminar la generación de predicciones respecto a la formación de estas categorías. La simulación en modelos neurocomputacionales ayudará a integrar el estudio de EE con áreas afines, permitiendo enriquecer distintas perspectivas; al mismo tiempo, el modelamiento neurocomputacional permitirá darle mayor generalidad al fenómeno de EE.

Los capítulos de esta tesis están estructurados de la siguiente manera: En el Capítulo 1 se presenta el área de EE. En el segundo capítulo se revisa la propuesta de investigación de procesos psicológicos desde el modelamiento neurocomputacional y se revisan los modelos que se han desarrollado para estudiar EE. En el Capítulo 3 se presenta la metodología, los resultados y discusiones de los experimentos y modelos generados. En el Capítulo 4 se discuten de manera general los resultados, alcances y limitaciones de esta tesis.

CAPÍTULO 1. EQUIVALENCIA DE ESTÍMULOS

En el análisis experimental del comportamiento, el estudio de Equivalencia de Estímulos (EE) comenzó con los trabajos de Murray Sidman en la década de 1970. En un artículo publicado en 1971 Sidman describe a un niño que sabía igualar dibujos con sus correspondientes palabras dictadas, pero era incapaz de igualar correctamente cuando los nombres de las palabras no eran dictados sino presentados visualmente (i.e., palabras impresas). Tras ser entrenado con procedimientos operantes a igualar palabras impresas con sus correspondientes palabras dictadas, el niño fue capaz de realizar nuevas igualaciones que no fueron entrenadas; pudo igualar palabras impresas con sus correspondientes dibujos, y pudo nombrar en voz alta las palabras impresas. El repertorio conductual final del niño estaba altamente relacionado con la conducta de lectura y sugería que todas las modalidades de los estímulos (i.e., palabras dictadas, impresas y dibujos) eran equivalentes.

Sidman continuó su programa de investigación centrándose en dotar de habilidades conductuales a sujetos con retardo mental, a la vez que analizó la emergencia de relaciones condicionales entre estímulos en los repertorios de sus participantes experimentales (Sidman & Cresson, 1973; Sidman, Cresson, & Willson-Morris, 1974).

Varios estímulos pueden ser considerados como equivalentes si las relaciones entre ellos muestran las tres propiedades definitorias de equivalencia matemática: reflexividad, simetría y transitividad (Sidman & Tailby, 1982). Estas propiedades pueden ser evaluadas con procedimientos de igualación a la muestra (IM). En un procedimiento de IM, una persona puede aprender a seleccionar dentro de varios estímulos de comparación ($B_1, B_2, B_3, \dots B_n$) aquel que correctamente iguala a un estímulo muestra (A_1); por ejemplo, supóngase que los estímulos de

comparación sean imágenes de distintos animales y el estímulo muestra sea el nombre hablado de un animal en particular; un niño aprenderá la igualación entre A1 y B1, si la selección del estímulo de comparación B1 en la presencia del estímulo muestra A1 es seguida de reforzamiento (aprobación, halagos, dulces); una vez que el niño ha aprendido la igualación entre A1 y B1 y entre B1 y C1, las propiedades definitorias de la relación de equivalencia pueden ser documentadas con ensayos de IM si en ausencia de instrucciones y reforzamiento, el niño selecciona A1 como comparación en presencia de A1 (reflexividad), selecciona A1 en presencia de B1, y B1 en presencia de C1 (simetría) y selecciona C1 en presencia de A1 (transitividad).

La simetría y la transitividad pueden ser combinadas en un ensayo denominado como “prueba de equivalencia” (Sidman, 1992), para el ejemplo actual el logro de la prueba de equivalencia consiste en seleccionar A1 en presencia de C1. La Tabla 1 muestra un ejemplo de la nomenclatura del tipo de relación y tipo de ensayo para un procedimiento típico de IM dentro de EE.

Tabla 1. Ejemplo de la nomenclatura de las relaciones entrenadas y evaluadas en un procedimiento experimental común de equivalencia de estímulos.

Relación	Tipo de Relación	Tipo de Ensayo
AB	Entrenada	Línea base
BC	Entrenada	Línea base
AA	Emergente	Prueba de reflexividad
BA	Emergente	Prueba de simetría
CB	Emergente	Prueba de simetría
AC	Emergente	Prueba de transitividad
CA	Emergente	Prueba de equivalencia

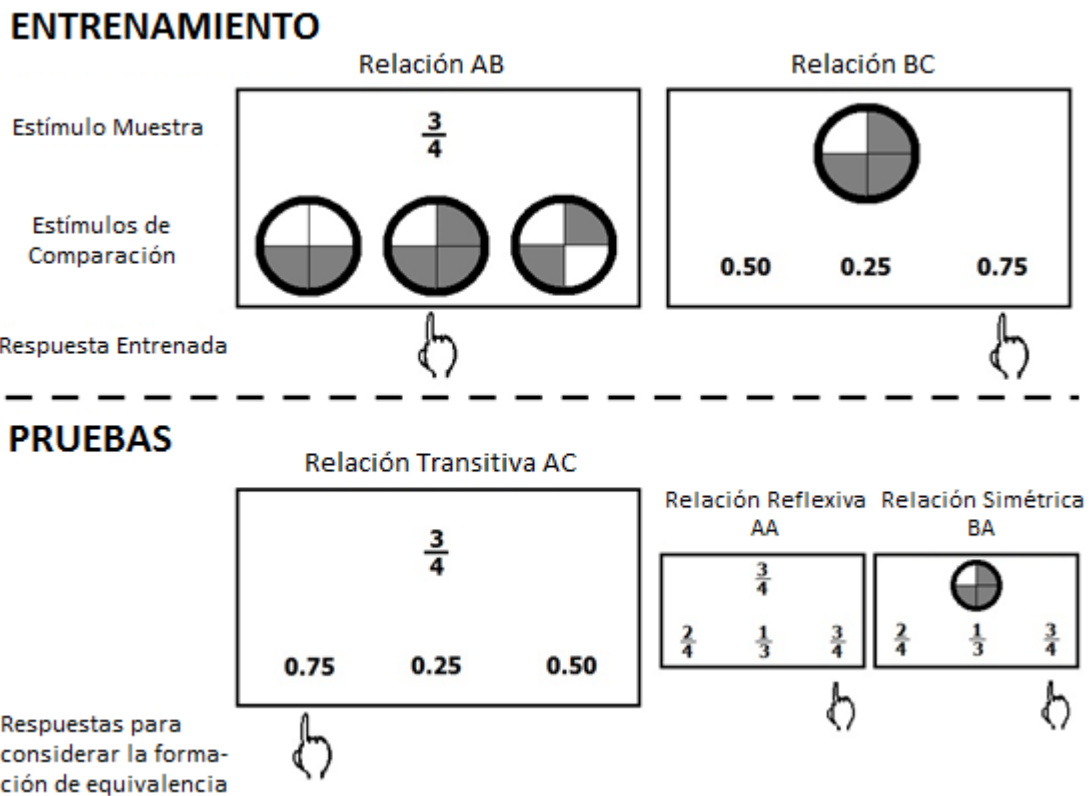
La relación entre los estímulos A1 y B1 se escribe A1B1, esta nomenclatura considera que el primer estímulo es el de muestra y el segundo el de comparación, las letras hacen referencia a estímulos distintos, los números hacen referencia a clases distintas.

Típicamente un estudio de EE está dividido en dos fases, la fase de entrenamiento y la fase de prueba. Durante el entrenamiento los participantes aprenden las llamadas *relaciones de línea base*, durante la fase de prueba se evalúan las ejecuciones ante ensayos de *relaciones emergentes*; aquellas que pueden ser derivadas a partir de las relaciones entrenadas. En la Figura 1 se muestra un ejemplo de las relaciones de línea base entrenadas y las relaciones emergentes derivadas, en un formato de IM, la imagen está inspirada en el estudio de Lynch y Cuvo (1995).

A partir de las respuestas emitidas por los participantes en pruebas de transitividad y equivalencia, se puede inferir la aparición de conducta novedosa funcional que no fue directamente entrenada (i.e., conducta emergente); por ejemplo, se puede entrenar a un participante ingenuo en la clase de *Frutas*, considerando como estímulo A la palabra “frutas”, como estímulo B una naranja y como estímulo C un plátano, tras enseñar a relacionar A con B y B con C el sujeto debería nombrar a los plátanos como frutas (C con A), aunque no haya aprendido eso directamente. Igualmente, se espera que las características que se aprendan para algún miembro de la clase *Frutas* sean asociadas con el resto de los miembros, es decir, se genere una transferencia de funciones entre los miembros de la clase; por ejemplo, si damos a probar una naranja a un participante y la asocia con un sabor dulce (B con D, considerando que el sabor dulce actúe como estímulo D dentro de la clase), esperaríamos que se formen también las relaciones AD, DA, DB, CD, y DC; es decir, el participante contará con 6 relaciones estímulo-estímulo nuevas dentro de su repertorio conductual, tras haber experimentado directamente solo

una asociación nueva. En una situación de prueba, el participante debería responder que el plátano tiene un sabor dulce aunque nunca lo haya probado.

Figura 1. Muestra un ejemplo de relaciones-estímulo entrenadas y derivadas en un formato de igualación a la muestra. Los estímulos A, B y C corresponden a diferentes representaciones de la misma cantidad. Sólo las relaciones AB y BC son entrenadas. AC, AA y BA se muestran como ejemplos de relaciones derivadas.



El ejemplo anterior ilustra cómo una situación de aprendizaje cotidiano se puede operacionalizar en el paradigma de EE, y cómo la inclusión o modificación de las características de un estímulo en una clase da como resultado un cambio en el control de estímulos de todos los componentes de la clase. Es decir, se altera el significado de una clase o concepto. El análisis de

las propiedades discriminativas de los estímulos, que pueden ser transferidas a otros miembros de las clases de equivalencia, y el estudio del establecimiento de relaciones transitivas, se han usado como modelos experimentales de la conducta simbólica humana (Dickins & Dickins, 2001; Wilkinson & McIlvane, 2001). Por ejemplo, se han generado propuestas de cómo una sintaxis correcta puede ser aplicada en el lenguaje dependiendo de las clases a las que pertenece cada palabra y del tipo de relaciones que mantiene con otras clases (Wulfert & Hayes, 1988). Se ha analizado también el grado de adquisición de significados en estímulos originalmente neutros mediante el entrenamiento de asociaciones entre expresiones emocionales y símbolos arbitrarios (Bortoloti & De Rose, 2009).

Los estímulos que conforman una clase de equivalencia pueden presentarse en cualquier modalidad sensorial; por ejemplo, se ha reportado equivalencia en clases que incluyen estímulos auditivos y visuales (e.g., Saunders, Wachter, & Spradlin, 1988; Sidman, 1971, Sidman, & Cresson, 1973), estímulos únicamente visuales (e.g., Devany, Hayes & Nelson, 1986; Mackay, Wilkinson, Farrel, & Serna, 2011; Spencer & Chase, 1996), estímulos únicamente auditivos (Dube, Green, & Serna, 1993), estímulos táctiles (e.g., O'Leary & Bush, 1996), estímulos interoceptivos (DeGrandpre, Bickel, Higgins, 1992), y estímulos gustativos (Hayes, Tilley, & Hayes, 1988).

Estudios de EE con Carácter Aplicado

La enseñanza de repertorios conductuales basados en procedimientos como el reforzamiento condicionado, la discriminación condicional y la IM, tiene una amplia historia en la psicología (Cooper, Heron, & Heward, 1987; Kerr, Meyerson, & Flora, 1977; Rehfeldt & Barnes-Holmes, 2009). El modelo de EE, implementado en este tipo de procedimientos, ha

enriquecido el área de evaluación y modificación conductual generando tecnologías aplicables a una amplia variedad de problemas de aprendizaje. Se ha trabajado tanto con personas que tienen un desarrollo normal, como con personas que presentan problemas de aprendizaje debidos a factores biológicos o sociales (Carr, Wilkinson, Blackman, & McIlvane, 2000; Eikeseth & Smith, 1992; Fiorentini, Arismendi, & Yorio, 2012; Hall, DeBernardis & Reiss, 2006; LeBlanc, Miguel, Cummings, Goldsmith, & Carr, 2003; O'Donnell & Saunders, 2003; Rehfeldt & Barnes-Holmes, 2009; Saunders, O'Donnell, Vaidya, & Williams, 2003; Sidman, 1971; Stromer, Mackay, McVay, & Fowler, 1998).

Rehfeldt (2011) notó en una revisión de la literatura que reporta procedimientos de EE de carácter aplicado, que cerca de la mitad de los participantes que fueron reportados entre 1992 y 2009 presentaba algún desorden de desarrollo, incluyendo lesión cerebral, síndrome de Down, discapacidad intelectual y autismo.

Se ha estudiado la formación de equivalencias en personas que viven en ambientes empobrecidos o con problemas de aprendizaje en situaciones de educación básica (De Rose, De Souza, & Hanna, 1996; Lynch & Cuvo, 1995, Matos & Hübner-D'Oliveira, 1992), o con personas que presentan un desarrollo normal, para evaluar los procedimientos de EE como una estrategia de aprendizaje alternativa (Fields, Travis, et al., 2009; Fienup, Covey & Critchfield, 2010; Fienup, Critchfield, & Covey, 2009; Miguel, Petursdottir, Carr, Michael, 2008).

Procedimientos Experimentales

La definición de una clase de estímulos equivalentes no especifica el uso de un procedimiento en particular para su formación (Fields, Reeve, Varelas, Rosen, & Belanich, 1997; Sidman, 1994; Sidman & Tailby, 1982); sin embargo, el procedimiento más común que se ha

utilizado para estudiar y/o establecer clases de estímulos equivalentes ha sido el de IM. Como se mencionó previamente, este procedimiento consiste en presentar al participante un arreglo donde aparece un estímulo de muestra y dos o más estímulos de comparación; el participante aprende, vía reforzamiento, a seleccionar el estímulo de comparación positivo ante un estímulo muestra determinado.

Se han utilizado dos variantes principales del procedimiento de IM; en una de ellas, la presentación del estímulo muestra está separada temporalmente de la presentación de los estímulos de comparación: igualación a la muestra demorada (e.g., Bentall, Dickins, & Fox, 1993; Stromer & Mackay, 1992). Dentro de esta variante, se han utilizado arreglos más complejos para estudiar formación de clases donde todos los estímulos son de carácter auditivo, de tal forma que la presentación de los mismos se va haciendo de forma secuencial para evitar confusión por solapamiento (Dube, Green, & Serna, 1993). En la otra variante, la aparición de los estímulos de comparación se da sin eliminar al estímulo muestra del campo perceptual del participante, este es el procedimiento que ha tenido mayor uso dentro de EE.

Se pueden derivar relaciones de equivalencia entre estímulos cuando los participantes aprenden relaciones de línea base con procedimientos de pares asociados, que incluyen respuestas de tipo *go/no-go*, o respuestas de tipo *yes/no*.

Debert y colaboradores reportaron formación de clases estímulo cuando solicitaron a participantes humanos emitir respuestas de tipo *go/no-go* ante distintos pares de estímulos (Debert, Huziwar, Faggiani, Simões de Mathis, & McIlvane, 2009; Debert, Matos & McIlvane, 2007). En sus procedimientos, los participantes aprendían a responder (*go*) en presencia de pares intra-clase; formados por estímulos que pertenecían a la misma clase, por ejemplo, A1B1 y B1C1. Y enseñaron también a no responder (*no-go*) cuando les mostraban pares cros-clase;

formados por estímulos que pertenecían a clases distintas, como A1B2 y B2C1. Tras este entrenamiento los participantes pudieron responder de forma correcta a nuevas configuraciones de pares de estímulos de acuerdo a la membresía de clase; por ejemplo, respondían ante los pares A1C1 y A2C2; y no respondían ante los pares A1C2 y A2C1 (Debert et al., 2007).

De forma similar, Fields, Doran y Marroquin (2009) reportaron formación de clases de equivalencia cuando utilizaron un procedimiento de pares de estímulos con respuestas tipo *yes/no*. En el entrenamiento utilizado, los participantes debían responder con la opción *yes* en presencia de pares intra-clase y debían responder con la opción *no* en la presencia de los pares cros-clase. La diferencia fundamental entre el procedimiento *go/no-go* y el *yes/no* es que con respuestas *yes/no* los participantes tienen la oportunidad de emitir rechazos explícitos en la presencia de pares cros-clase, en vez de simplemente retenerse de responder; y serán retroalimentados por su ejecución tras cada respuesta, a diferencia de los ensayos *no-go* en los que no hay reforzamiento.

Los procedimientos de pares de estímulos y respuestas de tipo *yes/no* o *go/no-go* han resultado una alternativa útil para el establecimiento y prueba de EE; por ejemplo, han sido útiles en estudios que buscan el correlato neurofisiológico asociado al responder ante una clase de estímulos equivalentes (Haimson, Wilkinson, Rosenquist, Ouimet, & McIlvane, 2009; Schlund, Cataldo, & Hoehn- Saric, 2008). En estas circunstancias un procedimiento de IM dificulta la interpretación de los datos, dado que en cada ensayo se presentan al sujeto estímulos de comparación pertenecientes a clases distintas; con lo que se complica el análisis de actividad asociada a eventos, pues cada evento resulta ambiguo. En cambio, un procedimiento de estímulos en pares y respuestas tipo *yes/no*, presenta en cada ensayo uno o dos estímulos que son de la misma clase o de clases distintas; así, los datos obtenidos por técnicas de electroencefalografía

(Haimson et al., 2009) o por resonancia magnética funcional (Schlund et al., 2008) han resultado más adecuados para ser interpretados en relación a una respuesta de categorización.

Los procedimientos mencionados hasta ahora generan aprendizaje basado en consecuencias donde el reforzamiento contingente con las respuestas de los participantes es responsable del establecimiento de las relaciones de línea base. Sin embargo, se ha argumentado que ni el reforzamiento, ni las respuestas son esenciales para que emerja la equivalencia; por ejemplo, Leader, Barnes y Smeets (1996) reportaron ejecuciones consistentes con formación de clases tras presentar a un grupo de participantes humanos con estímulos en secuencia que formaban pares, sin requerir ningún tipo de respuesta explícita. Tras la observación de las secuencias de estímulos relacionados como A1B1 y B1C1, los participantes respondieron correctamente en pruebas de simetría y equivalencia presentadas en un formato de IM (ver también Clayton & Hayes, 2004). Este tipo de experimentos resultan relevantes pues demuestran que la formación de asociaciones entre estímulos que no guardan similitud perceptual se puede derivar dependiendo de las regularidades ambientales con las que se presentan los estímulos y no serán dependientes necesariamente del despliegue y corrección de una conducta.

Estudios Neurofisiológicos

Los datos que se obtienen y analizan en la mayoría de los estudios de equivalencia son de exactitud y latencia de respuestas, a partir de éstos se hacen inferencias respecto a las condiciones bajo las cuales se aprende una clase de estímulos equivalentes y los procedimientos de entrenamiento que resultan más adecuados para establecer dichas clases; de igual forma, estos datos pueden ayudar a respaldar o cuestionar hipótesis teóricas sobre la formación de clases de equivalencia. Sin embargo, hacia los últimos años, en algunos estudios se han obtenido datos de

carácter neurofisiológico que han ayudado a esclarecer parte de la naturaleza o características de las clases de equivalencia. A continuación se mencionan los principales datos obtenidos en estos estudios.

Dickins y colaboradores (2001) analizaron los datos de resonancia magnética funcional (fMRI) de 11 participantes que realizaron tareas de IM para establecer y probar 6 clases de estímulos icónicos. Con el supuesto de que existe una relación entre EE y el lenguaje, Dickins y colaboradores compararon las activaciones generadas durante la resolución de tareas de equivalencia con activaciones generadas por tareas de fluidez verbal; sus resultados indicaron que los participantes que formaron las clases de equivalencia mostraron una activación lateralizada izquierda de la corteza prefrontal dorsolateral (DLPFC), tanto en las tareas de fluidez verbal como en las de equivalencia a diferencia de los sujetos que no formaron las clases, que mostraron una activación de la DLPFC menos lateralizada. Por otra parte, no encontraron activación del área de Broca durante las tareas de EE a diferencia de las tareas de fluidez verbal, lo cual no respalda la visión de que existe una mediación verbal abierta o cubierta para formar clases de equivalencia.

En cuanto a estructuras subcorticales, encontraron activación del núcleo caudado tanto para las relaciones entrenadas como para las derivadas, así como activación del tálamo para relaciones entrenadas y de equivalencia. Las activaciones encontradas en DLPFC y en tálamo resultan consistentes con los hallazgos reportados en otro tipo de estudios sobre categorización de estímulos (Grossman et al., 2002).

Los datos de fMRI del estudio de Dickins y colaboradores (2001) fueron analizados considerando bloques de 15 segundos en ensayos de IM, lo cual resulta problemático para la interpretación de la actividad cerebral asociada únicamente al aprendizaje de las clases, pues las

imágenes cerebrales obtenidas estuvieron relacionadas también con la búsqueda visual de estímulos de comparación, la elección de respuesta, y el intervalo entre ensayos. Ese y otros problemas del análisis fueron atendidos en un estudio de Schlund y colaboradores (2007) quienes entrenaron 2 clases compuestas de 3 estímulos que fueron caracteres ASCII en 12 participantes y analizaron las activaciones cerebrales generadas con fMRI considerando activaciones relacionadas a eventos, y comparaciones entre relaciones entrenadas contra relaciones derivadas. De forma general sus datos sugieren que el responder condicional tanto para relaciones entrenadas como derivadas es similar, habiendo una magnitud de activación mayor en las zonas frontales, principalmente de las áreas DLPFC y ventrolateral cuando las relaciones son entrenadas. En cuanto a las áreas subcorticales, encontraron que las activaciones son similares para relaciones entrenadas y derivadas con excepción de que las relaciones entrenadas muestran una activación mayor del núcleo caudado y del tálamo.

Schlund y colaboradores (2007) no encontraron un patrón de lateralización en el hemisferio izquierdo, sino una activación predominante del hemisferio derecho; lo que les hizo suponer que la resolución de tareas de equivalencia no está mediada por procesos verbales, aunque puede que esta diferencia en relación al estudio de Dickins y colaboradores (2001) se deba al tipo de estímulos utilizados. Una interpretación de dicha activación con predominancia derecha tiene que ver con que se trata de una tarea de ordenamiento serial de los estímulos; el entrenamiento utilizado estableció las relaciones AB y BC, generando una estructura lineal A-B-C que puede ser mejor representada por el hemisferio derecho, el responder consistente con la clase pudo reclutar circuitos de regiones relacionadas a procesos de memoria espacial.

En un siguiente estudio, Schlund y colaboradores (2008) analizaron la activación del hipocampo durante la resolución de tareas de equivalencia. Establecieron 2 clases de estímulos

equivalentes vía IM y analizaron con fMRI la activación generada durante las pruebas con un procedimiento de pares de estímulos re combinados (e.g., simetría B1A1, transitividad A1C1 y equivalencia C1A1) y respuestas de tipo *yes/no*. Los datos encontrados subrayan la activación del hipocampo durante el responder condicional a relaciones derivadas mediadas por estímulos nodales; es decir, hubo una mayor activación del hipocampo anterior en ensayos de transitividad y equivalencia, en comparación con ensayos de simetría o de pares de estímulos no relacionados.

Los datos de Schlund y colaboradores (2008) replican los hallazgos reportados en tareas de inferencia transitiva, entendida como la habilidad para inferir la relación entre estímulos que no han sido experimentados juntos, a partir de la relación que tienen con otro estímulo (Acuna, Eliassen, Donoghue, & Sanes, 2002). Las tareas de equivalencia pueden ser vistas como un caso de inferencia transitiva (Dickins, 2005). Heckers y colaboradores (2004) reportaron activación de la región anterior del hipocampo de participantes que respondieron a tareas de relaciones derivadas entre estímulos del tipo “mayor que” ($A > B > C > D > E$).

Por otra parte, estudios de equivalencia y electroencefalografía han reportado la presencia de componentes N400 durante la resolución de tareas para estímulos no equivalentes (Barnes-Holmes, Staunton et al., 2005; Haimson et al., 2009). El componente N400 se ha reportado consistentemente en estudios de corte cognitivo cuando se presentan palabras –o estímulos con algún significado asociado- que no tienen una relación semántica entre si, por ejemplo cuando se presentan la palabra “plátano” seguida de la palabra “tía” en comparación a cuando se presenta “sobrina” seguida de “tía”. El hallazgo de N400 en estudios de EE da apoyo a los argumentos de que las relaciones de equivalencia proveen un modelo útil para establecer y estudiar experimentalmente relaciones semánticas.

Explicaciones del Fenómeno de EE

La evidencia que se ha generado a lo largo de más de 40 años que lleva el estudio de EE, ha dado como resultado tres principales posturas teóricas de la formación de clases de equivalencia, particularmente respecto a la emergencia de relaciones entre estímulos que no fueron directamente entrenadas sino derivadas; la propuesta de formación de clases vía reforzamiento de Sidman (1990; 1994; 2000), la de nombramiento de Lowe (Dugdale & Lowe, 1990; Horne & Lowe, 1996), y la de Marcos Relacionales de Hayes (Barnes-Holmes, Rodríguez, & Whelan, 2005; Hayes, Barnes-Holmes, & Roche, 2002).

Sidman (1990; 1994; 1997; 2000) fue formalizando sus ideas para dar pie a un modelo descriptivo de EE. En su trabajo de 1990, Sidman presenta su postura respecto al origen de las relaciones equivalentes, analiza primero la posibilidad de que las relaciones emergentes se puedan derivar por un proceso lógico o por mediación de la conducta verbal. Sin embargo, Sidman argumenta que no existe una necesidad lógica para que las relaciones sean reflexivas, simétricas y transitivas; incluso, aventura la idea de que la equivalencia podría ser considerada como base para el razonamiento inductivo y no al revés. En cuanto al papel de la conducta verbal Sidman se centra en el nombramiento y las reglas; el nombramiento implica que dado que se puede aplicar un mismo nombre o una respuesta común para distintos estímulos, estos se pueden relacionar en una clase de equivalencia (ver más adelante propuesta de nombramiento); Sidman, nuevamente argumenta que la formación de equivalencias no requiere de un proceso de nombramiento, ni de ser establecida vía reglas, sino que el nombramiento y la generación de reglas pueden ser resultado de la capacidad de formación de equivalencias y no al contrario (Sidman, 1990; 1992).

Sidman llega a la conclusión de que la equivalencia es una función primitiva y que, al igual que los procesos como la discriminación, el reforzamiento condicionado y las discriminaciones condicionales, es un producto del reforzamiento (Sidman, 2000) y considera que la capacidad de formar equivalencias debe relacionarse a procesos adaptativos.

Por otra parte, la propuesta de nombramiento (Dugdale & Lowe, 1990; Horne & Lowe, 1996) considera que la capacidad para formar equivalencias se debe a que los humanos poseen un repertorio de conductas verbales más elaborado que el resto de los animales. Dugdale y Lowe (1990) proponen que el nombramiento puede estar en la base de la formación de equivalencias, su idea es que los estímulos pasan a formar parte de una misma clase debido a que un sujeto les aplica un nombre común; o generar una regla verbal que permite relacionar a todos los miembros de una clase. Parte de la justificación de esta postura se da por el hecho de que el fenómeno de equivalencia no se ha demostrado claramente en animales no humanos (pero ver Schusterman & Kastak, 1998, para una excepción con leones marinos) y que se ha reportado una importante relación entre las ejecuciones exitosas en tareas de equivalencia y los repertorios de lenguaje bien establecidos (Devany, Hayes & Nelson, 1986).

Sin embargo, algunos estudios recientes contradicen esta postura mostrando evidencia de la formación de equivalencias en sujetos con retraso en el desarrollo que no poseen un lenguaje bien establecido (O'Donnell & Saunders, 2003). Igualmente, datos provenientes de investigaciones apoyadas en técnicas de neuroimagen cuestionan la relevancia de un proceso de mediación por nombramiento en la resolución de tareas de equivalencia (Dickins et al., 2001; Schlund, Hoehn-Saric, & Cataldo, 2007).

La propuesta de Marcos Relacionales de Hayes y colaboradores (Hayes, 1991; Hayes et al., 2002) mantiene que las respuestas correctas a relaciones derivadas en pruebas de

equivalencia, resultan de un aprendizaje previo de relaciones bidireccionales entre distintos conjuntos de estímulos. Bajo esta perspectiva, la experiencia con diferentes clases de equivalencia, y las consecuencias aprendidas por tratar a diferentes estímulos como equivalentes, se generalizan hacia nuevas clases estímulo, como aquellas utilizadas en situaciones experimentales. Por ejemplo, si una persona aprendió de forma natural a relacionar todos los miembros de la clase ABC-1, es decir, aprendió las relaciones A1B1, B1C1, y A1C1. Es altamente probable que en una situación de laboratorio, tras aprender a relacionar A2 con B2, y B2 con C2, derive correctamente la relación entre A2 y C2, dado que generalizó el conocimiento de la clase ABC-1.

Parece difícil encontrar un conjunto de experimentos que apoyen o rechacen completamente alguna de las tres propuestas teóricas mencionadas, éstas no parecen ser completamente diferenciables en situaciones de laboratorio. Por ejemplo, en un estudio clásico donde se analizó la relación entre lenguaje y formación de equivalencias, Devany y colaboradores (1986) analizaron las ejecuciones de tres grupos de niños, el primero compuesto de niños con desarrollo normal, el segundo compuesto por niños con deficiencias en el desarrollo y algunas habilidades de lenguaje, y el tercero por niños con deficiencias en el desarrollo y sin habilidades de lenguaje. Los resultados mostraron que todos los niños con habilidades de lenguaje (grupos 1 y 2) respondieron correctamente a pruebas de transitividad, mientras que los niños sin habilidades de lenguaje no lograron responder correctamente a las pruebas de transitividad. Estos resultados parecerían apoyar la teoría de nombramiento de Lowe, sin embargo, no es posible determinar si la ausencia de lenguaje causó una ejecución deficiente en equivalencia, o los problemas para desarrollar habilidades de lenguaje y responder a tareas de equivalencia tienen una causa común subyacente.

En otro estudio, Luciano y colaboradores (2007) encontraron respuestas correctas de equivalencia en una niña de 19 meses de edad con desarrollo normal, aunque con un repertorio verbal mínimo. La niña fue entrenada con un protocolo de “ejemplares múltiples” consistente en aprender diversas relaciones bidireccionales entre estímulos, por ejemplo aprender las relaciones AB y BA, tras este entrenamiento la niña pasó al entrenamiento y prueba de relaciones de equivalencia. Los autores sugirieron que las respuestas de equivalencia fueron influidas por el entrenamiento de ejemplares múltiples, lo cual parece apoyar la teoría de Marcos Relacionales. Sin embargo en éstas condiciones no se puede descartar como explicación la teoría de formación de clases por reforzamiento de Sidman.

Transferencia Mediada por Estímulos Nodales y Efectos de Nodalidad

Una misma clase de equivalencia puede mantener una “estructura” distinta dependiendo de cómo haya sido aprendida, por ejemplo, si se contempla una clase con 4 miembros: A, B, C y D, su estructura puede ser considerada como lineal o serial si las relaciones directamente aprendidas fueron AB, BC y CD; en cambio puede tener una estructura de *cluster* o racimo si las relaciones entrenadas fueron AB, AC y AD, estas estructuras están representadas en la Figura 2.

Fields y Verhave (1987) describieron 4 parámetros que se pueden analizar con el fin de entender la estructura de una clase de estímulos equivalentes: 1. El tamaño de la clase, 2. La cantidad de estímulos nodales, 3. El arreglo de estímulos nodales con estímulos sencillos y 4. La direccionalidad del entrenamiento. Estos autores denominaron a un estímulo nodal como aquel que durante el entrenamiento para la formación de una clase, estuvo relacionado con 2 o más estímulos; y denominaron a los estímulos sencillos como aquellos que durante el entrenamiento

para la formación de una clase estuvieron relacionados únicamente con otro estímulo; en la Figura 2 se indican cuáles son los estímulos nodales y los sencillos.

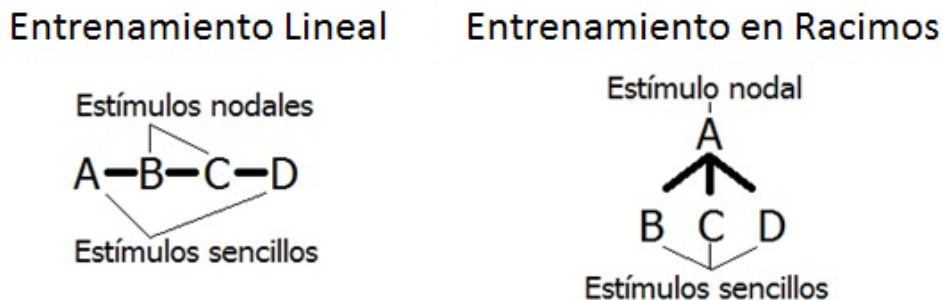


Figura 2. Ejemplo de dos estructuras de entrenamiento para la clase ABCD, se muestran los estímulos nodales y sencillos para el tipo de entrenamiento lineal y en racimos.

Conforme el tamaño de la clase aumenta (parámetro 1), se incrementa la cantidad de estímulos nodales posibles (parámetro 2), y la cantidad de posibles arreglos o estructuras entre estímulos nodales y sencillos (parámetro 3).

Se dice que hay efectos de nodalidad, cuando la estructura del entrenamiento genera una diferencia en el grado de relación que hay entre los miembros de una clase. Empíricamente se han demostrado efectos de nodalidad (Arntzen 2010; Bortoloti & de Rose, 2009; Fields, Adams, Verhave & Newman, 1993; Fields, Landon-Jimenez, Buffington & Adams, 1995; Kennedy, 1991; Kennedy, Itkonen & Lindquist, 1994, Moss-Lourenco & Fields, 2011; Spencer & Chase, 1996) consistentes en que mientras mayor es el número de nodos entre dos estímulos evaluados en una prueba de transitividad, mayor es el tiempo de respuesta a dicha prueba (Arntzen 2010; Bentall, Dickins & Fox, 1993; Fields, et al., 1995; Wulfert & Hayes, 1988).

Otra forma de medir efectos de nodalidad ha considerado el momento de aparición de las relaciones emergentes. Aquellas relaciones donde intervienen menos nodos presentan una mayor proporción de aciertos en los primeros bloques de prueba que las relaciones con mayor número

de nodos, que generan una mayor cantidad de errores durante los primeros bloques de prueba (Fields, et al., 1990; Kennedy, 1991; Kennedy, et al., 1994). Es decir, en general se puede determinar la presencia de efectos de nodalidad tras analizar las dos variables principales de respuesta observables en experimentos de EE: latencia y exactitud (Dymond & Rehfeldt, 2001).

Finalmente, se han descrito efectos de nodalidad utilizando evaluaciones de diferenciales semánticos, donde la cantidad de nodos está relacionada con el carácter semántico de un estímulo inicialmente neutro: a mayor cantidad de nodos, el estímulo adquiere un menor carácter semántico y viceversa (Bortoloti & de Rose, 2009).

Los datos reportados en estudios de nodalidad han intervenido en los debates respecto a la naturaleza de la formación de clases. La diferencia en el nivel de relación entre estímulos de una clase dependiente de la cantidad de nodos implica que la transitividad es un proceso mediado por los estímulos de la clase. Se ha debatido si los efectos de nodalidad son fundamentales para entender la estructura de las clases de equivalencia, el control de estímulos en una clase y el procesamiento que se tiene de éstos, o si más bien se dan como resultado de procedimientos metodológicos que los promueven (Sidman, 1994; 2000; Fieds et al., 1993; Fields et al., 1995; Imam, 2006).

El análisis de efectos de nodalidad ha generado vínculos entre el área de EE y el estudio del almacenamiento y procesamiento semántico (ver Fields, Adams, Verhave, & Newman, 1990). El hallazgo de efectos de nodalidad, particularmente el incremento en el tiempo de respuesta para relaciones-estímulo más distantes, resulta similar al incremento en tiempos de respuesta en recuperación de conceptos o características lejanas en una red semántica (Collins & Loftus, 1975). Resulta interesante que desde la perspectiva analítico conductual se puede entender entonces cómo se da la formación de redes de memoria semántica con elementos que

tienen mayor o menor lejanía, dependiendo de cómo se han aprendido las relaciones entre los elementos que conforman a dicha red.

CAPÍTULO 2. MODELOS NEUROCOMPUTACIONALES

Hay varias cosas que pueden ser consideradas como un modelo en ciencia: una descripción verbal, una representación gráfica o numérica, una formulación matemática, etc. Los modelos permiten estudiar y hablar de fenómenos en una forma simple y transparente, y se consideran exitosos en la medida en la que permiten alcanzar dos objetivos de todo desarrollo científico: descripción y predicción.

En muchos casos no se pretende que un modelo abarque la descripción de un fenómeno por completo; la simpleza que subyace a su desarrollo implica considerar sólo un conjunto de variables en la determinación de un fenómeno. Los modelos también sirven como una herramienta heurística, una tendencia dentro de las ciencias cognitivas es utilizar modelos como medios que permiten explorar las ideas que un investigador tiene sobre la naturaleza de un fenómeno (McClelland, 2009). Estas exploraciones pueden conducir a futuras investigaciones descriptivas o correlacionales mejor encaminadas.

Existen innumerables tipos de modelos en psicología, estos pueden ser clasificados como computacionales, matemáticos o conceptuales-verbales (Sun, 2008). Los computacionales son utilizados para modelar procesos mediante descripciones algorítmicas, los matemáticos modelan relaciones entre variables, y los modelos verbales describen variables, relaciones y procesos en lenguajes naturales informales.

Los modelos neuronales son predominantemente computacionales y en su mayoría están desarrollados dentro de una perspectiva asociacionista. Se componen de unidades de procesamiento denominadas neuronas, y de las interacciones entre ellas denominadas conexiones. Como característica principal su diseño y funcionamiento están inspirados en la

forma en la que el cerebro computa información; también son llamados modelos de redes neuronales artificiales, y la filosofía que primordialmente subyace su implementación es el conexionismo.

Los cómputos que llevan a cabo estos modelos implican cierta abstracción del funcionamiento cerebral, y se procura que los procesos que realizan sean biológicamente plausibles. Sin embargo, los objetivos de los modelos neurocomputacionales en psicología no están encaminados a describir o simular procesos puramente neurofisiológicos, en la mayoría de los modelos los objetivos tampoco se ven amenazados si se viola alguna parte del proceso biológico (en los últimos años se han dado desarrollos mejor apegados a la neurofisiología, e.g., O'Reilly & Munakata, 2000; Rolls, 2008). Los objetivos se centran en explorar y describir fenómenos cognitivos y conductuales como aprendizaje, atención, adquisición de lenguaje, desarrollo cognitivo, categorización y memoria entre otros (ver Houghton, 2005). Una visión relevante para esta tesis es que se trata de modelos de conducta, o modelos generales de aprendizaje, donde las neuronas y las conexiones artificiales, pese a su inspiración biológica, representan una abstracción de los procesos básicos y los canales de comunicación, respectivamente (Kohonen, 2001).

A continuación se presenta una revisión de los principales acontecimientos históricos que dieron lugar a la visión del conexionismo, después se mencionan las principales características de los modelos neurocomputacionales.

Desarrollo del Conexionismo

Se pueden encontrar diversas raíces de la correspondencia entre la dinámica de las redes neuronales y la dinámica de los procesos psicológicos (ver Spitzer, 1998, para una revisión).

Autores clásicos pertenecientes a distintas corrientes como William James, Sigmund Freud, y Edward Thorndike tuvieron aportes en este sentido, sin embargo, aquí considero dos eventos indispensables que dieron pie al desarrollo del conexionismo como es entendido actualmente: 1. El trabajo publicado por McCulloch y Pitts en 1943, donde presentaron una caracterización de la neurona como una unidad de procesamiento de información en términos de lógica proposicional, cuya concepción facilitó el poder simular la actividad de los grupos neuronales con ecuaciones simples, naciendo así la neurona artificial; y 2. La concepción de las asambleas neuronales de Hebb (1949) quien especuló, entre otras cosas, que los cambios en las conexiones sinápticas están en la base del aprendizaje.

La conjunción de dichas ideas inspiró la creación de los modelos neuronales computacionales. Pequeños conjuntos de neuronas artificiales como las descritas por McCulloch y Pitts trabajan para transformar un conjunto de patrones de entrada en otro conjunto de patrones de salida, estas transformaciones se daban tras la modificación del valor de las conexiones o sinapsis artificiales entre las neuronas. Los dos modelos con estas características que más influencia tuvieron fueron el Perceptrón de Rosenblatt y los ADALINE de Widrow y Hoff; estos modelos representaron un importante cambio en la forma de entender la computación, pues no se trataba de sistemas que ejecutaran tareas dictadas por un programador, sino que podían aprender y modificar su procesamiento mediante la experiencia. El poco avance tecnológico fue un impedimento para que el desarrollo de este tipo de sistemas continuara; en 1969 Minsky y Papert (citado en Bechtel & Abrahamsen, 1991) demostraron cómo los sistemas tipo Perceptron fallaban en responder a tareas donde no existían soluciones (o correspondencias) lineales entre el conjunto de datos de entrada y el conjunto de datos de salida, dicha crítica afectó al creciente interés en los modelos neuronales y lo mantuvo en niveles mínimos durante casi 15 años.

El interés en los modelos conexionistas resurgió a mediados de los años ochenta, Rumelhart y McClelland junto con el grupo PDP (1986) publicaron dos volúmenes donde destacaron dos principales avances con respecto a los antiguos perceptrones: 1. Se comenzó a trabajar con redes multicapa, llamadas así pues contenían unidades de procesamiento intermedias u ocultas, que trabajaban entre las unidades de representación de patrones de entrada (i.e., estimulación) y las unidades de representación de patrones de salida (i.e., respuesta). Las capas ocultas podían adoptar representaciones intermedias, lo que generaba al menos una transformación extra en el conjunto de datos, permitiendo dar solución a los problemas que no tenían una solución lineal; y 2. Se mostraba cómo el algoritmo de retropropagación o *backpropagation*, podía computar los cambios necesarios en los valores de todas las conexiones de estos sistemas multicapa.

Los modelos neuronales generan estados de activación óptimos al modificar el valor de las conexiones que hay entre las unidades o neuronas artificiales. La forma en la que se implementan los cambios en el valor de las conexiones es altamente relevante, pues es ahí donde se representa en gran medida el acercamiento teórico al problema estudiado, y es ahí donde quedan formalizados a nivel matemático los supuestos básicos y las limitaciones del modelo.

En general los algoritmos para la modificación de las conexiones se pueden dividir en dos grandes categorías, los supervisados y los no supervisados. Para los primeros hay un agente o una señal que determina un nivel deseado de activación a partir del cual se calcula una diferencia con la activación generada por el modelo; las diferencias entre el valor deseado y el obtenido en las activaciones marcan la magnitud y dirección de las modificaciones en los valores de conexión (e.g., Plunkett & Elman, 1997). Por otra parte, los algoritmos no supervisados por lo general extraen información de las regularidades de estimulación a las que son presentados, para en base

a dichas regularidades modificar sus conexiones generando un modelo estadístico del entorno (e.g., Kohonen, 2001).

Paralelamente al avance en el desarrollo de redes multicapa asociativas, que trabajan con algoritmos de aprendizaje supervisado, se desarrolló el trabajo en redes *competitivas*, que en su mayoría funcionan con algoritmos de aprendizaje no supervisado, las redes más representativas de esta última categoría son los mapas auto-organizables (SOM por Self Organizing Maps) trabajados inicialmente por el grupo de Teuvo Kohonen en 1984 (citado en Kohonen, 2001). Los SOM pueden extraer categorías de un espacio de entrada o input complejo sin supervisión, usando algoritmos de aprendizaje que son biológicamente plausibles (ver Mayor & Plunkett, 2010).

La idea misma de modificación sináptica de Hebb puede formalizarse en un algoritmo que comúnmente trabaja sin supervisión. A continuación se presentan los principales mecanismos de funcionamiento de las redes asociativas multicapa, de los mapas auto-organizables y de las redes Hebbianas.

Arquitectura y Funcionamiento de los Modelos Neurocomputacionales

Actualmente existen tantos algoritmos y tipos de redes como propósitos de simulación o investigación, por lo cual en esta sección se exponen las partes y funcionamiento básico que suponen la mayoría de los sistemas de redes neuronales artificiales.

Arquitecturas

Todas las redes están compuestas por dos tipos de elementos donde se lleva a cabo el procesamiento de las tareas: neuronas y conexiones. Las neuronas o unidades de procesamiento

están acomodadas en distintas capas que pueden estar representando ya sea estructuras cerebrales, o módulos globales en los que se divide el procesamiento; por ejemplo, en el primer caso se pueden diseñar capas especializadas para simular la acción de la corteza cerebral y capas para simular la actividad del hipocampo, para así evaluar posibles mecanismos de recuperación de información que resultan de la interacción de dichas capas (e.g., Murre, 1996; O'Reilly, 2005). En el segundo caso, se pueden diseñar capas que representen simplemente un sistema de entrada de información *-input-*, uno de procesamiento y uno de salida o respuesta *-output-*, varios ejemplos de este segundo tipo de redes son revisados por Plunkett y Elman (1997). Con estas redes se pueden explorar procesos cognitivos generales, más que fisiológicos particulares, como la formación de categorías o la adquisición de repertorios léxicos a través del tiempo.

Las neuronas están caracterizadas por un valor de activación que se representa de forma numérica, dependiendo de las propiedades particulares del modelo dicho valor puede estar entre -1 y 1, o entre 0 y 1; los valores elevados positivos representan una activación mayor y los valores negativos o de 0 (según sea el caso) representan inhibición. Esa misma lógica de correspondencia entre activación (1) y no activación (0 o -1), se utiliza para estimular o presentar información a las redes en las capas de entrada. Los valores de entrada se propagan a las demás neuronas, y a lo largo de varias iteraciones van cambiando sus activaciones como resultado de procesos de aprendizaje hasta encontrar los valores óptimos que les permiten resolver las situaciones problema que se les presentan o adaptarse a distintos ambientes. Los valores generados por las capas de salida se toman como las respuestas de la red y se interpretan con la misma lógica. Por ejemplo, en una simulación de condicionamiento clásico, los estímulos condicionados o incondicionados se presentan mediante la activación de las unidades de entrada que representan la presencia de señales medioambientales, de las unidades de entrada la

activación fluye hacia capas de procesamiento intermedio hasta llegar a unidades de salida cuya activación representará la emisión de una respuesta (Burgos, 2007).

Los valores de activación alcanzados por cada unidad dependen en gran medida de los valores de las conexiones pues la activación en las capas de entrada se propaga a todo el sistema por éstas.

Las conexiones, también llamadas pesos sinápticos por su referente biológico, también se representan por valores numéricos, que pueden ser tanto positivos como negativos, los valores que adquieren las conexiones simulan los distintos grados de interacción que se pueden dar en las redes neuronales biológicas, por ejemplo, valores negativos están asociados a la acción inhibidora de la neurona presináptica sobre la neurona postsináptica; valores positivos están asociados a efectos excitadores de la neurona presináptica sobre la neurona postsináptica, las conexiones normalmente inician con valores aleatorios y estos van cambiando por acción del aprendizaje.

Aprendizaje

El algoritmo más común de aprendizaje dentro de los modelos asociativos es el de retropropagación o backpropagation (Rumelhart, Hinton, & Williams, 1986). Es un algoritmo que logra modificar los valores de las conexiones de todas las capas de una red mediante la comparación del resultado generado por la red, con el resultado esperado; es decir, es un algoritmo de aprendizaje supervisado con modificación por corrección del error. Este algoritmo modifica los valores de las conexiones, comenzando por las últimas capas de pesos y terminando con las primeras; dicho funcionamiento supone que existe una conectividad hacia atrás en la red

neuronal. Distintos parámetros como tasa de aprendizaje o momentum pueden ser manipulados (McLeod, Plunkett, & Rolls, 1998) para simular dinámicas o situaciones de aprendizaje distintas.

Pese al amplio uso que ha tenido el algoritmo de retropropagación, ha sido criticado por su poca congruencia con procesos neurofisiológicos; no existe evidencia contundente de que en una red biológica la información vaya en un sentido y luego regrese en el sentido contrario (O'Reilly & Munakata, 2000, p. 162), sin embargo, han surgido algunos algoritmos más complejos para redes asociativas con aprendizaje supervisado que incluyen para sus cálculos una mayor plausibilidad biológica. El algoritmo LEABRA es un ejemplo de esto (O'Reilly & Munakata 2000), LEABRA considera cambios en los valores sinápticos a partir de un proceso de corrección del error que no requieren de una conectividad hacia atrás; sino que combina señales de corrección con un aprendizaje tipo Hebbiano, donde los grupos neuronales que se activan simultáneamente en respuesta a la estimulación (externa o interna) fortalecen las conexiones entre sí.

Hebb (1949) propuso una de las primeras formalizaciones para entender el aprendizaje asociativo desde una perspectiva neuropsicológica, su idea básica fue que cuando dos grupos neuronales se activan conjuntamente y uno de ellos participa en el disparo del otro, algún cambio metabólico toma lugar de tal forma que la eficacia de uno de los grupos para activar al otro se incrementa. Esta idea supone un cambio plástico entre los dos grupos neuronales que ha sido formalizado como un incremento en el peso de las conexiones que hay entre las neuronas que concurrentemente se activan. El fortalecimiento en las conexiones permite una mayor propagación e implica que la activación de una representación generará la recuperación y activación de las representaciones con las que ha sido concurrente. Existe evidencia neurofisiológica de que este tipo de aprendizaje es biológicamente viable, con los experimentos

de potenciación a largo plazo, que han demostrado mejoras en la eficiencia sináptica tras la estimulación con alta frecuencia entre dos grupos neuronales, también se ha demostrado el proceso contrario con experimentos donde la estimulación se presenta con baja frecuencia, dando como resultado una depresión a largo plazo, donde la eficiencia sináptica de un grupo neuronal para activar a otro disminuye (Bliss, Collingridge, & Morris, 2007; Malenka & Bear, 2004). El aprendizaje Hebbiano en la mayoría de los casos es considerado como un aprendizaje no supervisado, aunque puede adecuarse para actuar en situaciones supervisadas.

Por otra parte, las redes tipo SOM trabajan con un aprendizaje no supervisado, se trata de sistemas inspirados en la organización topológica de la corteza cerebral, que resultan adecuados para entender procesos de percepción, mediante la organización y visualización de datos que originalmente pudieron ser de varias dimensiones y son transformados a un mapa de dos dimensiones (Kohonen, 2001).

En estos modelos las neuronas se agrupan en capas bidimensionales donde cada una está caracterizada por un valor vectorial, la información entrante es también presentada con vectores. El sistema determina cuál de las neuronas tiene el vector más parecido al vector de entrada, y a partir de ello considera a una neurona como ganadora o como la neurona de mejor ajuste; dicha neurona modifica sus valores vectoriales en cada iteración o corrida del modelo para irse asemejando más al valor del vector de entrada. Igualmente, las neuronas vecinas a la ganadora modifican sus valores para generar cierta proximidad o semejanza con la neurona ganadora, de tal forma que tras la presentación de varios vectores de entrada durante varias corridas se va conformando un mapa topológico en la capa bidimensional de neuronas que representa cierta estructura de los vectores de entrada. Los modelos SOM agrupan información preservando las relaciones topológicas o métricas más importantes de los datos de entrada, se puede considerar

que producen cierto tipo de abstracción sobre estos datos, lo cual puede ser utilizado en procesos de análisis de datos, de percepción en sistemas artificiales, de control y de comunicación (Kohonen, 2001).

Existen también modelos mixtos, que pueden estar compuestos de módulos con aprendizaje supervisado y módulos con aprendizaje no supervisado, estos parecen ajustarse mejor cuando se pretende modelar una tarea que implica varias etapas de procesamiento (Mayor & Plunkett, 2010).

Modelos Neurocomputacionales de EE

Barnes y Hampson (1993) desarrollaron una red neuronal artificial de tipo asociativa y supervisada llamada RELNET (Network for Relational Responding) capaz de responder a tareas de equivalencia de manera semejante a como lo hacen participantes humanos. Realizaron una simulación que constó de un pre-entrenamiento, donde el modelo RELNET aprendió todas las relaciones posibles en 7 clases de estímulos equivalentes (clases de práctica), después le fueron enseñadas solo algunas relaciones condicionales para el establecimiento de una nueva clase (clase 8 de prueba) y la red pudo responder de forma correcta en pruebas de relaciones simétricas y transitivas.

Los resultados positivos del trabajo de Barnes y Hampson (1993) llevaron a dos trabajos más donde se usaron modelos conexionistas para simular tareas de formación de equivalencias y de transferencia de secuencias de respuesta (Cullinan, Barnes, Hampson, & Lyddy, 1994; Lyddy, Barnes-Holmes, & Hampson, 2001); estos procesos desde una perspectiva analítico conductual constituyen la base de algunos procesos lingüísticos como la sintaxis.

Lyddy y Barnes-Holmes (2007), dejaron el tema de la sintaxis para centrarse en los procesos más básicos de formación de clases de equivalencia. Analizaron las diferencias sobre la formación de clases debidas a dos protocolos de entrenamiento distintos: el protocolo de series lineales (entrenamiento de relaciones AB y BC) y el protocolo de racimos o de uno a muchos (BA y BC). De forma general, sus resultados mostraron que en los ensayos de prueba de relaciones emergentes, bajo el protocolo de series lineales hubo una mayor cantidad de errores en comparación al protocolo de racimos.

Un análisis detenido del funcionamiento de RELNET y los modelos de ahí derivados, muestra que la emergencia de relaciones entre estímulos no se da a partir del entrenamiento de relaciones de línea base, sino que se debe a un error en el diseño de la red, que lejos de poder generar ejecuciones novedosas funcionales, genera una mala interpretación o una simulación incompleta. El error se deriva por la forma en la que se presenta la información en la capa de entrada, cada patrón de entrada contiene la información de los estímulos presentes en cada ensayo, así como de la función o ubicación que tiene cada uno; como estímulo muestra o estímulos de comparación, dado que se trata de simulaciones de ensayos de IM. Las unidades encargadas de señalar la función como estímulo muestra o comparación son denominadas unidades "Z".

El problema crítico es que los ensayos de prueba tienen la misma activación en las unidades Z que algunos ensayos de entrenamiento (o ensayos de las clases de práctica). Por ejemplo, al presentar la prueba de transitividad A2C2, hay una serie de unidades Z que tendrán la misma activación que el ensayo entrenado A1C1; las respuestas que generan los modelos tipo RELNET en las fases de prueba se basan únicamente en las unidades Z, así que en términos

estrictos la ejecución en dicha fase no es pertinente de ser evaluada como si se tratara de ensayos nuevos, sino que son ensayos repetidos y previamente aprendidos por el modelo.

En otras áreas de estudio, se ha analizado la formación de relaciones-estímulo arbitrarias con modelos conexionistas, por ejemplo, Mayor y Plunkett (2010) estudiaron el aprendizaje temprano de palabras a partir de la formación de correspondencias entre la representación neuronal de palabras y la representación neuronal de los objetos designados por dichas palabras. Utilizaron mapas auto-organizables, para que el modelo lograra la abstracción de los distintos estímulos y utilizaron conexiones tipo Hebbianas entre los mapas para generar la correspondencia entre palabras y objetos; sin embargo, en este tipo de modelos no se han estudiado relaciones derivadas o emergentes entre estímulos.

Son pocos los modelos conexionistas desarrollados para modelar el proceso de formación de clases de estímulos equivalentes, los modelos existentes se han restringido a trabajar con redes asociativas y simulaciones de procedimientos de IM, por lo cual las interpretaciones derivadas a partir de una estrategia conexionista para tener un mejor entendimiento de la formación de clases de estímulos equivalentes son aún muy limitadas. El uso de distintos modelos y la simulación de distintos procedimientos, puede generar acercamientos que resulten más adecuados para analizar la formación de clases estímulo.

JUSTIFICACIÓN

Las motivaciones para desarrollar modelos neurocomputacionales de la formación de clases de estímulos equivalentes se puntualizan a continuación.

No existe un consenso acerca de los mecanismos que subyacen la habilidad para derivar relaciones transitivas de equivalencia entre estímulos y no hay modelos cuantitativos que describan las diferencias en el nivel de relación entre los miembros de una clase de equivalencia.

Tras analizar los modelos conexionistas hasta ahora existentes que simulan tareas de EE fue imposible determinar si estos sistemas artificiales podían demostrar emergencia de relaciones transitivas entre estímulos como las que comúnmente se observan en ejecuciones de participantes humanos. Esta situación cuestiona la pertinencia de usar modelos conexionistas para estudiar EE.

Toda vez que puedan simularse tareas de EE en modelos conexionistas será posible aprovechar las facilidades metodológicas que permiten este tipo de simulaciones como son el tener un completo conocimiento de la historia de aprendizaje en un sistema, tener un alto control sobre variables extrañas, poder manipular el efecto de distintos parámetros de aprendizaje para observar su efecto sobre la formación de una clase, poder generar predicciones del éxito en la formación de clases dependiente de un tipo de entrenamiento particular, y analizar las teorías existentes sobre formación de clases de equivalencia a la luz de un modelo cuantitativo.

La simulación de EE permitirá vincular los hallazgos en EE con otras áreas donde comúnmente se utiliza modelamiento neurocomputacional, pues los algoritmos utilizados en los distintos modelos tienden a representar procesos generales de aprendizaje subyacentes a varios fenómenos psicológicos, como son la categorización, la memoria y la discriminación de estímulos.

OBJETIVOS

General. Desarrollar modelos neurocomputacionales del fenómeno conductual de EE, que simulen datos apegados a la evidencia empírica y permitan analizar el aprendizaje y la ejecución que se observa en humanos.

Particulares. Analizar la viabilidad de simular la formación de clases de estímulos equivalentes en modelos neurocomputacionales.

Generar descripciones cuantitativas del nivel de relación que mantienen los estímulos pertenecientes a una clase.

Simular y explorar los efectos de nodalidad, entendidos como una diferencia en el nivel de relación entre los estímulos de una clase dependiente del número de estímulos que median la transitividad.

Tener modelos que ayuden a generar predicciones respecto al éxito en la formación de una clase de estímulos equivalentes dependiente del tipo de entrenamiento simulado.

Desarrollar modelos que simulen un tipo de procesamiento biológicamente plausible.

Poder discutir las teorías existentes del fenómeno de EE a la luz del funcionamiento y los datos de modelos neurocomputacionales.

CAPÍTULO 3. EXPERIMENTOS, MODELOS NEUROCOMPUTACIONALES, RESULTADOS Y DISCUSIONES

Experimento 1

Los modelos para simular EE hasta ahora existentes (Barnes & Hampson, 1993; Cullinan et al., 1994; Lyddy et al., 2001; Lyddy & Barnes-Holmes, 2007), no generan respuestas que revelen un proceso de derivación de relaciones de equivalencia entre estímulos; como se explicó en el Capítulo 2, las unidades “Z” de estos modelos son igualmente utilizadas en los ensayos de entrenamiento y de prueba, y por si solas pueden activar la respuesta correcta durante las fases de prueba, por lo cual, dichas respuestas resultan originalmente entrenadas y no derivadas.

Con el primer modelo que se presentará en esta tesis se pretende analizar la formación de clases de equivalencia en redes conexionistas tradicionales (redes multicapa con aprendizaje de tipo backpropagation). Para este fin se consideró que el procedimiento de IM sería poco conveniente de ser simulado, pues requiere codificar la información de los estímulos que participan en un ensayo, y de su función (como estímulo de muestra o de comparación), esta situación inevitablemente genera la necesidad de utilizar unidades extra para señalar las funciones de los estímulos, como en el caso de las unidades Z de los modelos descritos, que pueden generar interpretaciones confusas y cuestionables.

Se consideró entonces simular un procedimiento que no requiriera la señalización de las funciones estímulo, en este contexto los procedimientos con estímulos compuestos y respuestas tipo *yes/no* resultaron alternativas adecuadas al procedimiento IM.

Para tener datos de ejecuciones de humanos con los cuales comparar al modelo se realizó una réplica sistemática del estudio de Debert y colaboradores (2007). Se establecieron dos clases

estímulo de tres estímulos cada una, a diferencia del procedimiento original con respuestas tipo *go/no-go*, en esta réplica se utilizaron respuestas de tipo *yes/no*.

Método

Participantes. Seis mujeres estudiantes de psicología sin conocimientos de EE, con edades entre 19 y 22 años participaron de forma voluntaria. Ellas recibieron créditos en una materia del semestre cursado tras su participación independientemente de su ejecución.

Aparatos y Estímulos. Las sesiones experimentales se llevaron a cabo en un cuarto de 4m x 6m, cada participante se sentó frente a una pantalla sensible al tacto y utilizó unos audífonos. Un programa desarrollado en Visual Basic controló la presentación de estímulos y registró todas las respuestas.

Los estímulos fueron seis figuras abstractas como las utilizadas en los estudios de Debert y cols. (2007); y Markham y Dougher (1993). Fueron designadas como A1, B1, C1, A2, B2, y C2, como se muestra en la Figura 3. Los participantes nunca vieron estas etiquetas.

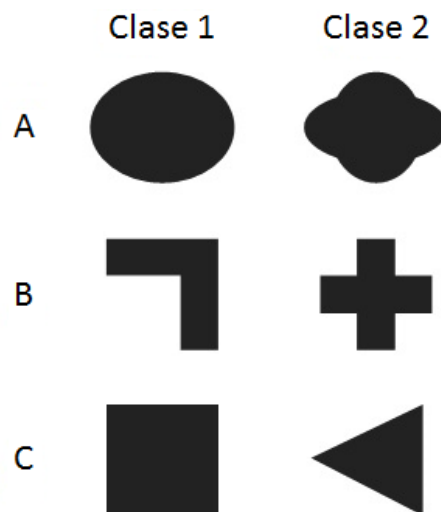


Figura 3. Figuras abstractas utilizadas para la formación de dos clases estímulo.

Procedimiento. El experimento se llevó a cabo en dos sesiones que ocurrieron en dos días consecutivos, un total de cuatro fases fueron aplicadas. La Fase 1 se aplicó durante la Sesión 1, el resto de las fases se aplicaron en la Sesión 2. En las fases de entrenamiento los participantes aprendieron las relaciones A-B y B-C, a partir de las cuales se podían derivar las relaciones simétricas B-A y C-B, y transitivas A-C y C-A. Estas fueron evaluadas durante las fases de prueba.

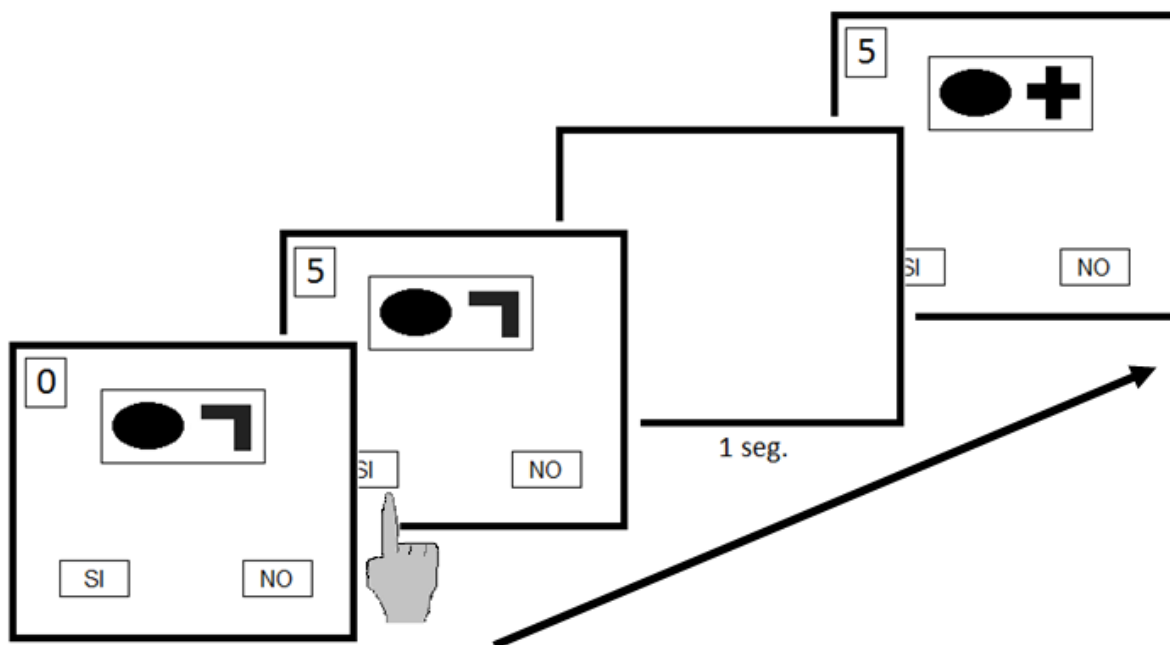


Figura 4. Ejemplo de los eventos presentados durante los ensayos de entrenamiento. Al centro aparecen los estímulos que componen al par, abajo las opciones de respuesta, arriba a la izquierda el contador de puntos. La imagen representa un ensayo A1B1 con una respuesta correcta, luego un intervalo entre ensayos y luego el inicio del ensayo A1B2.

Sesión 1, Fase 1. Entrenamiento de Relaciones de Línea Base. Las relaciones de línea base A-B, y B-C fueron entrenadas con el uso de estímulos compuestos o pares de estímulos; durante esta fase se presentaron distintos estímulos compuestos ante los cuales los participantes

tenían que responder con la opción “sí” o la opción “no”. La elección de “sí” fue reforzada en presencia de los pares intra-clase A1B1, B1C1, A2B2 y B2C2; mientras que la elección de la respuesta “no” fue reforzada en presencia de los pares cross-clase A1B2, B1C2, A2B1 y B2C1. La nomenclatura de los pares de estímulos obedece a la forma en la que se presentaron a los participantes, cuando se lee A1B1, indica que el participante vio un estímulo compuesto donde del lado izquierdo estaba A1 y del lado derecho estaba B1. Un ejemplo de un ensayo puede apreciarse en la Figura 4.

Al inicio de la Sesión 1 los participantes leyeron las siguientes instrucciones:

¡Gracias por tu participación!

Este estudio no es una prueba de inteligencia y no evaluará ningún aspecto de tus habilidades intelectuales, al finalizar recibirás una explicación.

Un investigador permanecerá cerca por si surge algún problema técnico, pero en ningún momento podrá darte ayuda para resolver la tarea.

Tu meta es obtener la mayor cantidad de puntos, estos puntos aparecerán en un contador ubicado en la esquina superior izquierda de la pantalla. En el centro de la pantalla aparecerán dos figuras, tu tarea es seleccionar la opción “SI” cuando consideres que las figuras están relacionadas; y seleccionar la opción “NO” cuando consideres que las figuras no están relacionadas. Podrás aprender cuáles figuras están relacionadas y cuáles no poniendo atención a los cambios en el contador de puntos.

La tarea parece sencilla pero aumentará su dificultad conforme avance así que pon atención.

Por favor repite estas instrucciones al experimentador y cuando él te lo indique toca el botón de “Avanzar”.

Tras tocar el botón “Avanzar” los ensayos comenzaron, en cada uno apareció un estímulo compuesto al centro, las opciones de respuesta abajo y un contador de puntos arriba a la izquierda, ver Figura 4.

Cuando los participantes realizaron una elección correcta el contador de puntos aumentó 5 puntos su valor mientras parpadeaba en verde, al mismo tiempo se escuchó una melodía

ascendente en los audífonos. Cuando los participantes realizaron una elección incorrecta el contador de puntos disminuyó en 5 puntos su valor mientras parpadeaba en rojo, al mismo tiempo se escuchó una melodía descendente. Durante los dos segundos que duró el parpadeo en colores ninguna respuesta surtió efecto. Hubo un intervalo entre ensayos de 1 segundo. El siguiente ensayo comenzó con el valor del contador en negro.

La Fase 1 estuvo dividida en seis etapas, en la Tabla 2 se muestra la secuencia de etapas, así como las relaciones entrenadas en cada una y el criterio de aprendizaje establecido para pasar de una etapa a otra. Los ensayos dentro de cada etapa se presentaron en orden aleatorio cuidando que no se presentara el mismo ensayo más de 2 veces consecutivas. Cada etapa se repitió hasta que los participantes lograron el criterio establecido.

Durante las etapas 1 a 5, todas las respuestas generaron las consecuencias de reforzamiento programadas. En la Etapa 6 se informó a los participantes que ya no sabrían si sus respuestas fueron correctas o incorrectas, pero que debían seguir haciendo su mejor esfuerzo. El retiro de las consecuencias durante la Etapa 6 se hizo para preparar a los participantes a responder durante las pruebas, en las que nunca se presentó reforzamiento.

Sesión 2, Fase 2, Reentrenamiento de Relaciones de Línea Base. Durante esta fase las relaciones de línea base fueron reentrenadas, se presentaron las mismas 6 etapas de entrenamiento utilizadas en la Fase 1, aunque en cada etapa se presentó la mitad de los ensayos, En la Tabla 2 se muestra la cantidad de ensayos y criterios de respuesta de la Fase 2.

Tras completar la Fase 2 los participantes fueron evaluados en las pruebas de relaciones emergentes, las fases 3 y 4 evaluaron las relaciones de simetría y transitividad, respectivamente.

Fase 3. Pruebas de Simetría. Los componentes de cada par de estímulos entrenados se presentaron en la posición contraria a la que fueron entrenados, por ejemplo, el par entrenado

A1B1, fue evaluado durante la Fase 3 como B1A1, apareciendo B1 al lado izquierdo y A1 al lado derecho. Esta evaluación permite determinar si los participantes están respondiendo a las propiedades funcionales de un par y no a su configuración perceptual. Existieron 8 ensayos de simetría, cada uno se presentó 4 veces generando un total de 32 ensayos en esta fase.

Tabla 2. Secuencia de etapas de entrenamiento y reentrenamiento.

Etapa	Estímulo Compuesto	Respuesta Correcta	Porcentaje de Reforzamiento	Fase 1		Fase 2	
				Entrenamiento		Reentrenamiento	
				Número de Ensayos	Criterios	Número de Ensayos	Criterios
1	A1B1	SI	100%	12	21/24	6	11/12
	A1B2	NO	100%	12		6	
2	B1C1	SI	100%	12	21/24	6	11/12
	B1C2	NO	100%	12		6	
3	A2B2	SI	100%	12	21/24	6	11/12
	A2B1	NO	100%	12		6	
4	B2C2	SI	100%	12	21/24	6	11/12
	B2C1	NO	100%	12		6	
5	A1B1	SI	100%	3	23/24	2	15/16
	B1C1	SI	100%	3		2	
	A2B2	SI	100%	3		2	
	B2C2	SI	100%	3		2	
	A1B2	NO	100%	3		2	
	B1C2	NO	100%	3		2	
	A2B1	NO	100%	3		2	
	B2C1	NO	100%	3		2	
6	A1B1	SI	0%	3	23/24	2	15/16
	B1C1	SI	0%	3		2	
	A2B2	SI	0%	3		2	
	B2C2	SI	0%	3		2	
	A1B2	NO	0%	3		2	
	B1C2	NO	0%	3		2	
	A2B1	NO	0%	3		2	
	B2C1	NO	0%	3		2	

Fase 4, Pruebas de Transitividad. En esta fase se evaluó la emergencia de las respuestas “sí” y “no” ante nuevas combinaciones de pares que representaron relaciones transitivas. Se evaluaron las respuestas ante los pares A1C1, A1C2, A2C2, y A2C1 y a sus simétricos C1A1, C2A1, C2A2, C1A2. Cada uno de los pares evaluados se presentó 12 veces generando un total de 96 ensayos en la Fase 4.

Resultados y Discusión

Las seis participantes completaron el experimento, cuatro de seis formaron la dos clases estímulo ABC-1 y ABC-2, pues respondieron durante las pruebas con una exactitud de al menos 85%.

En la Tabla 3 se muestra la cantidad de repeticiones que requirió cada participante de cada una de las etapas del entrenamiento y el reentrenamiento para poder completarlos, así como la cantidad y porcentaje de respuestas correctas durante las fases de prueba 3 y 4. Durante el entrenamiento y reentrenamiento las participantes mostraron una ejecución con un número mínimo de repeticiones de etapas, salvo la Participante 1 que requirió cinco repeticiones de la Etapa 5.

Durante las pruebas de relaciones emergentes, las participantes 1, 5 y 6 mostraron la ejecución más alta al responder de forma correcta en al menos 96% de los ensayos. La Participante 3 logró 93.7% y 96.8% de ensayos correctos en las fases 3 y 4, respectivamente. Estas cuatro participantes fueron las que formaron las dos clases estímulo.

La Participante 2 respondió correctamente a todos los ensayos de simetría, sin embargo, logró sólo el 60% de ensayos correctos en las pruebas de transitividad. La Participante 4 respondió correctamente a todos los ensayos de simetría, pero en los ensayos de transitividad

alcanzó sólo el 56% de respuestas correctas. Hacia el final de la Fase 4 mostró errores en todos los ensayos de los pares relacionados y mostró correctos todos los ensayos de los pares no relacionados, lo cual indica que utilizó la opción de respuesta “no” para todos estos ensayos. Este tipo de elección de respuesta por default en participantes humanos ha sido reportada con procedimientos similares (Fields et al., 2009).

Tabla 3. Ejecución de los seis participantes durante las sesiones 1 y 2.

Participantes	Sesiones	Entrenamiento y Reentrenamiento (Fases 1 y 2)								Pruebas (Fases 3 y 4)					
		Etapa 1		Etapa 2		Etapa 3		Etapa 4		Etapa 5 Todas las relaciones ref. 100%	Etapa 6 Todas las relaciones ref. 0%	Simetría	Transitividad		
		1	2	1	2	1	2	1	2						
1	1	1	(21/24)	1	(23/24)	1	(24/24)	1	(24/24)	5	(86/120)	3	(59/72)		
	2	1	(12/12)	1	(12/12)	1	(12/12)	1	(11/12)	2	(28/32)	1	(16/16)	(32/32) 100%	(94/96) 97.90%
2	1	2	(39/48)	1	(23/24)	1	(21/24)	1	(22/24)	2	(44/48)	1	(23/24)		
	2	1	(11/12)	1	(12/12)	1	(12/12)	1	(12/12)	1	(15/16)	2	(30/32)	(32/32) 100%	(60/96) 62.00%
3	1	1	(22/24)	1	(23/24)	1	(24/24)	1	(23/24)	1	(22/24)	1	(24/24)		
	2	1	(12/12)	1	(12/12)	1	(12/12)	1	(11/12)	1	(16/16)	1	(16/16)	(30/32) 93.70%	(93/96) 96.80%
4	1	1	(22/24)	1	(22/24)	1	(23/24)	1	(24/24)	1	(23/24)	1	(24/24)		
	2	1	(12/12)	1	(11/12)	1	(12/12)	1	(12/12)	1	(15/16)	1	(16/16)	(32/32) 100%	(54/96) 56.00%
5	1	2	(44/48)	1	(22/24)	1	(23/24)	1	(24/24)	1	(23/24)	1	(24/24)		
	2	1	(12/12)	1	(12/12)	1	(12/12)	1	(12/12)	1	(15/16)	1	(16/16)	(31/32) 96.80%	(96/96) 100%
6	1	1	(21/24)	1	(23/24)	1	(24/24)	1	(24/24)	1	(24/24)	1	(24/24)		
	2	1	(12/12)	1	(12/12)	1	(12/12)	1	(12/12)	1	(16/16)	1	(16/16)	(32/32) 100%	(96/96) 100%

Con el procedimiento utilizado se observó la formación de dos clases de estímulos equivalentes. Un aspecto importante a considerar es que las contingencias de reforzamiento experimentales establecieron la base para que durante las pruebas de transitividad, los ensayos A1C1 y A2C2 pudieran evocar tanto la respuesta “si” como la respuesta “no”. Hay que considerar que durante el entrenamiento, la respuesta “si” se reforzó ante los pares A1B1 y

B1C1, lo cual estableció la línea base para que emergiera la respuesta “sí” ante A1C1; al mismo tiempo se reforzó la elección de “no” ante los pares A1B2 y B2C1, lo cual también estableció la línea base para que emergiera la respuesta “no” ante A1C1. Es decir, se establecieron trayectorias de entrenamiento ambiguas y la mayoría de los participantes respondió con la opción “sí”, lo cual requiere una explicación.

Una hipótesis es que probablemente las palabras “sí” y “no” utilizadas como respuestas, generaron un sesgo en la ejecución de los participantes a favor de la elección de “sí”. Dicho sesgo tendría su origen en una historia de aprendizaje diferencial para cada una de esas palabras, la forma en la que frecuentemente se haya reforzado el uso de “sí” y de “no” en situaciones extra-experimentales de categorización tendría entonces un efecto sobre la forma en la que los participantes usan esas opciones durante la tarea experimental. Sin embargo, esta historia de aprendizaje es desconocida para el experimentador.

Una forma de evaluar si esta hipótesis es correcta es comparar la ejecución de los participantes que utilizan las opciones de respuesta verbales “sí” y “no” con la ejecución de participantes que usen opciones de respuesta no verbales, como pueden ser simplemente dos botones de colores distintos. Bajo el supuesto de que dichos colores no son comúnmente utilizados en situaciones extra-experimentales de categorización, a diferencia de las palabras.

Otra forma de evaluar si la historia diferencial de aprendizaje para las opciones de respuesta tiene un efecto favorable sobre la formación de clases cuando se presenta un entrenamiento ambiguo es con un modelo neurocomputacional que pueda simular esta tarea y pueda simular una historia diferencial de aprendizaje asociado a las opciones de respuesta.

Ambas formas de evaluar la hipótesis del sesgo en las opciones de respuesta serán abordadas en esta tesis, tanto en el Experimento 2, como en la segunda simulación del Modelo Neurocomputacional 1.

Modelo Neurocomputacional 1

Se desarrolló en una red multicapa con aprendizaje por retropropagación, se analizó si este modelo logra aprender y derivar relaciones-estímulo en una situación que simula al Experimento 1. Una descripción gráfica del Modelo Neurocomputacional 1 puede verse en la Figura 5.

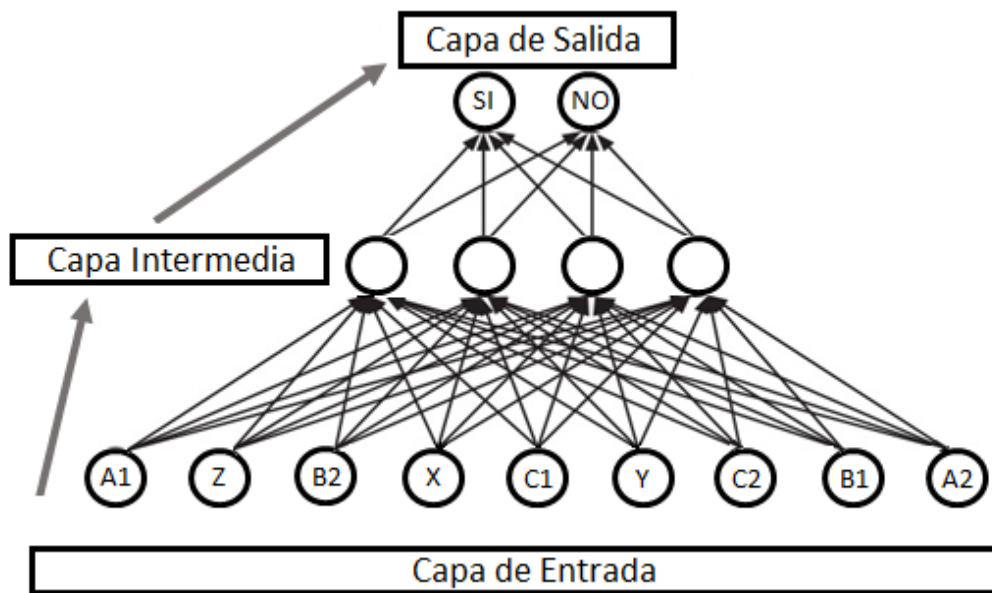


Figura 5. Arquitectura del Modelo Neurocomputacional 1.

Arquitectura

El modelo consta de 3 capas, una de entrada o estimulación, una capa intermedia de procesamiento y una de salida o respuesta. La activación o no activación de las unidades de entrada representó la presencia o ausencia de cada uno de los seis estímulos que podían formar

pares de estímulos. Por ejemplo, el estímulo compuesto A1B1 fue representado por la activación de las unidades correspondientes con A1 y B1, generando el vector de estimulación 100000010. En la Tabla 4 se muestran todos los patrones de estimulación utilizados durante el entrenamiento. La posición de las unidades que representaban a cada uno de los estímulos en la capa de entrada se determinó de forma semialeatoria, cuidando que no hubiera dos componentes de la misma clase juntos.

Tabla 4. Patrones de entrada presentados durante el entrenamiento y patrones de salida esperados.

Estímulos Compuestos	Estímulos Representados en la Capa de Entrada									Unidades de Respuesta en la Capa de Salida	
	A1	Z	B2	X	C1	Y	C2	B1	A2	SI	NO
	Patrones de estimulación en la capa de entrada									Patrones correctos/ esperados de salida	
A1B1	1	0	0	0	0	0	0	1	0	1	0
A1B2	1	0	1	0	0	0	0	0	0	0	1
B1C1	0	0	0	0	1	0	0	1	0	1	0
B1C2	0	0	0	0	0	0	1	1	0	0	1
A2B2	0	0	1	0	0	0	0	0	1	1	0
A2B1	0	0	0	0	0	0	0	1	1	0	1
B2C2	0	0	1	0	0	0	1	0	0	1	0
B2C1	0	0	1	0	1	0	0	0	0	0	1
XY	0	0	0	1	0	1	0	0	0	1	0
YZ	0	1	0	0	0	1	0	0	0	1	0
XZ	0	1	0	1	0	0	0	0	0	1	0

La capa intermedia consistió de cuatro unidades cuyos valores de activación fueron generados por el modelo e interpretados como la “representación” desarrollada por el modelo ante cada tipo de estimulación. La capa de salida consistió de dos neuronas que representaron la acción motora sobre una de las dos posibles opciones de respuesta, la opción “si” y la opción “no”. En la parte derecha de la Tabla 4 se muestran los patrones de activación correctos o esperados para las unidades de respuesta ante cada uno de los ensayos simulados. Las tres capas

del modelo estaban completamente conectadas con dirección de la capa de entrada a la capa de salida. Las conexiones comenzaron en valores aleatorios entre 1 y -1.

Funcionamiento

Cada neurona de la capa intermedia y de salida generaba un nivel de activación que resultaba de ponderar todas las entradas que recibía desde otras neuronas y transformar ese valor de entrada en uno de salida con una función sigmoide que permitió establecer la activación de cada neurona en un rango entre 0 y 1. El cálculo de la entrada de información y el de transformación de entradas en niveles de activación están expresados en las ecuaciones 1 y 2, respectivamente.

$$(1) \text{netinput}_i = \sum_j a_j * w_{ij}$$

$$(2) \text{act} = 1/(1+\exp(-\text{netinput}))$$

El proceso de aprendizaje en este modelo implica que tras cada presentación de estimulación, o ensayo, se obtiene un valor de error, que es la diferencia entre la respuesta generada por el modelo y la respuesta esperada para cada ensayo, esto resulta similar a lo que ocurre con humanos cuando generan una respuesta ante un ensayo y el contador de puntos les indica si la respuesta fue correcta y debe mantenerse o si fue incorrecta y debe cambiar.

Tras calcular el valor del error, se genera un proceso de asignación de crédito donde se evalúa con qué magnitud está colaborando cada neurona al valor de error, esto implica que neuronas muy activas serán consideradas como más responsables que neuronas menos activas y por ende, la modificación en las conexiones que partan de neuronas más activas serán más drásticas. Este proceso se repite hasta que el valor del error llega a un nivel criterio (e.g., error cuadrático medio ≤ 0.05).

Simulación 1

Primero se entrenaron en el modelo las mismas relaciones de línea base que aprendieron los humanos; el modelo aprendió a activar la unidad de respuesta “si” cuando los vectores de estimulación representaron a los pares intra-clase A1B1, B1C1, A2B2 y B2C2, y aprendió a activar la unidad de respuesta “no” cuando los vectores de estimulación representaron a los pares cros-clase A1B2, B1C2, A2B1 y B2C1.

Durante esta fase el modelo también aprendió las relaciones existentes entre los tres estímulos de una tercer clase denominada como clase XYZ, es decir, el modelo aprendió a activar la unidad “si” ante los pares XY, YZ y XZ. Esta parte del entrenamiento fue determinante para que el modelo respondiera adecuadamente, pues en ausencia de este entrenamiento las dos opciones de respuesta son indiferenciables. Aun cuando una opción esta etiquetada por el experimentador como “si” y la otra opción como “no”, el modelo no es sensible ni es informado acerca de estas etiquetas, para el modelo simplemente son dos opciones de respuesta; el entrenamiento de XYZ permite que el modelo aprenda que una de esas opciones de respuesta (la que adquirirá la función de “si”) pueda ser utilizada ante cualquier combinación de elementos de una clase y no aprende esto para la otra opción. El entrenamiento de XYZ es determinante para entender la lógica de la Simulación 2, donde no hay entrenamiento de XYZ, con lo cual se puede simular el uso de respuestas sin etiquetas verbales con participantes humanos, en vez de respuestas con un sesgo verbal como las respuestas “si/no”

El criterio de entrenamiento utilizado fue que el modelo lograra un valor de error cuadrático medio igual o menor a 0.05. Una vez logrado dicho valor se procedió con la fase de prueba.

Dado que el modelo no fue diseñado a ser sensible a la posición de los estímulos en un espacio determinado, sino a las relaciones entre éstos, no se corrieron ensayos de simetría, sino que se pasó directamente a las pruebas de transitividad. Durante la fase de prueba el algoritmo de aprendizaje fue desactivado de tal forma que no se dieron modificaciones en las conexiones, el modelo utilizó el valor de las conexiones a las que llegó al final del entrenamiento para procesar la información de los ensayos de prueba. Se evaluó la ejecución del modelo al presentarle los pares de prueba A1C1, A1C2, A2C2, A2C1. Los vectores utilizados para cada uno de estos pares pueden ser observados en la Tabla 5.

Tabla 5. Patrones de estimulación utilizados para los ensayos de prueba.

Estímulos Compuestos	Estímulos Representados en la Capa de Entrada								
	A1	Z	B2	X	C1	Y	C2	B1	A2
Patrones de estimulación en la capa de entrada									
A1C1	1	0	0	0	1	0	0	0	0
A1C2	1	0	0	0	0	0	1	0	0
A2C2	0	0	0	0	0	0	1	0	1
A2C1	0	0	0	0	1	0	0	0	1

El modelo se corrió en seis ocasiones, en cada una se comenzó con valores aleatorios distintos en las conexiones.

Resultados y Discusión

Las seis corridas de esta simulación lograron aprender la tarea, pues llegaron a un valor de error cuadrático medio menor a 0.05. Para las pruebas se consideró que las ejecuciones del modelo serían indicativas de un proceso de formación de clases cuando la activación de la unidad de respuesta “si” alcanzara un valor de al menos 0.85 ante los pares de prueba intra-clase A1C1 y A2C2, y la unidad de respuesta “no” alcanzara dicho valor ante los pares cross-clase A1C2 y A2C1.

Basados en dicho criterio, en la Figura 6 se puede observar que las corridas 1 a 5 mostraron la formación de dos clases estímulo, mientras que la sexta corrida no logró alcanzar todos los valores criterio.

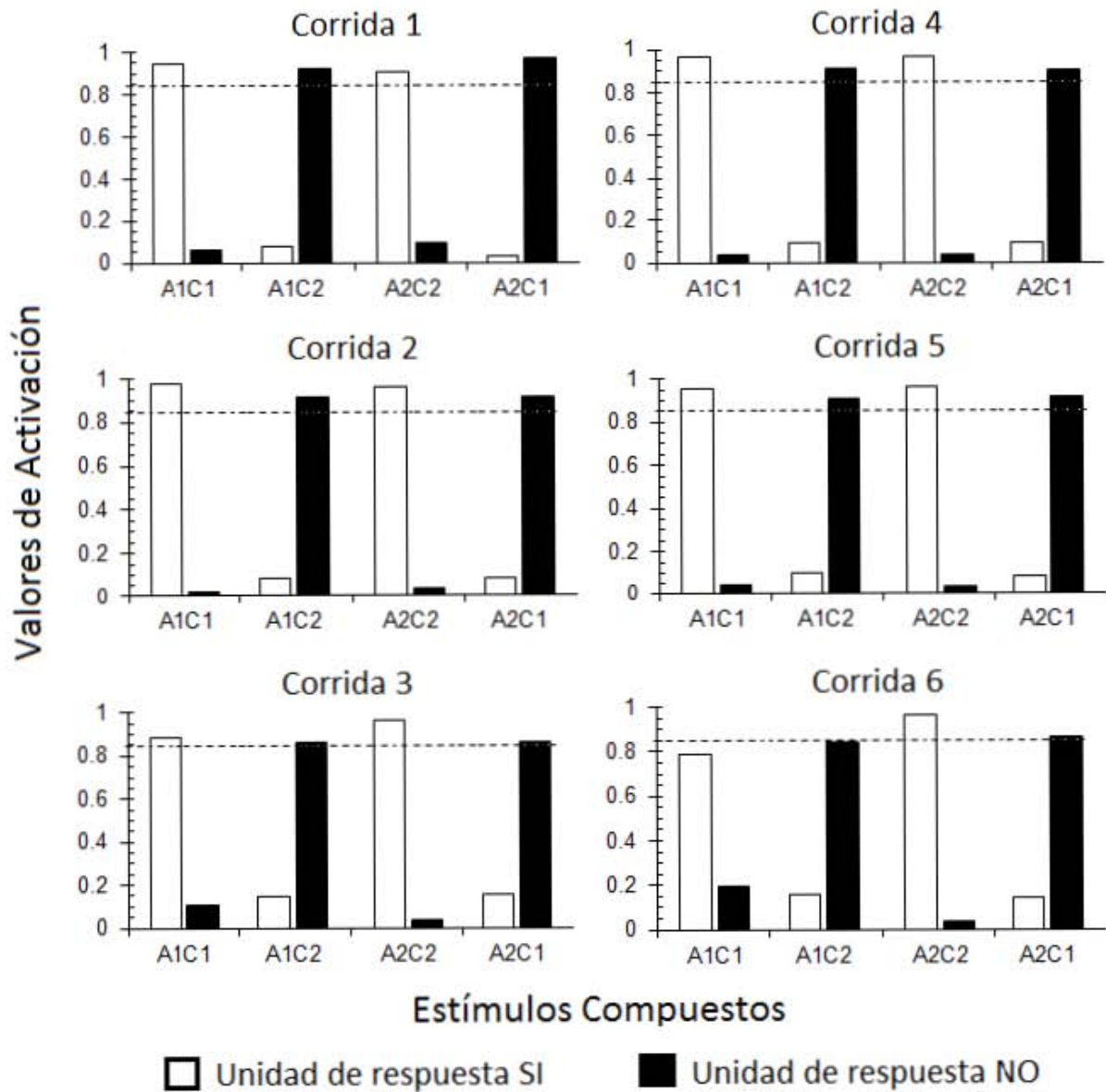


Figura 6. Las barras muestran los valores de activación alcanzados por las unidades de respuesta “si” en blanco y “no” en negro, para cada uno de los cuatro ensayos de prueba en las seis corridas del Modelo Neurocomputacional 1.

Análisis de Racimos en la Capa Intermedia. Como se mencionó en el Capítulo 2, en las capas intermedias de los modelos asociativos multicapa se realiza la primer transformación de los vectores de estimulación, esto implica que ahí se genera una representación intermedia entre la estimulación y la respuesta (Thomas & McClelland, 2008). Esto corresponde con una etapa de la resolución de una tarea que es inobservable en experimentos conductuales. Una forma común de analizar el proceso de representación en la capa intermedia es con un análisis de racimos, con el que es posible observar cuáles patrones de estimulación son transformados para generar representaciones intermedias similares. Este análisis agrupa patrones intermedios similares al computar la distancia Euclidiana que hay entre ellos. En la Figura 7 se muestra una representación gráfica de un análisis de racimos aplicado a la segunda corrida del modelo, los análisis generados para el resto de las corridas generaron patrones similares.

Se pueden observar tres racimos principales en la Figura 7; el primero corresponde a la representación de los pares de la primer clase: A1B1, B1C1 y A1C1, el segundo corresponde a los pares de la segunda clase: A2B2, B2C2 y A2C2; el tercero corresponde a los pares cross-clase A2C1, B1C2, A2B1, A1B2, B2C1 y A1C2. Esto es, los pares que conforman a los primeros dos racimos son los asociados con la activación de la unidad “si”, los pares del tercer racimo están asociados con la activación de la respuesta “no”. Los racimos indican que durante el entrenamiento, la capa intermedia se encargó de transformar los distintos patrones de entrada en activaciones similares dependiendo de la membresía de clase. Por ejemplo, el par A2B2 fue representado en la capa de entrada con 001000001, el par B2C2 fue representado en la capa de entrada con el vector 001000100. Sin embargo, ambos pares generaron un nivel de activación de 1001 en las cuatro unidades de la capa intermedia. La red estableció los valores de conexión

adecuados que permitieron que el par de prueba A2C2 presentado en la capa de entrada como 000000101, también generara una activación de 1001 en la capa intermedia durante las pruebas.

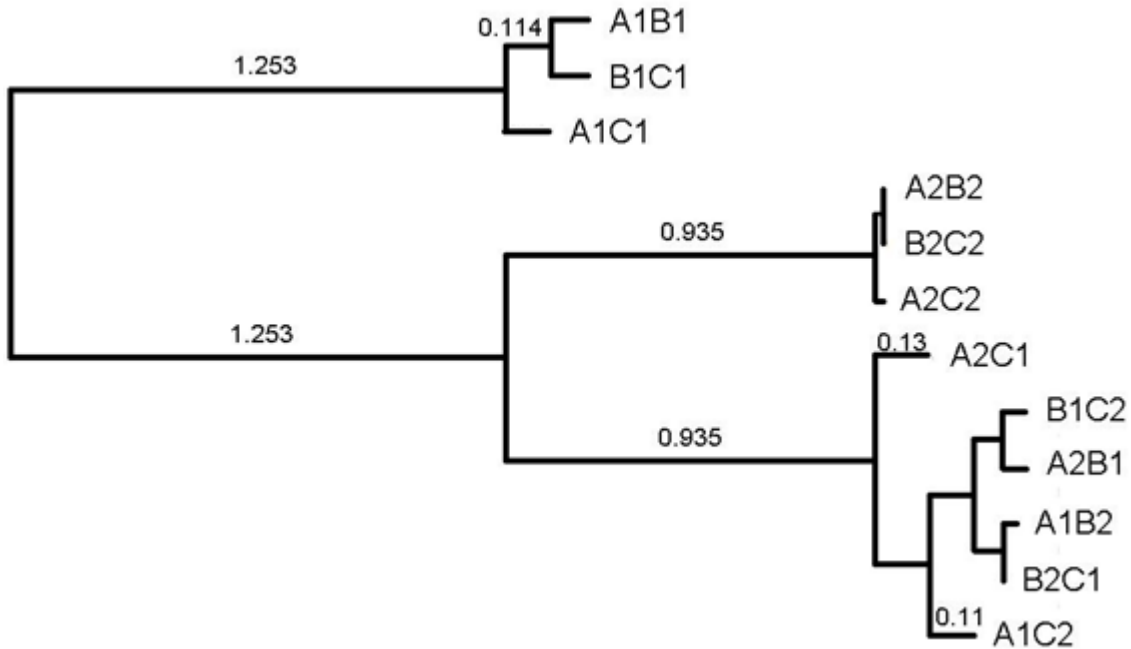


Figura 7. Diagrama de los clusters o racimos generados en la capa intermedia durante la Corrida 2. Se aprecian algunos valores que representan distancia Euclidiana. Mientras más cercanos están horizontalmente los pares de estímulos generaron activaciones más parecidas en la capa intermedia.

El modelo aprendió a generar la misma representación intermedia para pares que conformaron la misma clase, esto puede ser interpretado como una estrategia que probablemente es utilizada por participantes humanos para responder en pruebas de EE; sin importar las características perceptuales de los estímulos, pueden formar parte de una misma clase al ser representados de la misma forma. Es decir, puede considerarse que cada clase o categoría tiene una representación asociada, congruente con un mismo nivel de activación en la capa intermedia independiente de la estimulación inicial; la equivalencia se da cuando cualquier miembro de la clase es capaz de disparar dicha representación. Si originalmente una clase estaba conformada

únicamente por el estímulo A1 y su asociación con B1 generó que B1 se volviera un disparador de la misma representación que A1, luego la asociación ente B1 y C1 generó la misma consecuencia, más tarde se podrá evaluar la relación A1 y C1, y dado que ambos disparan la misma representación es más probable que se responda correctamente a esa relación en comparación a si se presenta A1 con algún estímulo de otra clase. Esta interpretación estaría de acuerdo a la propuesta de nombramiento, donde se propone que el mismo nombre es asignado a elementos distintos y eso permite la formación de una categoría (Horne & Lowe, 1996)

Simulación 2

Para poder evaluar la hipótesis respecto al sesgo verbal derivado del aprendizaje diferencial asociado a las opciones de respuesta, se ejecutaron seis nuevas corridas del modelo (corridas 7 a 12) en las que no se presentó el entrenamiento de la clase XYZ, de esta forma en vez de tener la simulación de una respuesta “sí”, asociada a cualquier par de estímulos intra-clase, simplemente se tuvieron dos respuestas denominadas en esta simulación como R-1 y R-2. Esto resulta similar a utilizar respuestas arbitrarias (e.g., sólo colores) con participantes humanos. En esta simulación se entrenó la activación de R-1 para los pares intra-clase A1B1, B1C1, A2B2 y B2C2, y la activación de R-2 para los pares cros-clase A1B2, B1C2, A2B1 y B2C1.

Resultados y Discusión

En esta segunda simulación las seis corridas del modelo nuevamente aprendieron todas las relaciones de entrenamiento logrando el valor criterio establecido. Las pruebas de relaciones emergentes revelaron que el modelo no pudo responder de forma consistente en las seis corridas. Sólo en las corridas 7 y 8, la R-1 se activó ante los pares emergentes intra-clase y la R-2 se activó

para los pares emergentes cros-clase. En las corridas 9, 11 y 12 el modelo invirtió el patrón anterior, activando R-1 para pares cros-clase emergentes y R-2 para pares intra-clase emergentes. En la corrida 10 las activaciones de las unidades de respuesta R-1 y R-2 fueron intermedias ante los estímulos de prueba.

Este resultado apoya la hipótesis del sesgo verbal en el procedimiento con humanos y sugiere la siguiente predicción: cuando se entrenan dos clases estímulo de tres miembros, en un procedimiento de estímulos compuestos, el uso de respuestas arbitrarias (i.e., sin una historia de aprendizaje explícita en tareas de categorización) genera un patrón de respuesta heterogéneo durante las pruebas de relaciones emergentes.

Experimento 2

Para poder evaluar y dar apoyo empírico a la predicción generada a partir del Modelo Neurocomputacional 1, se corrió un experimento con participantes humanos que fue similar al Experimento 1. Se trató de una réplica sistemática donde una de las diferencias más importantes fue, que en vez de utilizar respuestas “si/no”, se utilizaron respuestas sin carga verbal, particularmente se utilizaron como opciones de respuesta un cuadrado morado en vez de la opción “sí” y un cuadrado café en vez de la opción “no”.

Método

Participantes. Participaron seis estudiantes sin experiencia en EE, un hombre y cinco mujeres con edades entre 19 y 20 años. Fueron numerados como participantes 7 a 12. El participante hombre fue el 10. Obtuvieron créditos en una materia del semestre que cursaban independientemente de su ejecución.

Aparatos y Estímulos. Los estímulos que conformaron las clases ABC-1 y ABC-2, el escenario, material, software y consecuencias de reforzamiento, fueron los mismos que los utilizados en el Experimento 1.

Procedimiento. Se utilizó un procedimiento similar al Experimento 1, las diferencias fueron las siguientes: cada participante se corrió en una sola sesión, a diferencia del Experimento 1 donde la primera sesión fue de entrenamiento y la segunda de reentrenamiento y prueba. No hubo fase de reentrenamiento. Las opciones de respuesta fueron un cuadro morado en vez de la opción “si” y un cuadro café en vez de la opción “no”. El procedimiento estuvo dividido en tres fases. La primera fue de entrenamiento de relaciones de línea base, la segunda de pruebas de simetría y la tercera de pruebas de transitividad junto con ensayos de evaluación de mantenimiento de línea base.

Fase 1: Entrenamiento de Relaciones de Línea Base. Durante esta fase los participantes aprendieron a responder con la opción de respuesta morada ante los pares intra-clase A1B1, B1C1, A2B2 y B2C2; y con la opción de respuesta café ante los pares cross-clase A1B2, B1C2, A2B1 y B2C1. Dichas relaciones las aprendieron a través de seis etapas que tuvieron la misma cantidad de ensayos y criterios que las seis etapas de entrenamiento de la Fase 1 del Experimento 1.

Al inicio de la Fase 1 los participantes leyeron las siguientes instrucciones:

¡Gracias por tu participación!

El siguiente estudio no es una prueba de inteligencia y no evaluará ningún aspecto de tus habilidades intelectuales, al finalizar recibirás una explicación

Un investigador permanecerá cerca por si surge algún problema técnico, pero en ningún momento podrá darte ayuda para resolver la tarea.

Tu meta es obtener la mayor cantidad de puntos, estos puntos aparecerán en un contador ubicado en la esquina superior izquierda de la pantalla. En el centro de la pantalla aparecerán dos figuras, en la parte de abajo aparecerán dos opciones de respuesta, un cuadrado morado y uno café, tu

tarea es aprender a relacionar las dos figuras del centro con una de las opciones de respuesta; la morada o la café. Podrás aprender cuáles figuras están relacionadas y cuáles no poniendo atención a los cambios en el contador de puntos.

La tarea parece sencilla pero aumentará su dificultad conforme avance así que pon atención.

Por favor repite estas instrucciones al experimentador y cuando él te lo indique toca el botón de “Avanzar”.

Fase 2. Pruebas de Simetría. Se evaluaron las respuestas ante nuevas configuraciones de los estímulos compuestos al presentarlos en una ubicación espacial inversa a la del entrenamiento. Cada uno de los 8 ensayos de simetría se presentó 4 veces para lograr un total de 32 ensayos.

Fase 3. Pruebas de Transitividad. Se evaluó la emergencia de respuestas en las pruebas de transitividad para A1C1, A1C2, A2C2, A2C1, C1A1, C2A1, C2A2 y C1A2, cada uno de estos ensayos se presentó ocho veces, y se presentó cada uno de los ensayos de línea base dos veces, dando un total de 80 ensayos en esta fase.

Resultados y Discusión

Los seis participantes completaron el entrenamiento como puede observarse en la Figura 8. Mostraron un número mínimo de errores durante las primeras cuatro etapas, pero durante la Etapa 5, cuando se presentaron bloques de entrenamiento con todas las relaciones aprendidas, los participantes requirieron más repeticiones para lograr el criterio. Por ejemplo, los participantes 11 y 12 requirieron de 216 ensayos en 9 bloques para completar la Etapa 5.

Cuando se comparan los resultados de adquisición de la Fase 1 entre los participantes del Experimento 1 y el Experimento 2, donde la única diferencia fue el tipo de respuesta, se encuentra una diferencia estadísticamente significativa en el total de bloques requeridos para completar la Etapa 5 (Wilcoxon test, $W = 25$, $p = 0.02$ para $n_1 = n_2 = 6$ participantes); cuatro de

los seis participantes que respondieron con “si/no” requirieron menos bloques que cualquier participante que respondió con colores. Esto hace suponer que el uso de respuestas arbitrarias con etiquetas verbales facilita el aprendizaje de relaciones-estímulo.

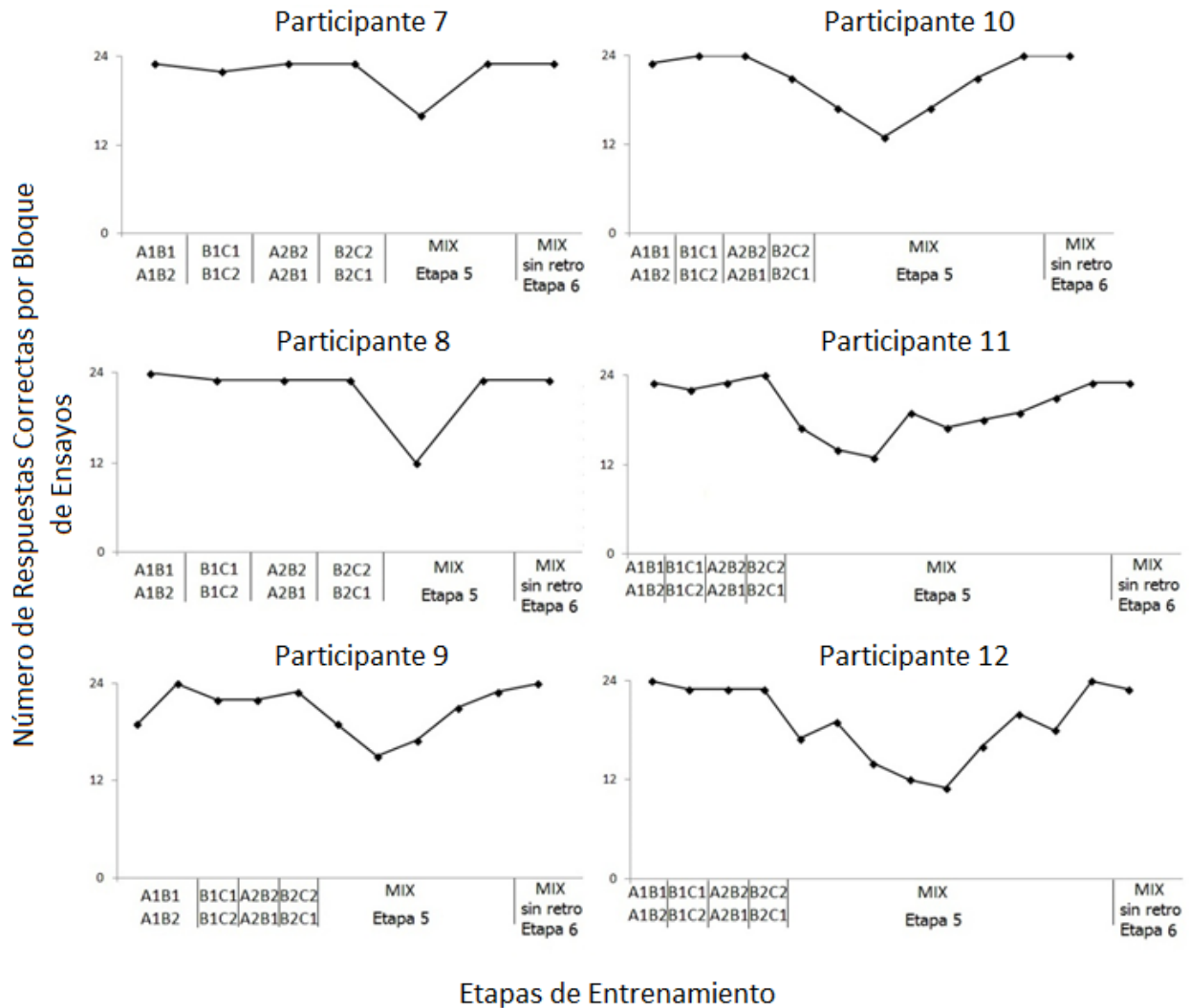


Figura 8. Cada punto muestra la cantidad de ensayos correctos por bloque para cada una de las seis etapas de entrenamiento de la Fase 1.

En la Tabla 6 se compara el número de bloques requerido por cada participante para completar la Etapa 5 y para completar todas las etapas de la Fase 1 en los Experimentos 1 y 2.

Ejecución durante las Pruebas de Simetría. Cinco de 6 participantes; los participantes 7 a 10 y el Participante 12, mostraron al menos 90% de ensayos correctos durante las pruebas de simetría, es decir, para ellos no importó la configuración perceptual del par sino los elementos que lo componían. Estos datos resultan similares a los correspondientes del Experimento 1, donde 6 de 6 participantes respondieron con altos niveles de exactitud en las pruebas de simetría.

Tabla 6. Para cada participante de los experimentos 1 y 2 se muestra la cantidad de bloques requeridos para cumplir el criterio en la Etapa 5 y en todas las etapas de la Fase 1 del entrenamiento.

Experimento 1			Experimento 2		
Participantes	Bloques Requeridos en la Etapa 5	Bloques Requeridos en la Fase 1	Participantes	Bloques Requeridos en la Etapa 5	Bloques Requeridos en la Fase 1
1	5	12	7	2	7
2	2	8	8	2	7
3	1	6	9	5	11
4	1	6	10	5	10
5	1	7	11	9	14
6	1	6	12	9	14

Ejecución durante las Pruebas de Transitividad. Los participantes pudieron formar clases de distinta forma, ya que tanto la respuesta morada como la café podían ser seleccionadas en los ensayos de transitividad debido a las dos trayectorias de relaciones-estímulo establecidas durante la Fase 1. Para considerar que su ejecución mostraba un proceso de formación de clases se consideraron los siguientes criterios: 1. Lograr al menos 85 % de ensayos correctos de mantenimiento de línea base; 2. Para los pares emergentes de prueba, responder al menos a 85% de los ensayos usando la opción morada para pares intra-clase y la opción café para pares cros-

clase, o responder al menos a 85% de los ensayos con la opción morada para pares cross-clase y la opción café para pares intra-clase. Un resumen del segundo criterio se presenta en la Tabla 7.

Tabla 7. Opciones de respuesta para satisfacer el segundo criterio de formación de clases estímulo.

Estímulos Compuestos para Evaluar Transitividad	Opción 1 de Formación de Clases	Opción 2 de Formación de Clases
A1C1	morado	café
C1A1	morado	café
A2C2	morado	café
C2A2	morado	café
A1C2	café	morado
C2A1	café	morado
A2C1	café	morado
C1A2	café	morado

Considerando los criterios mencionados, ningún participante formó clases estímulo. En la Figura 9 se muestra el número de respuestas a morado y a café para cada uno de los 8 ensayos de prueba. No hay constancia en las respuestas a lo largo de los ensayos, los participantes usaron la opción morada y café ante el mismo par de prueba en distintos ensayos. Las excepciones más sobresalientes son los participante 9 y 11; el Participante 9 respondió a todos los ensayos A1C1 y C1A1 con la opción café; ante esos mismos pares el Participante 11 respondió 14 de 16 ensayos con café. Sin embargo ambos participantes respondieron al resto de las relaciones sin un patrón claro. La Fase 2 mostró que los participantes 7, 8, 9, 10 y 12 pudieron responder a ensayos de simetría sin necesidad de instrucciones o retroalimentación; sin embargo, ellos no mantuvieron un criterio de simetría cuando se les presentaron nuevos pares estímulo. Por ejemplo, el Participante 10 respondió a todos los ensayos A1C2 con la opción café y a todos los ensayos C2A1 con la opción morada.

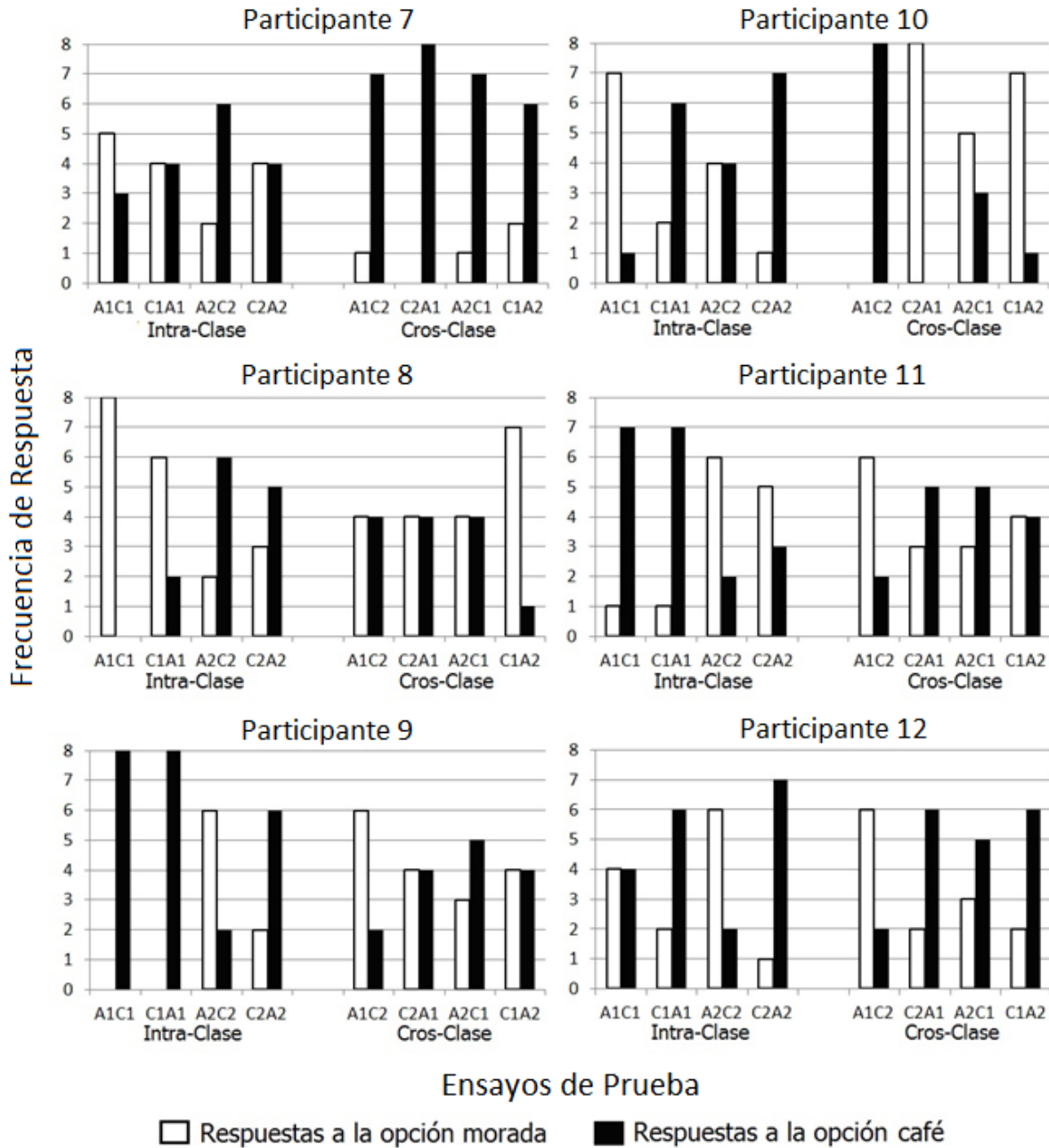


Figura 9. Número de respuestas con la opción morada y con la opción café para cada una de las ocho relaciones evaluadas en los seis participantes del Experimento 2.

Los resultados del Experimento 2 apoyan la predicción generada a partir el Modelo Neurcomputacional 1; el uso de respuestas arbitrarias generó un patrón de respuesta heterogéneo

durante las pruebas de transitividad. Con el procedimiento que implementa las opciones de respuesta arbitrarias ningún participante logró formar clases de estímulos equivalentes

Si las contingencias de reforzamiento permiten distintas formas de categorizar y los participantes no derivan una regla consistente con todos los casos que aprenden, entonces no se manifiesta la emergencia de relaciones entre estímulos que demuestre un proceso de categorización. Esto apunta a considerar que bajo condiciones de ambigüedad en las relaciones entre estímulos, la derivación de relaciones nuevas no es únicamente una cuestión de recombinación de estímulos, sino que se requiere de la generación de una regla consistente; las etiquetas verbales en las opciones de respuesta pueden ayudar a generar ese tipo de reglas.

Resulta interesante notar que la formación o no formación de clases estímulo en los experimentos 1 y 2, en el Modelo 1, así como en los estudios de Campos, Debert, Da Silva Barrios y McIlvane (2011), Debert et al. (2007) y Fields et al. (2009), parecen mostrar una interacción entre las contingencias de reforzamiento, la experiencia previa en tareas de categorización similares y los mediadores verbales. Estas tres variables están en relación con las tres propuestas teóricas respecto a la formación de clases revisadas en el Capítulo 1, la propuesta de reforzamiento de Sidman (2000), los Marcos Relacionales de Hayes (Hayes et al., 2002) y la propuesta de nombramiento de Lowe (Dugdale & Lowe, 1990; Horne & Lowe, 1996). El reforzamiento es esencial para aprender cómo se asocian los distintos estímulos, la experiencia previa en categorización de estímulos es aplicada cuando los participantes responden usando las opciones “sí” y “no” y estas respuestas podrían promover el uso de descripciones intraverbales que generen reglas para la formación de clases.

Las investigaciones futuras deberán analizar si existen influencias similares en procedimientos con estímulos compuestos y respuestas *go/no-go*, si bien este sistema de

respuesta está libre de etiquetas verbales, las instrucciones verbales asociadas a su uso pueden afectar la formación de clases. Por ejemplo, en un estudio reciente en pichones se observó simetría pero no transitividad (como en el Experimento 2) cuando dos clases de tres miembros fueron entrenadas con estímulos compuestos y respuestas *go/no-go* (Campos et al., 2011), el entrenamiento utilizado pudo generar la línea base ambigua descrita aquí. Cuando los humanos responden con un sistema tipo *go/no-go* (e.g., Debert et al., 2007) las instrucciones verbales pueden determinar la formación de clases, pero en el caso de los pichones, la imposibilidad de usar instrucciones verbales, y la ausencia de una historia extra-experimental asociada con las respuestas *go/no-go* puede llevar a las fallas en las pruebas de transitividad.

Modelo Neurocomputacional 2

Con el Modelo Neurocomputacional 1 se pudo determinar que los modelos asociativos multicapa con aprendizaje por retropropagación son capaces de simular formación de clases de equivalencia, pues demuestran el establecimiento de relaciones entrenadas y derivadas como lo hacen los humanos. La simulación requirió el uso de un procedimiento con estímulos compuestos y dos opciones de respuesta; se argumentó que el tipo de opciones de respuesta genera un sesgo sobre la formación de clases. Sin embargo, dicho modelo tiene algunas limitaciones que impiden extender su uso para abarcar los objetivos de esta tesis. Las principales limitaciones son: 1. No hay ningún valor cuantitativo que directamente hable del grado de relación entre los estímulos que conforman una clase. 2. No es posible observar algún indicador cuantitativo que hable de efectos de nodalidad en EE. 3. Resulta difícil simular varios estudios existentes de EE que están diseñados para procedimientos de IM, por lo cual varias de las conclusiones que se puedan derivar serán aplicables únicamente a procedimientos con estímulos compuestos. Y 4. Los

algoritmos de retropropagación han sido criticados por ser poco plausibles biológicamente, no hay evidencia a nivel biológico que indique que una señal de error puede ser conducida hacia atrás para modificar la conectividad hacia delante de futuras iteraciones (O'Reilly & Munakata, 2000, p. 162).

Por estas razones se desarrolló un modelo donde se pudiera simular el aprendizaje de clases con un procedimiento tanto de IM como de estímulos compuestos. Se consideró una arquitectura parsimoniosa; una sola capa de neuronas donde las neuronas representen a los estímulos pertenecientes a las clases y donde el valor de las conexiones que hay entre dichas neuronas represente el grado de relación que hay entre los estímulos, se implementó un algoritmo de aprendizaje Hebbiano biológicamente plausible.

Una consideración derivada del Modelo Neurocomputacional 1, es que los estímulos de una clase se vuelven equivalentes pues logran generar la misma representación, en dicho modelo la representación simplemente es una transformación de los vectores de entrada en los vectores de la capa intermedia, para el Modelo Neurocomputacional 2 se toma la idea de la representación común a una clase, pero considerando que la naturaleza de dicha representación resulta del fortalecimiento de conexiones entre representaciones particulares para generar pequeñas asambleas de neuronas que corresponden con las categorías.

Los supuestos básicos con los que se diseñó, y a partir de los cuales se analiza la ejecución del Modelo Neurocomputacional 2 son los siguientes:

1. *El nivel de representación es localista*; es decir cada unidad representa a un estímulo. Dado que en EE los estímulos son elegidos generalmente para minimizar los traslapes perceptuales las representaciones distribuidas no son necesarias.

2. *Las conexiones entre neuronas incrementan o disminuyen siguiendo un aprendizaje de tipo Hebbiano supervisado*; es decir, este modelo es sensible a las frecuencias de co-ocurrencia de los estímulos en el ambiente y a las consecuencias (i.e., reforzamiento) que se presentan tras categorizarlos.

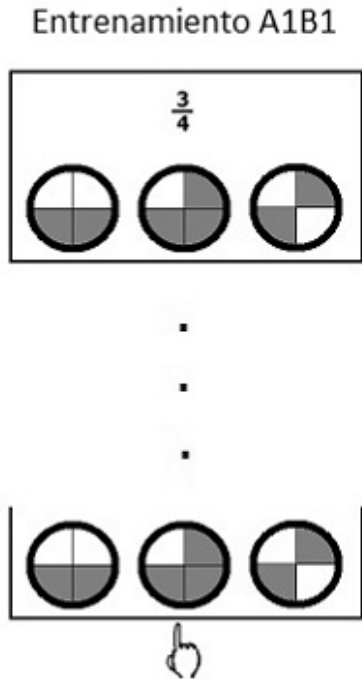
3. *El valor de la conexión entre dos neuronas (e.g., A y B) representa el valor con el que los estímulos (A y B) están relacionados*, dado que el nivel de relacionalidad de los estímulos se considera dependiente de la frecuencia y el reforzamiento con la que han sido experimentados.

De estos supuestos se deriva que, si la exactitud y velocidad de respuesta en participantes humanos para tareas de EE dependen de la fuerza de asociación o nivel de relación entre estímulos, entonces *El valor de conexión observado en el modelo mantiene una correspondencia directa con la exactitud y la velocidad de respuesta de participantes humanos.*

Arquitectura

El modelo consta de una sola capa de neuronas completamente interconectadas (ver Figura 10). La activación de las unidades ($X_i, X_j \dots X_n$) representa a los distintos estímulos y los valores de las conexiones entre unidades ($W_{ij}, \dots W_{mn}$) representan el nivel de relación entre los estímulos. Las unidades se pueden activar por uno de dos posibles mecanismos: por exposición a patrones de estimulación, o por la activación que fluye desde otras unidades. Los valores de activación están entre 0 y 1; 0 = no activación, 1 = máxima activación. La forma de trabajo de este modelo será explicada con las siguientes dos etapas: Interacción y Aprendizaje; considerando la simulación de ensayos de IM.

Estimulación vía ensayos de igualación a la muestra



Capa neuronal con conectividad completa

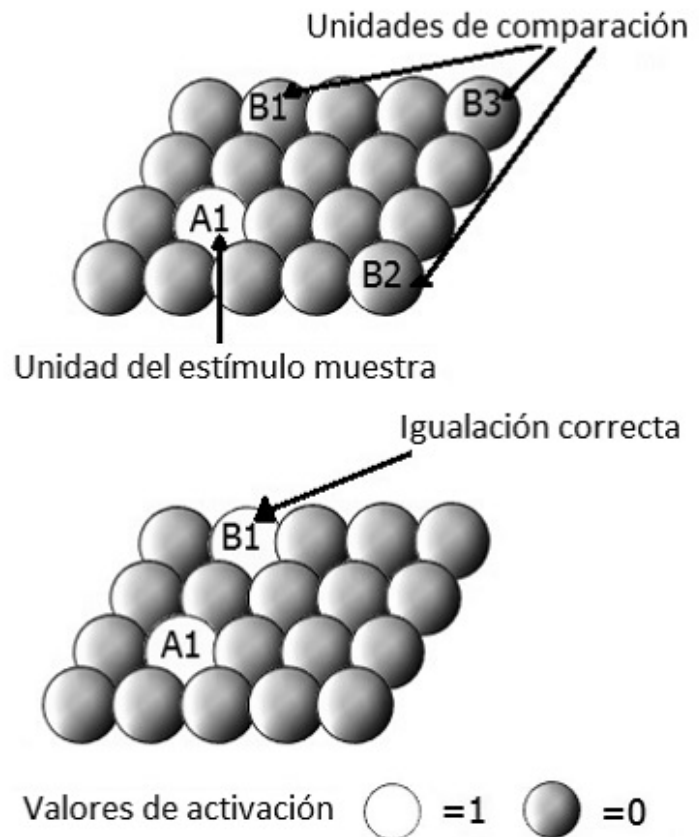


Figura 10. Arquitectura del Modelo Neurocomputacional 2. Del lado izquierdo se muestra un ejemplo de un ensayo de entrenamiento A1B1 utilizado con participantes humanos. Del lado derecho se muestra la arquitectura y una representación gráfica de la activación en la simulación del entrenamiento de A1B1.

Etapa de Interacción

En esta etapa se presentan patrones de estimulación que representan ensayos de IM, estos se consideran como información ambiental. Por cada ensayo hay dos tipos de estímulo que pueden presentarse; estímulos de muestra y estímulos de comparación (de forma simple o como estímulos compuestos). Cuando un estímulo muestra es presentado, por ejemplo A1, la unidad

que lo representa automáticamente se activa y propaga su activación hacia el resto de las neuronas dependiendo de la conexión que mantiene con cada una. Cuando los estímulos de comparación se presentan, por ejemplo B1, B2 y B3, el modelo selecciona uno de ellos como la respuesta, esta selección depende de cuál estímulo presenta el mayor nivel de activación.

Las unidades de los estímulos de comparación computan el input que reciben desde la unidad del estímulo muestra, de tal forma que aquella unidad que tenga la conexión más fuerte con la unidad del estímulo muestra será activada como la respuesta. Solamente una de las unidades de comparación será seleccionada y quedará activada. El resto de las unidades permanecerán apagadas (ver Figura 10). Este procedimiento simula competencia implementada a través de inhibición lateral entre las unidades de comparación.

Los valores de las conexiones comienzan en 0 para simular que no existe experiencia previa de las relaciones entre estímulos, así que durante los primeros ensayos el modelo selecciona el estímulo de comparación de forma aleatoria. Durante el entrenamiento, las conexiones pueden adquirir valores entre -1 y 1; los valores de 0 representan ausencia de propagación de actividad y se interpreta que no hay relación entre los estímulos representados en las unidades conectadas, los valores de 1 son interpretados como el nivel de relación más elevado entre dos estímulos, los valores negativos (inhibición) resultan de selecciones incorrectas de estímulos de comparación. Las unidades suman todos los inputs y generan un valor de activación con las ecuaciones 1 y 2.

Etapas de Aprendizaje

Tras cada selección de un estímulo de comparación (respuesta) el modelo es informado acerca de su ejecución. De forma similar a como ocurre con participantes humanos en IM. En un

experimento con humanos, si la selección de B1 es seguida de reforzamiento cuando A1 es la muestra, entonces los participantes pondrán atención a B1 en presencia de A1 lo que llevará a estos dos estímulos a producir representaciones concurrentes, mientras que el resto de los estímulos B serán ignorados.

La conexión que vincula a las unidades coactivas incrementa su valor siguiendo el principio de aprendizaje Hebbiano, y eventualmente se vuelve lo suficientemente fuerte como para permitir que la activación de una unidad dispare la activación de la otra. Así que, tras percibir por ejemplo el estímulo A1, su activación se propagará para activar la representación del estímulo B1 y viceversa. Las respuestas incorrectas llevan al decremento en el valor de la conexión de las unidades coactivas.

Como resultado del proceso de aprendizaje descrito, las conexiones entre unidades son simétricas, lo que implica que la relación de A1 a B1 se interpreta con el mismo valor que la relación de B1 a A1.

La formalización del aprendizaje Hebbiano se expresa en la ecuación 3.

$$(3) W_{ij}(t+1) = W_{ij}(t) + \beta (X_i * X_j)$$

Donde $W_{ij}(t+1)$ es el valor que la conexión entre la neurona i y la neurona j adquiere tras añadir a su estado previo, en (t), el resultado de la actividad concurrente entre i y j; β es la tasa de aprendizaje.

Nótese que si la ecuación 3 se implementa en un proceso continuo, el resultado es un incremento infinito en W_{ij} , dado que cada coactivación de X_i y X_j generará un incremento en la conexión lo cual llevará a mayor coactivación en un loop reiterativo.

Para controlar dicho crecimiento, se considera un valor máximo para W_{ij} que dependerá del valor máximo de coactivación para cada iteración, igualmente se considera que cada cambio en W_{ij} tomará lugar únicamente cuando el valor de la coactivación haya superado un umbral (θ).

La cantidad de incrementos o decrementos en W_{ij} durante cada iteración depende de la relación entre W_{ij} y el valor de la coactivación; λ computa esta relación al substrair el valor de W_{ij} del valor de la coactivación.

En la ecuación 4 se muestra cómo se calcula λ y en la ecuación 5 se muestra la actualización del algoritmo de aprendizaje Hebbiano.

$$(4) \quad \lambda = (X_i * X_j) - W_{ij}(t)$$

$$(5) \quad W_{ij}(t+1) = W_{ij}(t) + \lambda [\beta (X_i * X_j \geq \theta)]$$

Nótese que tras varias presentaciones de un par de estímulos, por ejemplo A1B1, la conexión entre la neurona que representa A1 y la neurona que representa a B1 será fuerte (i.e., cercana a 1). Considérese el caso de entonces presentar el estímulo A1 solo en el patrón de estimulación (ambiente); su activación podrá fluir vía la conexión entre A1 y B1 y entonces B1 se activará; sin embargo, dado que A1 y B1 no están juntos en el ambiente, no hay razón para que la conexión entre ellos se fortalezca, es decir, esta relación no está siendo reforzada. En esos casos λ toma un valor negativo (ver ecuación 4) y la conexión entre A1 y B1 es debilitada. Sin embargo, no resulta adaptativo el aprender que dos cosas están relacionadas y desperdiciar dicho aprendizaje en cuanto una de ellas aparece sola; dada esta consideración, la tasa de cambios en las conexiones, dada por el parámetro β , que siga a valores de λ negativos será menor que la tasa de incrementos cuando λ tiene un valor positivo. Para las simulaciones se determinó empíricamente que los decrementos en W_{ij} que resulten de una λ negativa serán al 25% de los incrementos.

El balance entre incrementos y decrementos lleva a estabilizar a W_{ij} de tal forma que en los valores de las conexiones se almacena la información respecto a la ocurrencia y co-ocurrencia de estímulos. Las conexiones entonces deberían generar una descripción cuantitativa de las regularidades ambientales. Una simulación de aprendizaje entre A y B, y B y C muestra que las conexiones modifican sus valores con incrementos iniciales rápidos y tardíos lentos, como describe la curva tradicional de aprendizaje, ver Figura 11.

Plausibilidad Biológica del Algoritmo de Aprendizaje

Existe evidencia neurofisiológica de la plausibilidad del aprendizaje Hebbiano, principalmente observada en los experimentos de potenciación a largo plazo (LTP) y depresión a largo plazo (LTD), donde se reporta que la eficacia sináptica depende de la historia de activación de las redes neuronales biológicas. Tras una estimulación de alta frecuencia en un grupo de neuronas que está conectado con un segundo grupo se observa LTP, entendida como la intensificación en la propagación de activación; en cambio, al utilizar estimulación de baja frecuencia se observa una reducción en la eficacia sináptica de un grupo de neuronas sobre el otro grupo, esto es descrito como LTD (Bliss, Collingridge, & Morris, 2007; Bear, 2003; Malenka & Bear, 2004).

La estabilidad lograda por W_{ij} en el algoritmo también corresponde con cómo los cambios de dirección y magnitud en las conexiones sinápticas son combinados para lograr un equilibrio o *set-point* en las redes neuronales biológicas; se ha indicado que la plasticidad Hebbiana es regulada en neuronas biológicas por un balance entre LTP y LTD, así como por aspectos bioquímicos del funcionamiento de la red que la modulan y mantienen valores

homeostáticos (ver Turrigiano & Nelson, 2000 para una descripción detallada del equilibrio sináptico).

El valor de la diferencia entre coactivación y W_{ij} calculado en λ hace posible que el mismo tipo de estimulación conlleve a potenciación o depresión, dependiendo de las últimas experiencias con los estímulos, este tipo de dinámica ha sido sugerida en modelos de modificación sináptica biológica (e.g., Bienestock, Cooper & Munro, 1982).

El límite inferior de coactivación descrito (umbral de coactivación, θ), que debe ser superado para que el fortalecimiento de las conexiones tenga lugar, asegura que los incrementos en las conexiones se den únicamente tras valores elevados de coactivación, tal cual ocurre en las sinapsis biológicas tras la estimulación de alta frecuencia (Froc & Racine, 2005).

Estas consideraciones cobran relevancia para comprender las posibles extensiones del modelo. En la simulación 3 se presentará un caso de simulación de deficiencias cognitivas, para dicha simulación los parámetros del modelo son ajustados de acuerdo a la evidencia neurofisiológica disponible.

Simulaciones

Se corrió una simulación del aprendizaje de las relaciones AB y BC para determinar los parámetros θ y β , y para observar cómo incrementa el nivel de relación entre los distintos estímulos de la clase. Luego se simularon tres estudios clásicos en EE; primero se simuló el estudio de Sidman y Tailby (1982), luego el de Devany, Hayes y Nelson (1986), y finalmente el de Spencer y Chase (1996). Se seleccionaron estos tres estudios por las siguientes razones: 1. Han tenido un alto impacto en la literatura de EE. 2. En ellos se dan descripciones detalladas de los procedimientos que permiten replicas. 3. Los resultados que publicaron los estudios

originales pueden ser comparados con los valores de las conexiones en el Modelo Neurocomputacional 2.

Los objetivos específicos de estas simulaciones fueron evaluar la capacidad del modelo para aprender y derivar relaciones-estímulo, esto fue abordado en las simulaciones 1 y 2. El segundo objetivo específico fue simular participantes con deficiencias cognitivas para replicar fallas en las respuestas a ensayos de transitividad, esto fue abordado en la Simulación 3. El tercer objetivo específico, abordado en la Simulación 4, fue determinar la capacidad del modelo para simular efectos de distancia nodal.

Todas las simulaciones fueron corridas en Matlab 2011a. Los estímulos simulados fueron presentados al modelo como patrones de números binarios. Para los ensayos de entrenamiento se utilizó retroalimentación informativa (reforzamiento), para comparar al estímulo seleccionado por el modelo como respuesta con el estímulo de comparación correcto. Durante los ensayos de prueba no se presentó retroalimentación informativa.

Simulación 1. Aprendizaje de AB y BC

Se presentaron al modelo patrones de estimulación que representaron el tipo de relaciones mostradas en la Figura 1. Dos relaciones-estímulo fueron directamente entrenadas: A1B1 y B1C1. El entrenamiento de A1B1 comenzó con la presentación del estímulo de muestra A1 y los estímulos de comparación B1, B2 y B3. Tras la activación de la unidad de respuesta el modelo fue presentado con la retroalimentación informativa y los valores de las conexiones se modificaron de acuerdo al algoritmo descrito. Después se presentó el siguiente ensayo con el estímulo B1 como muestra y los estímulos C1, C2 y C3 como comparaciones. El entrenamiento continuó hasta que se completaron 30 repeticiones de cada tipo de relación.

Resultados y Discusión

La red mostró aprendizaje de las relaciones entrenadas mediante incrementos en los valores de conexión entre A1 y B1, y B1 y C1 (Figura 11). En un punto de la simulación (cerca del ensayo 11) las repeticiones de la relación A1B1 llevaron a la red a evocar cualquier estímulo cuando el otro era presentado, de tal forma que la unidad de A1 quedó activada cuando se presentó el ensayo B1C1. Esto resultó en la activación concurrente de A1, B1 y C1 tras cada presentación de cualquier ensayo que contuviera B1, dándose así incrementos en las conexiones entre A1 y C1, pese a que estos estímulos no se presentaran juntos en ningún ensayo. El nivel de relación entre A1 y C1 siempre fue menor al nivel alcanzado por las relaciones entrenadas, como puede notarse en la Figura 11

Tras esta simulación se estableció el valor de los parámetros del algoritmo de la siguiente manera: umbral de coactivación (θ) = 0.7, y tasa de aprendizaje (β) = 0.2. Con estos parámetros el modelo mostró estabilidad en las conexiones tras 25 repeticiones de cada ensayo de entrenamiento.

El resultado principal de esta simulación fue dar cuenta de la emergencia de relaciones transitivas en clases de estímulos equivalentes. La simulación también mostró que el algoritmo de aprendizaje genera cambios en las conexiones con incrementos iniciales rápidos y tardíos lentos, como lo describen las curvas de aprendizaje tradicionales. Esta dinámica de modificación se debe al rango dinámico de fortalecimiento establecido por λ . Durante los primeros ensayos hay una gran diferencia entre la coactivación y el peso de la conexión, de tal forma que el cambio en el peso es rápido, a lo largo de los ensayos esta diferencia se reduce llevando progresivamente a modificaciones más pequeñas en las conexiones.

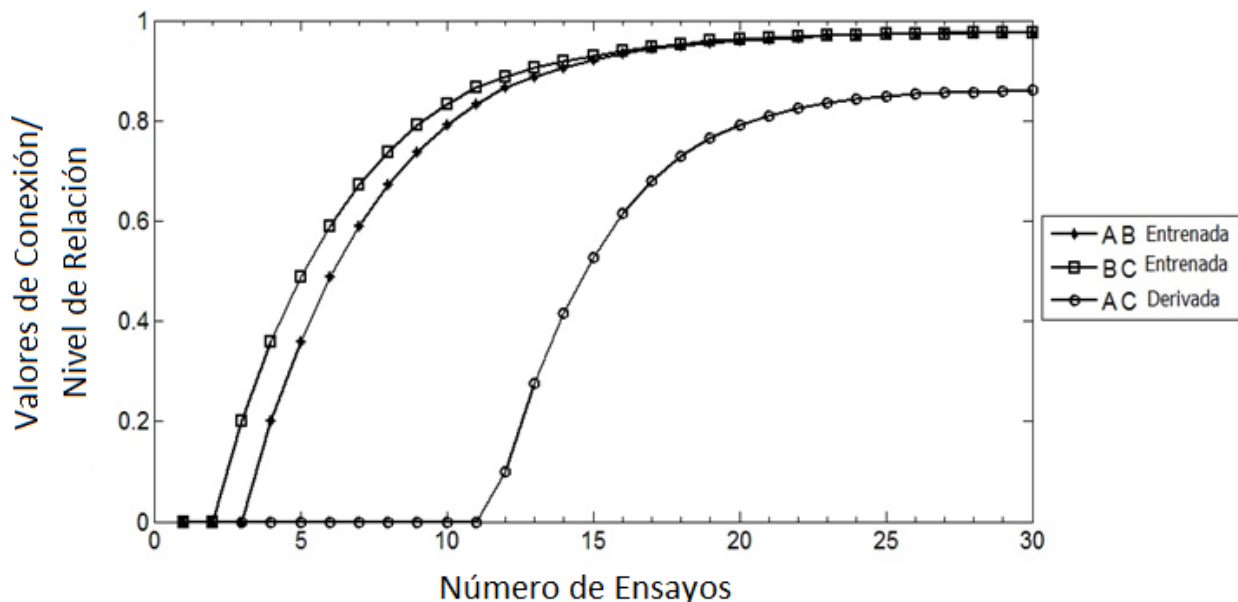


Figura 11. Muestra los incrementos en el valor de las conexiones para las relaciones entrenadas AB y BC, y para la relación derivada AC.

Simulación 2. Sidman y Tailby (1982)

Los primeros estudios de EE trabajaban con clases de tres estímulos. El propósito principal del estudio original fue analizar la inclusión de un cuarto estímulo en una clase de estímulos equivalentes.

Los participantes fueron ocho niños con desarrollo normal. Los estímulos fueron un conjunto A de letras griegas dictadas; y los conjuntos B, C y D con diferentes letras griegas impresas. Un total de tres clases de cuatro miembros fueron entrenadas con un procedimiento de IM. Los niños y el modelo aprendieron las relaciones AB, AC y DC y fueron evaluados en las relaciones DB, BD, AD, BC, CB y CD. En el Apéndice A se muestra la secuencia de ensayos de entrenamiento y prueba usados en el estudio original y en el Modelo Neurocomputacional 2.

Los niños del estudio original aprendieron primero las relaciones AB y AC con ensayos de IM auditivo-visuales. Fueron reforzados por elegir las letras griegas impresas de los conjuntos B y C que correspondían con las letras griegas dictadas del conjunto A. Después los niños

siguieron con ensayos de IM visuales-visuales para aprender las relaciones DC. En esta etapa fueron reforzados por elegir la letra del conjunto C que correspondía con la muestra del conjunto D. Las 9 relaciones-estímulo (AB, AC, DC, en las clases 1, 2 y 3) fueron entrenadas de forma progresiva, la secuencia de entrenamiento es mostrada en el Apéndice A. Un número mínimo de ensayos correctos era requerido en cada etapa del entrenamiento para poder continuar, estos criterios también se muestran en el Apéndice A.

Una vez que los niños alcanzaron el nivel de ejecución requerido mediante los criterios, la probabilidad de reforzamiento fue gradualmente disminuida para después presentar los ensayos de prueba de las relaciones derivadas DB, BD, AD, BC, CB, y CD. Durante el entrenamiento los ensayos de IM incluyeron una muestra y dos o tres estímulos de comparación; durante las pruebas todos los ensayos de IM incluyeron una muestra y tres estímulos de comparación.

Para igualar el entrenamiento del Modelo Neurocomputacional 2 con el entrenamiento del estudio original de Sidman y Tailby (1982), se diseñaron patrones de estimulación que correspondieron con las 9 relaciones entrenadas. Cada uno de estos patrones se presentó de acuerdo a la secuencia y número de repeticiones que reporta el estudio original. Por ejemplo, para simular la primera etapa de entrenamiento de las relaciones AB, se presentó al modelo un bloque de 20 ensayos, donde 10 correspondieron con la relación A1B1 y 10 correspondieron con la relación A2B2. Si la unidad de respuesta activada por el modelo fue la correcta en al menos 19 ensayos, como fue el criterio en el estudio original, entonces se dio paso al siguiente bloque de entrenamiento donde se presentaron 20 ensayos donde 10 correspondieron a la relación A1B1 y otros 10 a la relación A3B3, y así sucesivamente.

En el modelo no se simuló la etapa de disminución gradual de la probabilidad de reforzamiento; este procedimiento es usado en humanos para mantener un nivel de ejecución independiente de la presencia de reforzamiento durante las pruebas, donde no se da ninguna señal informativa sobre la ejecución de los participantes. En el modelo no hay necesidad de motivar las respuestas durante las situaciones de extinción.

Para evaluar las relaciones derivadas, en la simulación se presentó un bloque de ensayos con las 9 relaciones entrenadas y las 18 relaciones de prueba, todo esto en ausencia de reforzamiento.

Resultados y Discusión

En el estudio original 6 de los 8 niños respondieron demostrando formación de clases durante las fases de prueba. En el Panel A de la Figura 12 se muestra el porcentaje de las respuestas correctas para las relaciones entrenadas y derivadas de los seis participantes que formaron clases.

El modelo aprendió la tarea y respondió demostrando formación de clases durante la fase de prueba como se puede apreciar en el Panel B de la Figura 12, donde los valores de las conexiones logrados por el modelo son visualmente comparables con el porcentaje de respuestas correctas del Participante "A.D." La secuencia de ensayos de mantenimiento de línea base presentada durante la fase de prueba fue la misma en el modelo y en el Participante A.D. (dicha secuencia varió entre los seis participantes reportados). Hay incrementos de la relación AB a la AC a la DC tanto en el modelo como en el Participante A.D. En el modelo esto se generó por efecto de la secuencia de relaciones entrenadas; se evaluaron distintas secuencias generando distintas valores finales en las conexiones. Probablemente el Participante A.D. fue sensible a la

secuencia de ensayos de tal forma que ella pudo responder mejor a las últimas relaciones entrenadas, de forma similar a como ocurrió en el modelo.

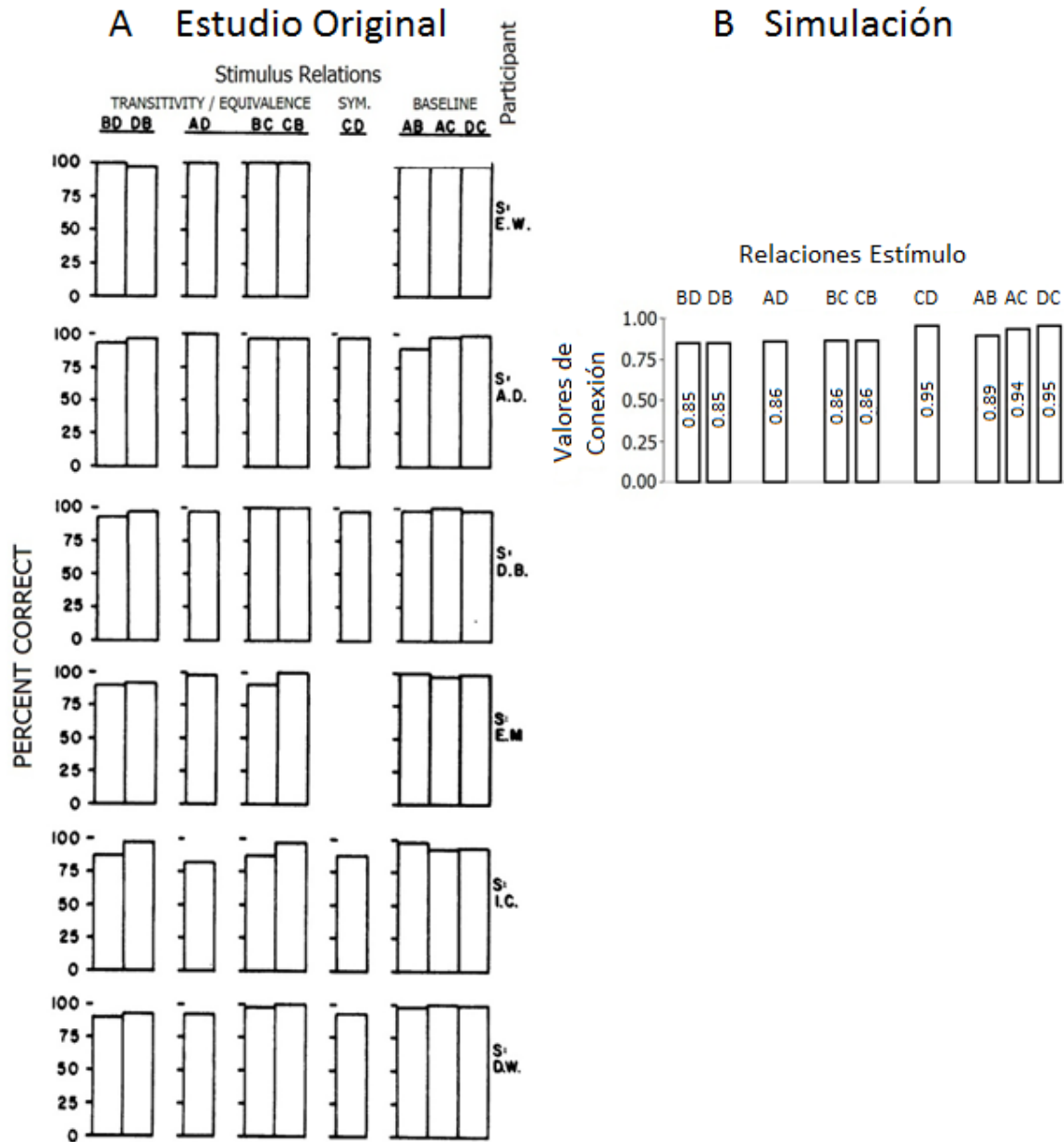


Figura 12. El Panel A es del estudio de Sidman y Tailby (1982). Cada línea de barras da el porcentaje de respuestas correctas de un niño a ensayos de transitividad/equivalencia, simetría y línea base. El Panel B muestra los valores de conectividad finales en el Modelo Neurocomputacional 2 para las relaciones derivadas y de línea base.

El principal resultado de la Simulación 2 fue demostrar que cuando un entrenamiento específico de EE es simulado, los valores de las conexiones que dan cuenta del aprendizaje en la red pueden ser comparados con la exactitud de respuesta de los participantes humanos. En este caso los valores de las conexiones son similares a la exactitud reportada de 6 niños participantes con desarrollo normal.

Simulación 3. Devany, Hayes y Nelson (1986)

Las ejecuciones en tareas verbales y en EE parecen estar correlacionadas. Devany y colaboradores (1986) mostraron uno de los primeros estudios que develó esta correlación. Su principal objetivo fue determinar si había diferencias entre las ejecuciones de EE de niños con desarrollo normal y niños con deficiencias en el desarrollo. Un total de 12 niños participaron en este estudio. Como se mencionó en el Capítulo 1, los niños pertenecieron a uno de tres grupos: 1) Niños con desarrollo normal, 2) Niños con deficiencias de desarrollo y con algunas habilidades de lenguaje, 3) Niños con deficiencias de desarrollo y sin habilidades de lenguaje.

El resultado principal de este estudio fue que los niños que no tenían habilidades de lenguaje no pudieron responder correctamente a las pruebas de transitividad, mientras que los niños con habilidades de lenguaje, ya fueran con desarrollo normal o con deficiencias de desarrollo sí pudieron responder a los ensayos de transitividad.

Dado lo anterior, un objetivo particular de esta simulación fue evaluar las condiciones bajo las cuales el modelo no puede ejecutar respuestas correctas durante las fases de prueba. En el Modelo Neurocomputacional 2 no hay forma de simular la influencia de mayores o menores habilidades de lenguaje pues estas se consideran una salida o “outcome” de un sistema y no una parte del proceso; sin embargo, algunas ideas provenientes del perfil general neurofisiológico del “retraso mental” pueden ser implementadas. Los déficits de aprendizaje asociados al retraso

mental en modelos animales están relacionados con un nivel elevado de inhibición, responsable de los fallos en la inducción de LTP (Kleschevnikov et al., 2004). Así como una magnitud mayor de LTD (Scott-McKean & Costa, 2011; Huber, Gallagher, Warren & Bear, 2002). Se ha sugerido que un desequilibrio entre LTP y LTD puede resultar en un aumento en el umbral de estimulación que debe alcanzarse para producir LTP, de tal forma que los mecanismos neuronales asociados a la inducción de LTP son funcionales pero requieren de una mayor actividad neuronal para ser disparados (Meredith, Holmgren, Weidum, Burnashev, & Mansvelder, 2007).

El perfil neurofisiológico descrito fue simulado en el Modelo Neurocomputacional 2 considerando el aumento en el umbral para la producción de LTP, este umbral se considera análogo al umbral de coactivación (θ), que restringe el fortalecimiento de las conexiones a valores de coactivación elevados; se corrió el Modelo Neurocomputacional 2 con los parámetros regulares para simular las ejecuciones de niños con desarrollo normal; y después se corrió el modelo con un umbral de coactivación más elevado (θ pasó de 0.7 a 0.72); y con una tasa de aprendizaje más baja (β pasó de 0.2 a 0.1) para simular a los niños con deficiencias en el desarrollo.

Durante el entrenamiento, los niños y el modelo aprendieron las relaciones AB y AC de dos clases de estímulos vía un procedimiento de IM, en cada ensayo se presentó un estímulo muestra y dos de comparación. Durante las pruebas fueron evaluadas las relaciones transitivas BC y CB. Los estímulos del estudio original fueron figuras parecidas a animales impresas en hojas de papel. Las respuestas correctas de los niños del grupo 1 generaron reforzamiento con alagos, cantos y tiempo para soplar burbujas de jabón; los reforzadores para los niños de los

grupos 2 y 3 incluyeron un tiempo de flashes de luz, burbujas de jabón, globos, jugo o galletas con queso.

Durante las primeras 7 etapas de entrenamiento los ensayos de las distintas relaciones se introdujeron gradualmente en bloques de 10 ensayos. Tras 9 de 10 respuestas correctas consecutivas los niños avanzaban hacia las siguientes etapas. Las relaciones entrenadas en cada etapa se muestran en el Apéndice B. Tras las primeras 7 etapas de entrenamiento los niños pasaron a bloques con ensayos mezclados de todos los tipos de relación de entrenamiento donde la probabilidad de reforzamiento fue disminuyendo hasta que llegó a un nivel cercano al 25%.

En la fase de prueba las relaciones BC y CB para las clases 1 y 2 fueron evaluadas en un bloque de 40 ensayos. Cada tipo de relación se presentó en total 10 veces en una secuencia aleatoria. No se usó reforzamiento contingente con las respuestas a estos ensayos

Para la simulación se presentaron patrones de estimulación que correspondieron con las 4 relaciones de entrenamiento en la misma secuencia u el mismo número de ensayos propuestos en el estudio original. Cuando el modelo lograba 9 de 10 ensayos con la activación de la unidad de respuesta correcta se pasaba a la siguiente etapa de entrenamiento. Al igual que en la simulación previa no se implementó la disminución gradual de reforzamiento, en vez de ello se presentó un último bloque con todas las relaciones de entrenamiento sin retroalimentación informativa.

Durante las pruebas se presentaron 4 patrones de estimulación que representaron la evaluación de las relaciones BC y CB de las clases 1 y 2, estos fueron presentados en orden aleatorio y sin reforzamiento.

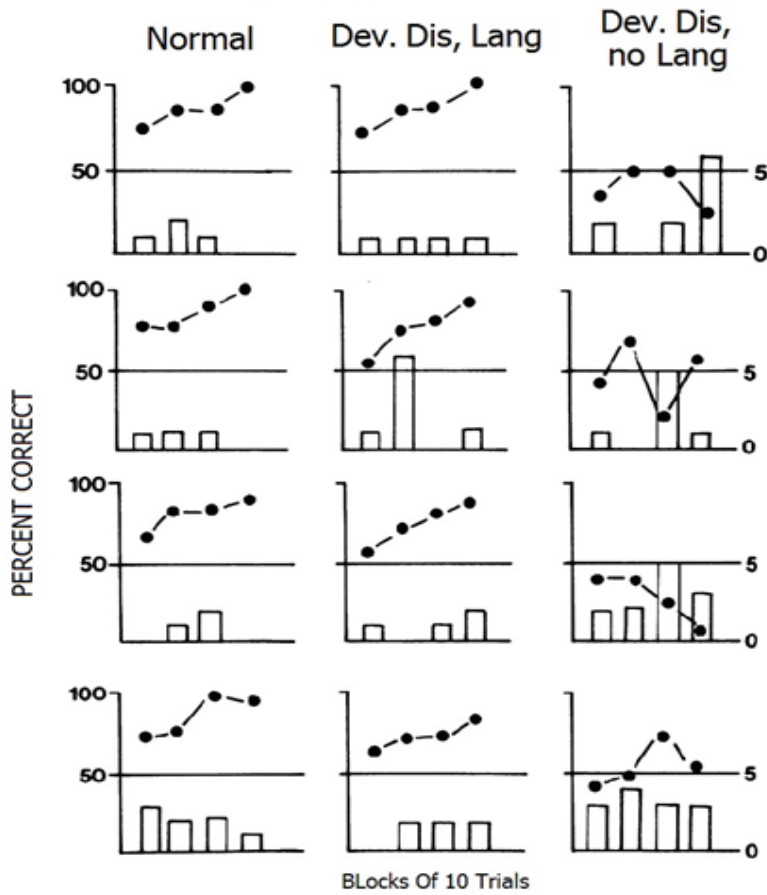
Resultados y Discusión

En la Figura 13 panel A se reproducen los resultados del estudio original durante las pruebas por cada uno de los 12 niños, cada gráfica muestra en las barras el número de respuestas no emitidas por los niños en grupos de 10 ensayos; los puntos muestran el porcentaje de respuestas correctas (dentro de todas las respuestas emitidas) en grupos de 10 ensayos. Los niños que formaron parte del grupo sin habilidades de lenguaje mostraron el nivel de ejecución más bajo, la mayoría de ellos respondió a nivel de azar durante las pruebas, mientras que los niños de los otros dos grupos respondieron de forma similar entre ellos acercándose a una ejecución sin errores conforme avanzaron los ensayos.

En el Panel B de la Figura 13, se muestra la representación gráfica de los valores de las conexiones para las relaciones entrenadas y derivadas cuando las condiciones de niños con desarrollo normal y niños con deficiencias en el desarrollo son simuladas.

En ambas condiciones las relaciones AB y AC fueron aprendidas, sin embargo, los valores de las conexiones son más fuertes en la simulación de niños normales que en la simulación de niños con deficiencias, lo cual resulta similar al estudio original, donde los niños normales aprendieron las relaciones de línea base en menos bloques que los niños con deficiencias. Puede notarse que la relación transitiva BC muestra virtualmente un nulo incremento en la simulación de niños con deficiencias, demostrando que no hay un proceso de inferencia transitiva aunque si hubo aprendizaje de las relaciones de línea base.

A Estudio Original



B Simulación

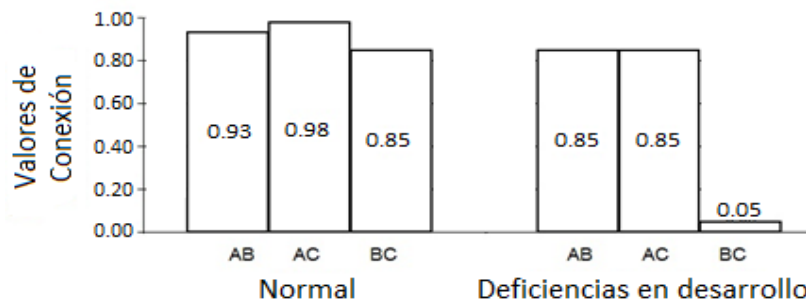


Figura 13. El Panel A es del estudio original de Devany y colaboradores (1986), las barras y los puntos representan los datos de los niños en los distintos grupos. Los puntos muestran el porcentaje de respuestas correctas y las barras muestran la cantidad de respuestas no emitidas en bloques de 10 ensayos. En el Panel B se muestran los valores de las conexiones logrados por el modelo para las relaciones entrenadas y derivadas en la simulación de niños normales y niños con deficiencias en desarrollo sin lenguaje.

La explicación mecanicista de la falla en transitividad en el modelo es la siguiente: el valor elevado del umbral de coactivación condujo a una limitada adaptación de las conexiones; la estimulación directa para los ensayos de entrenamiento generó una completa activación (i.e., activación = 1) y la coactivación resultante logró superar el umbral para que se modificaran los valores de las conexiones, de esta forma el modelo si aprendió las relaciones entrenadas. Sin embargo, las unidades activadas por un proceso de propagación (en vez de estimulación directa) alcanzan niveles de activación menores, de tal forma que en esta simulación no alcanzaron el umbral de coactivación. Es decir, una vez que AB y AC fueron entrenadas, las conexiones de A con B y C permitieron propagar la activación hacia cualquiera de estas neuronas en presencia de cualquier ensayo donde estuviera A, pero no fue posible fortalecer la conexión entre B y C (relación transitiva) dado que éstas no generaron coactivaciones por encima del umbral.

Una posible forma de superar esta situación en el modelo se describe a continuación: la activación generada como resultado de la propagación si puede generar valores de coactivación por encima del umbral toda vez que las conexiones resultantes del entrenamiento hayan adquirido valores cercanos a 1. Si la relación AB es la única relación entrenada, la conexión entre A y B alcanzará un valor de 1 (aun cuando el modelo trabaje con los parámetros de simulación de deficiencias en el desarrollo). La razón por la que la conexión de AB no se acerca a 1 es porque el entrenamiento mezclado con la relación AC genera un fenómeno de interferencia; durante los ensayos AC el modelo no solo aprende dicha relación sino que también se debilita la relación entre A y B.

En el modelo, cuando AB se entrena sin la interferencia de AC, hasta alcanzar valores de conexión cercanos a 1 (i.e., 0.99 tras 45 ensayos), y hasta entonces se presentan los bloques de

ensayos con relaciones AB y AC (30 ensayos), entonces la conexión de las relaciones entrenadas se mantiene fuerte y la relación derivada BC comienza a fortalecerse.

Este resultado puede ser tomado como una predicción del modelo, se debe analizar en investigaciones futuras el efecto de evitar la interferencia de aprendizaje de relaciones-estímulo en niños con deficiencias intelectuales para ver si esto permite una mejora en su ejecución en las fases de prueba.

Simulación 4, Spencer y Chase (1996)

Distintos niveles de relación entre los miembros de una clase estímulo pueden depender de la cantidad de estímulos nodales; mientras el número de nodos incrementa, el nivel de relación disminuye. Una suposición es que conforme el nivel de relación entre estímulos incrementa, la velocidad para responder a esas relaciones también aumenta. Spencer y Chase (1996) estaban particularmente interesados en medir las velocidades de respuesta durante las respuestas a relaciones entre estímulos equivalentes. Su propósito fue caracterizar las respuestas a ensayos de línea base y de prueba a través de un análisis de tiempos de respuesta.

En el estudio original se formaron 3 grupos experimentales. El primer grupo fue instruido verbalmente acerca de cómo los estímulos estaban relacionados, el segundo grupo fue interrogado durante sus respuestas en las pruebas acerca de sus razones para responder de una u otra forma. La Simulación 4 se centró en el tercer grupo, denominado como Grupo Estándar, formado por 4 estudiantes universitarios que no fueron instruidos ni cuestionados en las fases de prueba.

Los estímulos fueron 21 figuras sin sentido acomodadas en tres clases de 7 miembros cada una. Las tareas se corrieron en una computadora donde los participantes aprendieron seis

conjuntos de relaciones; las AB, BC, CD, DE, EF, y FG para las clases 1, 2 y 3, con ensayos de IM que tenían un estímulo muestra y 3 estímulos de comparación. El entrenamiento se dividió en 7 etapas con 48 ensayos cada una. Cada conjunto de relaciones se entrenó durante cada una de las primeras 6 etapas; por ejemplo, el conjunto de relaciones AB, compuesto por A1B1, A2B2 y A3B3 fue entrenado durante la primer etapa, una vez que los participantes lograron el criterio pasaron a la segunda etapa de entrenamiento, donde 24 ensayos consistieron en entrenar las relaciones BC y otros 24 ensayos se usaron para mantener las relaciones aprendidas previamente.

Las respuestas correctas generaron retroalimentación verbal y la ganancia de puntos; cada punto fue equivalente a \$0.01 dólar. El criterio de aprendizaje para avanzar entre etapas fue tener al menos 90% de respuestas correctas y no más de un error en cada tipo de relación. Durante la etapa 7 todas las relaciones fueron presentadas en un bloque mixto y sin reforzamiento.

Se requirieron 5 repeticiones de la etapa 7 con una ejecución de al menos 90% de ensayos correctos para terminar el entrenamiento. El Apéndice C muestra las relaciones entrenadas y la secuencia utilizada en este estudio.

Durante las pruebas se evaluó un total de 108 relaciones; cuarenticinco correspondieron con relaciones transitivas (e.g., AC), otras 45 correspondieron con ensayos llamados “pruebas combinadas” dado que se combinaba la transitividad con la simetría (e.g., CA, prueba de equivalencia), y otros 18 ensayos correspondieron con pruebas de simetría (e.g., BA). Estos ensayos también se muestran en el Apéndice C. No se presentó reforzamiento durante estas pruebas.

El entrenamiento con el modelo consideró el número de ensayos, secuencia de entrenamiento por etapas y criterios del estudio original. Un total de 18 patrones de estimulación

fueron diseñados para el entrenamiento de los 6 conjuntos de relaciones. En la etapa 7 se presentaron todos los ensayos de entrenamiento mezclados de forma aleatoria sin reforzamiento

Durante las pruebas la red fue estimulada con un total de 126 patrones de entrada, correspondientes con el total de ensayos posibles; dieciocho de entrenamiento y 108 de prueba.

Resultados y Discusión

Spencer y Chase encontraron una tendencia a disminuir la velocidad de respuesta conforme la distancia nodal aumentaba, demostrando la presencia de efectos de nodalidad.

En la Figura 14, paneles A y B se presentan las gráficas del estudio original, donde la tendencia descrita puede ser observada, la Figura 14 Panel C muestra los valores de las conexiones en el modelo para las relaciones de línea base y emergentes, dependiendo del número de nodos que hay entre cada relación. Se puede observar que el nivel de relación aparece como una función inversa de la distancia nodal; mientras el número de nodos incrementa, el nivel de asociación entre estímulos disminuye. El Panel D muestra la media de los valores de conexiones lograda por el modelo para todas las relaciones de línea base y para todas las relaciones derivadas.

El principal resultado de la Simulación 4 es demostrar que el Modelo Neurocomputacional 2 pudo simular los efectos de nodalidad que son comúnmente reportados en estudios de EE (e.g., Bortoloti & de Rose, 2009; Fields, Adams, Verhave & Newman, 1993; Fields, Landon-Jimenez, Buffington & Adams, 1995; Kennedy, 1991; Kennedy, Itkonen & Lindquist; 1994, Moss-Lourenco & Fields, 2011). Cabe mencionar que los valores de las conexiones pueden deberse no sólo a la cantidad de nodos que conectan a cada par de estímulos, sino a la cantidad de veces que cada relación de entrenamiento fue presentada. Como se aprecia

en el Apéndice C, hay una frecuencia desigual en la presentación de ensayos de entrenamiento, en una simulación complementaria el modelo mostró que cuando se iguala la cantidad de ensayos de entrenamiento las diferencias en los valores de conexión de las relaciones transitivas son explicadas tanto por la distancia nodal como por la frecuencia de entrenamiento.

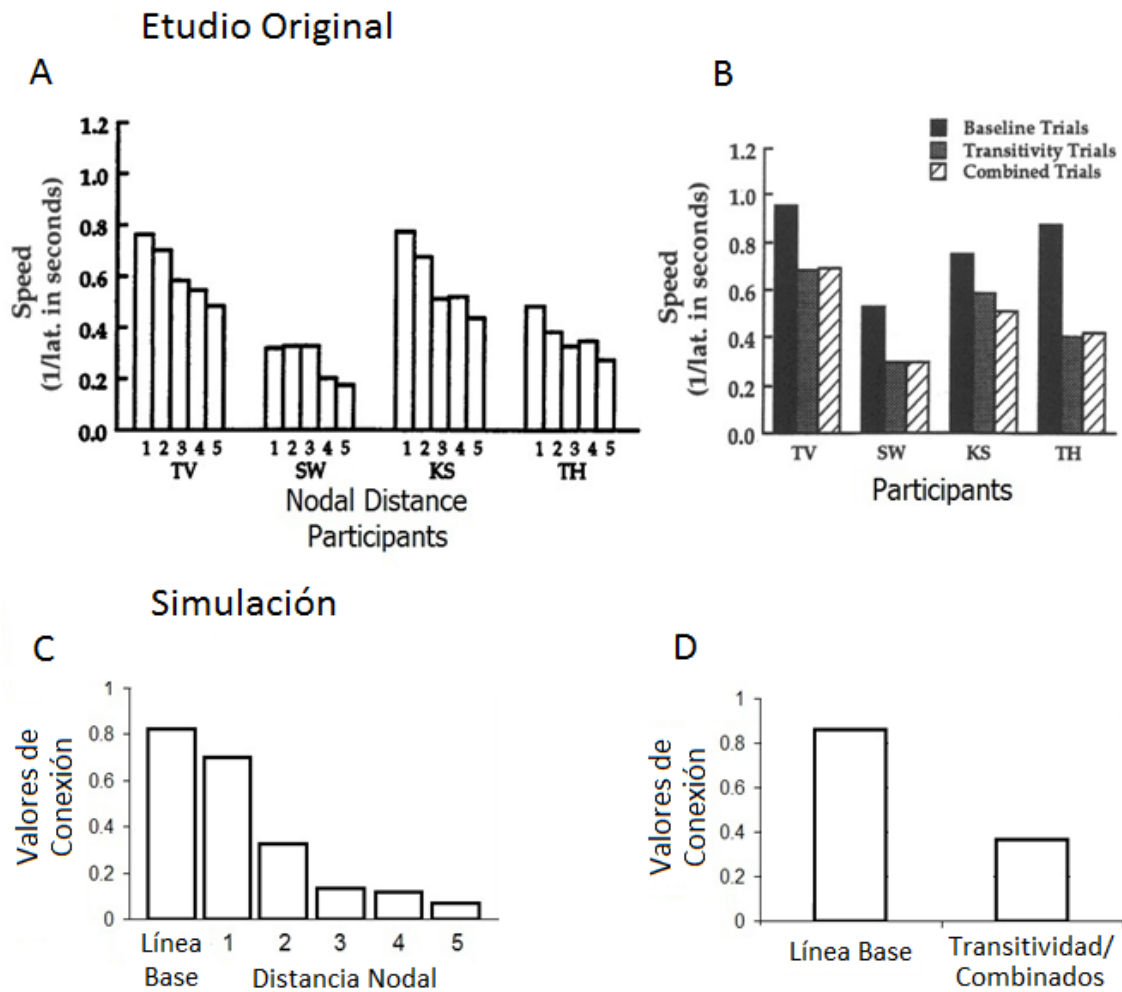


Figura 14. Los paneles A y B son del grupo estándar de participantes del estudio de Spencer y Chase (1996). En A se muestra la media de velocidad para respuestas correctas en relaciones transitivas de 1 hasta 5 nodos. En B se muestra la media de velocidad de respuesta para pruebas de línea base, transitividad y combinadas. En C se muestra el valor final de las conexiones en el modelo para ensayos de línea base y transitividad de uno hasta cinco nodos. En D se muestra la media final de los valores de conexión en todas las relaciones de línea base y todas las relaciones emergentes.

Como principal limitación del Modelo Neurocomputacional 2 se puede mencionar que la forma en la que éste aprende y asocia estímulos está principalmente guiada por un proceso bottom-up, el grado de relación entre estímulos depende de la información respecto a ocurrencia y co-ocurrencia en el medio ambiente, sin embargo, los procesos top-down pueden tener una fuerte influencia en la formación de clases (Wisniewski & Medin, 1994) como se mostró en el Modelo Neurocomputacional 1, y en el Experimento 2. Por ejemplo, si un participante experimenta a los estímulos A y B junto a una etiqueta que dice “No relacionados” es probable que no derive relaciones transitivas entre A y cualquier otro estímulo relacionado a B.

CAPÍTULO 4. DISCUSIÓN GENERAL

El paradigma de EE permite estudiar la adquisición y despliegue de conductas simbólicas que se caracterizan por el establecimiento de relaciones arbitrarias entre estímulos y la transferencia de respuestas, asociadas a los estímulos, vía dichas relaciones. Un considerable cuerpo de investigación se ha enfocado a determinar las circunstancias bajo las cuales se establecen las relaciones de equivalencia entre estímulos. En esta tesis se analizó y exploró el establecimiento de dichas relaciones con un acercamiento neurocomputacional.

Los modelos neurocomputacionales capturan la forma básica del procesamiento de información de los sistemas biológicos para analizar procesos cognitivos y conductuales en simulaciones computacionales totalmente manipulables. Este enfoque permite explorar las ideas que un investigador tiene sobre un fenómeno, plantear hipótesis y generar descripciones y predicciones del fenómeno.

En este trabajo se presentaron dos modelos neurocomputacionales que simularon tareas de EE. El Modelo Neurocomputacional 1 se desarrolló tras notar que no existían simulaciones inequívocas del proceso de aprendizaje y derivación de relaciones-estímulo en modelos multicapa con aprendizaje por retropropagación; se simuló un procedimiento de estímulos compuestos (Debert et al., 2007) y respuestas tipo *yes/no* (Fields et al., 2009).

Con este modelo fue posible demostrar el aprendizaje y la derivación de relaciones-estímulo, los datos generados resultaron similares a las ejecuciones observadas de los participantes humanos de los experimentos 1 y 2.

Para el Experimento 1 y la Simulación 1, seis participantes humanos y 6 corridas del modelo trabajaron con un procedimiento de estímulos compuestos. En el caso de los humanos se

utilizaron las palabras “sí” y “no” como opciones de respuesta. En la simulación se utilizaron dos opciones de respuesta que fueron utilizadas en otra tarea de categorización previa a la simulación experimental. Las ejecuciones del modelo fueron similares a las de los humanos, los datos obtenidos apuntaron a considerar que la experiencia previa y las instrucciones asociadas a las opciones de respuesta (palabras “sí” y “no”) tienen una influencia sobre la formación consistente de las clases estímulo.

Los resultados del Experimento 1 y la Simulación 1 fueron comparados con un segundo experimento y una segunda simulación, donde se utilizaron los mismos estímulos compuestos para formar dos clases estímulo, pero con un sistema de respuesta que implementó opciones sin etiquetas verbales (o sin experiencia previa asociada), consideradas como opciones de respuesta arbitrarias. Seis humanos participaron en este segundo experimento, y se hicieron 6 corridas más como parte de la Simulación 2.

Ninguno de los participantes humanos pudo formar las 2 clases estímulo de 3 miembros. La Simulación 2 mostró que el modelo no pudo formar las clases estímulo de forma consistente a lo largo de las 6 corridas, solamente en 2 corridas el modelo formó clases de forma similar a la Simulación 1. Los resultados de los experimentos 1 y 2 resultaron consistentes con los datos generados por las simulaciones 1 y 2 del primer modelo, esto permite considerar que la arquitectura y aprendizaje de este modelo son adecuados para simular tareas de EE con estímulos compuestos.

El contraste entre mayor o menor experiencia asociada a las opciones de respuesta en las simulaciones 1 y 2 mostró que cuando se utilizan opciones de respuesta novedosas, que carecen de una historia explícita en tareas de categorización, se pierde consistencia en la formación de clases. Estos datos cuestionan la pureza metodológica de procedimientos similares que son

utilizados con humanos (e.g., Fields et al., 2009), pues en ellos no se considera la historia de aprendizaje asociada a las opciones de respuesta, ni su influencia en la resolución de la tarea.

Los datos mencionados también cuestionan si el fracaso en las ejecuciones de animales no humanos en tareas de EE similares se debe a la metodología de entrenamiento utilizada. Las fallas en experimentos de EE con tareas de estímulos compuestos en animales no humanos (e.g., Campos et al., 2011), se dan como resultado de procedimientos donde se entrenan relaciones entre estímulos sin que se establezca un entrenamiento diferencial entre las opciones de respuesta. Quizá este entrenamiento sea necesario para que dichos experimentos puedan funcionar como un buen modelo de lo que ocurre con la formación de clases con humanos.

El Modelo Neurocomputacional 1 también permitió observar que las distintas combinaciones de estímulos que forman parte de una clase de equivalencia generan el mismo tipo de representación intermedia. Las representaciones comunes para estímulos que componen a la misma clase son consistentes con la hipótesis de que los estímulos de las clases de equivalencia adquieren el mismo significado resultando en un buen modelo de relaciones semánticas (e.g., Barnes-Holmes, Staunton et al., 2005).

En el Modelo Neurocomputacional 2 se intentaron superar algunas limitaciones del Modelo Neurocomputacional 1. Se desarrolló un modelo Hebbiano con un algoritmo de aprendizaje biológicamente plausible en donde se simulaban ensayos de IM. Con este modelo se observó cómo se desarrollan las relaciones-estímulo a lo largo del entrenamiento y como uno de los datos destacados de este modelo y esta tesis, se observó el desarrollo de las relaciones emergentes de transitividad.

Este segundo modelo estuvo diseñado con una arquitectura simple que consistió de una capa de neuronas donde se representaron de forma localista los estímulos y donde las conexiones

entre neuronas se interpretaron como el nivel de relación entre estímulos. El aprendizaje Hebbiano implementado, con el que se actualizaban los valores de las conexiones, fue congruente con la revisión de experimentos neurofisiológicos sobre LTP y LTD. Las conexiones generaron una descripción cuantitativa de las regularidades ambientales de estimulación presentadas en las simulaciones. Entre las variables que se observaron con un efecto sobre los valores de conexión están: frecuencia de ensayos de una relación, secuencia de ensayos, y estructura de entrenamiento (e.g., estructura lineal y de racimos).

El Modelo Neurocomputacional 2 se acerca a la forma en la que los humanos son sensibles a las regularidades de estimulación; los valores de conexión en el modelo resultaron comparables con la exactitud y tiempo de respuesta de los 3 experimentos simulados de participantes humanos. Esta congruencia resulta relevante pues sitúa a este modelo como una herramienta para generar predicciones; la simulación de un tipo de entrenamiento particular llevará a generar valores de conexión que puedan tomarse como predictivos del nivel de relación entre estímulos que debería establecerse si el mismo entrenamiento se utiliza con humanos.

El proceso de aprendizaje de relaciones entrenadas es directamente observable tanto en las ejecuciones de humanos como en las simulaciones del Modelo Neurocomputacional 2. Sin embargo, hay partes de las tareas de equivalencia que son difícilmente observables durante el transcurso de los experimentos con humanos, como es el establecimiento de las relaciones derivadas, que en la mayoría de los casos sólo pueden ser observadas hasta el final del experimento, sin embargo en este modelo resultaron directamente observables a lo largo de toda la simulación. Esto brinda la ventaja de poder generar inferencias que en experimentos con humanos resultan complicadas, por ejemplo, respecto al momento y rapidez con la que se establece una relación de transitividad.

La Simulación 3 del Modelo Neurocomputacional 2 mostró fallas en la derivación de relaciones transitivas al simular a participantes con desarrollo atípico. Esta simulación se generó al considerar modificaciones paramétricas en el algoritmo de aprendizaje de acuerdo a la evidencia de las deficiencias neurofisiológicas encontradas en modelos animales de retraso mental. (Kleschevnikov et al., 2004; Meredith, et al., 2007; Scott-McKean & Costa, 2011). De tal forma que este modelo pudo simular poblaciones típicas y atípicas manteniendo congruencia biológica en todo momento. El análisis de la Simulación 3 mostró que la incapacidad para derivar relaciones transitivas resulta de un proceso de interferencia de aprendizaje que lleva al establecimiento de relaciones entrenadas con un nivel de relación bajo. Cuando se mezclan varias relaciones-estímulo en el entrenamiento el aprendizaje se vuelve deficiente. Si las relaciones se entrenan de forma separada, evitando que el aprendizaje de una de ellas interfiera con el aprendizaje de otra, entonces el nivel de relación entre estímulos alcanza valores más elevados y permite entonces derivar correctamente relaciones de transitividad.

Las simulaciones de poblaciones atípicas resultan relevantes pues dan generalidad y validez al modelo, a la vez que lo convierten en un buen candidato para simular distintos procedimientos aplicados a poblaciones con desarrollo atípico para generar estrategias de entrenamiento y modificación conductual específicas.

Del segundo modelo se pueden extraer hipótesis respecto a la naturaleza de las relaciones transitivas, un tema que ha sido de gran relevancia en la literatura de EE. En este modelo las relaciones de transitividad se observaron cuando las conexiones directamente reforzadas fueron lo suficientemente fuertes como para permitir la propagación de la activación hasta alcanzar neuronas con las que no se había reforzado una conexión directamente, es decir, emergieron

redes funcionales como resultado de la estimulación directa y de un proceso de propagación de la activación.

La propagación de la activación dotó a este modelo con la capacidad de “representar” relaciones entre estímulos o eventos sin necesidad de percibirlos, este mecanismo fue utilizado para modificar la conectividad entre neuronas que representaban estímulos que nunca fueron presentados juntos. De esta forma la relación derivada AC, tras el entrenamiento de AB y BC es explicada al considerar que cada vez que B es presentado, su relación con A y C, produce la activación de los tres estímulos y entonces el incremento en la relación entre A y C toma lugar, con el mismo mecanismo que explica los incrementos en las relaciones de entrenamiento.

Cuando se considera la ocurrencia de este proceso a lo largo de varias relaciones-estímulo de varias categorías, se puede pensar en una dinámica de propagación y activación neuronal que da como resultado el surgimiento de representaciones que van reclutando o inhibiendo a otras representaciones para conformar lo que entendemos como un procesamiento ordenado y congruente. Las características de este modelo permiten explorar varios procesos psicológicos como el almacenamiento y recuperación de información, la memoria semántica y la formación de conceptos.

A la luz del trabajo teórico existente en EE, el Modelo Neurocomputacional 2 tiene más coincidencias con la propuesta de reforzamiento de Sidman (2000), pues no fue necesario ningún otro proceso o aprendizaje previo para que se derivaran relaciones de transitividad como han argumentado las otras propuestas teóricas (Dugdale & Lowe, 1990; Hayes et al., 2002), aunque tampoco se descarta que el desarrollo de otras habilidades, como las de lenguaje, o una mayor experiencia en formación de categorías tengan una influencia sobre la formación de clases estímulo.

La idea del fortalecimiento de conexiones o asociaciones entre representaciones co-activas tiene una larga tradición en la psicología, se pueden ubicar propuestas en este sentido desde William James (citado en Spitzer, 1998), resulta interesante que la evidencia neurofisiológica apoya la viabilidad de este proceso, por lo cual, el tomar la tarea de EE y llevarla a simulaciones que contemplan este mecanismo de aprendizaje, permite ubicarla como un caso más que se apega a los principios generales de aprendizaje, pese a que el establecimiento de muchas de las relaciones-estímulo no devenga de experiencias directamente observables.

La formación de clases de estímulos equivalentes fue analizada desde una perspectiva neurocomputacional. Se demostró que es posible simular experimentos de EE y que los modelos pueden acompañar las investigaciones del área para explorar el proceso de aprendizaje de las clases. Esto extiende el uso de modelos neurocomputacionales a otra área de interés para los psicólogos experimentales, y da generalidad a las tareas de equivalencia; los modelos utilizados no estuvieron diseñados para simular específicamente EE, sino tareas en el contexto general del aprendizaje asociativo.

Los modelos presentados mostraron utilidad en el contexto de estudios básicos y aplicados. En el contexto teórico permitieron esbozar explicaciones mecanicistas del aprendizaje de las clases estímulo, basadas en la formalización de supuestos sobre el procesamiento de información en redes neuronales. Por otra parte, la relevancia para los estudios aplicados se dio al mostrar la viabilidad de los modelos como herramientas predictivas del éxito o fracaso en la formación de clases; debe extenderse la investigación y adecuación de estos modelos para que las simulaciones permitan determinar las condiciones que resultarán más exitosas para establecer clases de estímulos equivalentes.

REFERENCIAS

- Acuna, B. D., Eliassen, J. C., Donoghue, J. P., & Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cerebral Cortex, 12*, 1312-1321.
- Arntzen, E. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *The Psychological Record, 60*, 437-462.
- Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149-178.
- Barnes, D. & Hampson, P. J. (1993). Stimulus equivalence and connectionism: Implications for behavior analysis and cognitive science. *The Psychological Record, 43*, 617-638.
- Barnes-Holmes, D., Rodríguez, M., Whelan, R. (2005). La teoría de los marcos relacionales y el análisis experimental del lenguaje y la cognición, *Revista Latinoamericana de Psicología, 37- 2*, 255-275.
- Barnes-Holmes, D., Staunton, C., Whelan, R., Barnes-Holmes, Y., Commins, S., Walsh, D., et al. (2005). Derived stimulus relations, semantic priming, and event-related potentials: testing a behavioral theory of semantic networks. *Journal of the Experimental Analysis of Behavior, 84*, 417-433.
- Bear, M. F. (2003). Bidirectional synaptic plasticity: from theory to reality. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358* (1432), 649-655.
- Bechtel, W., Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks*. Oxford. Basil Blackwell.
- Bentall, R. P., Dickins, D. W. & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology, 46B*, 187-214.
- Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience, 2*, 32-48.

- Bliss, T., Collingridge, G., & Morris, R. (2007). Synaptic plasticity in the hippocampus. In: Andersen P., Morris R., Amaral D., Bliss T., O'Keefe J., (Eds.), *The Hippocampus Book*. New York: Oxford University Press. pp. 343–474.
- Bortoloti, R. & de Rose, J. (2009). Assessment of the relatedness of equivalent stimuli through a semantic differential, *The Psychological Record*, 59, 563–590.
- Burgos, J. (2007). Autoshaping and automaintenance: a neural-network approach. *Journal of the Experimental Analysis of Behavior*, 88, 115-130.
- Campos, H. C., Debert, P., da Silva Barros, R., McIlvane, W. J. (2011). Relational discrimination by pigeons in a go/no-go procedure with compound stimuli: A methodological note. *Journal of the Experimental Analysis of Behavior*, 96(3): 417-26.
- Carr, D., Wilkinson, K. M., Blackman, D. & McIlvane, W. J. (2000). Equivalence classes in individuals with minimal verbal repertoires. *Journal of the Experimental Analysis of Behavior*, 74, 101-114.
- Clayton, M. C., & Hayes, L. J. (2004). A comparison of match-to-sample and respondent-type training of equivalence classes. *The Psychological Record*, 54, 579-602.
- Cohen, H., & Lefebvre, C. (2005). *Handbook of Categorization in Cognitive Science*. Maryland Heights, MO: Elsevier Science.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review* 82(6), 407-428.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied Behavior Analysis*. New York: Macmillan.
- Cullinan, V., Barnes, D., Hampson, P.J. & Lyddy, F. (1994). A transfer of explicitly and non-explicitly trained sequence responses through equivalence relations: An experimental demonstration and connectionist model. *The Psychological Record*, 44, 559-586.
- De Rose, J. & De Souza, D., & Hanna, E.S. (1996). Teaching reading and spelling: Exclusion and stimulus equivalence. *Journal of Applied Behavior Analysis*, 29, 451-469.
- Debert, P., Huziwara, E., Faggiani, R., Simoes de Mathis, M., & McIlvane, W. (2009). Emergent conditional relations in a go/no-go procedure: Figure-ground and stimulus-position compound relations. *Journal of the Experimental Analysis of Behavior*, 92, 233-243.

- Debert, P., Matos, M. A., & McIlvane, W. J. (2007). Conditional relations with compound abstract stimuli using a go/no-go procedure. *Journal of the Experimental Analysis of Behavior*, *87*, 89-96.
- DeGrandpre, R. J., Bickel, W. K., & Higgins, S. T. (1992). Emergent equivalence relations between interoceptive (drug) and exteroceptive (visual) stimuli. *Journal of the Experimental Analysis of Behavior*, *58*, 9-18.
- Devany, J. M., Hayes, S. C. & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disable children. *Journal of the Experimental Analysis of Behavior*, *46*, 243-257.
- Dickins, D. W. (2005). On aims and methods in the neuroimaging of derived relations. *Journal of the Experimental Analysis of Behavior*, *84*, 453-483.
- Dickins, D. W., Singh, K. D., Roberts, N., Burns, P., Downes, J. J., Jimmieson, P., & Bentall, R. P. (2001). An fMRI study of stimulus equivalence. *Neuroreport*, *12*, 405-411.
- Dickins, T. E., & Dickins, D. W. (2001). Symbols, stimulus equivalence and the origins of language. *Behavior and Philosophy*, *29*, 221-244.
- Dube, W., Green, G., & Serna, R. W. (1993). Auditory successive conditional discrimination and auditory stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, *59*, 103-114.
- Dugdale, N. & Lowe, C. F. (1990). Naming and stimulus equivalence. En D. E. Blackman & H. Lejeune. (Eds.) *Behavior Analysis in Theory and Practice: Contributions and controversies*. (pp. 115-138). Hove, England: Erlbaum.
- Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin*, *19*, 8-12.
- Eikeseth, S. & Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior*, *58*, 123-133.
- Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior*, *48*, 317-332.
- Fields, L., Adams, B., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior*, *53*, 345-358.

- Fields, L., Adams, B., Verhave, T., & Newman, S. (1993). Are stimuli in equivalence classes equally related to each other? *The Psychological Record*, *45*, 85-105.
- Fields, L., Doran, E., & Marroquin, M. (2009). Equivalence class formation in a trace stimulus pairing two-response format: Effects of response labels and prior programmed transitivity induction. *Journal of the Experimental Analysis of Behavior*, *92*, 57–84.
- Fields, L., Landon-Jimenez, D., Buffington, D. & Adams, B. (1995). Maintained nodal-distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior*, *64*, 129-145.
- Fields, L., Reeve, K., Varelas, A., Rosen, D. Belanich, J. (1997) Equivalence class formation using stimulus-pairing and yes-no responding. *The Psychological Record*, *47(4)*, 661-686.
- Fields, L., Travis, R., Roy, D., Yadlovker, E., Aguiar-Rocha, L., & Sturmey, P. (2009). Equivalence class formation: a method for teaching statistical interactions. *Journal of Applied Behavior Analysis*, *42*, 575-593.
- Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching brain-behavior relations economically with stimulus equivalence technology. *Journal of Applied Behavior Analysis*, *43*, 19-33.
- Fienup, D. M., Critchfield, T. S., & Covey, D. P. (2009). Building contextually-controlled equivalence classes to teach about inferential statistics: a preliminary demonstration. *Experimental Analysis of Human Behavior Bulletin*, *27*, 1-10.
- Fiorentini, L., Arismendi, M., & Yorrio, A. A. (2012). Una revision de las aplicaciones del paradigma de equivalencia de estímulos. *International Journal of Psychology and Psychological Therapy*, *12*, 261-275.
- Froc, D. J., & Racine, R. J., (2005) Interactions between LTP- and LTD-inducing stimulation in the sensorimotor cortex of the awake freely moving rat. *Journal of Neurophysiology*, *93*, 548–556.
- Goldstone, R. L., & Kersten, A.W. (2003). Concepts and categorization. In A. F. Healy & R. W. Proctor (Eds.). *Comprehensive Handbook of Psychology, Volume 4: Experimental Psychology*. (pp. 599-621). New Jersey: Wiley.
- Grossman, M., Smith, E. E., Koeing, P., Glosser, G., DeVita, C., Moore, et al. (2002). The neural basis of categorization in semantic memory. *NeuroImage*, *17*, 1549-1561.

- Haimson, B., Wilkinson, K. M., Rosenquist, C., Ouimet, C., & McIlvane, W. J. (2009). Electrophysiological correlates of stimulus equivalence processes. *Journal of the Experimental Analysis of Behavior, 92*, 245-256.
- Hall, S.S., Debernardis, G.M. & Reiss, A.L. (2006). The acquisition of stimulus equivalence in adolescents with fragile X syndrome. *Journal of Intellectual Disability Research, 50*: 643-651.
- Hayes, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on Verbal Behavior* (pp. 19-40). Reno, NV: Context Press.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.) (2002). Relational frame theory: A post-Skinnerian analysis of human language and cognition. New York. Plenum.
- Hebb, D. O. (1949) *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.
- Heckers, S., Zalesak, M., Weiss, A. P., Ditman, T., Titone, D. (2004). Hippocampal activation during transitive inference in humans. *Hippocampus, 14*, 153-162.
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*, 185-241.
- Houghton, G. (Ed.). (2005). *Connectionist Models in Cognitive Psychology*. London. Psychology Press.
- Huber, K. M., Gallagher, S. M., Warren, S. T., & Bear, M. F. (2002). Altered synaptic plasticity in a mouse model of fragile X mental retardation. *Proceedings of the National Academy of Sciences, USA 99*, 7746–7750.
- Imam, A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*, 107-124.
- Kennedy, C. H. (1991). Equivalence class formation influenced by the number of nodes separating stimuli. *Behavioural Processes, 24*, 219–245.
- Kennedy, C. L., Itkonen, T., & Lindquist, K (1994). Nodality effects during equivalence class formation: An extension to sight-word reading and concept development. *Journal of Applied Behavior Analysis, 27*, 673-683.

- Kerr, N., Meyerson, L., & Flora, J.A. (1977). The measurement of motor, visual, and auditory discrimination skills. *Rehabilitation Psychology, 24*, 95-112.
- Kleschevnikov, A. M., Belichenko, P. V., Villar, A. J., Epstein, C. J., Malenka, R. C., & Mobley, W. C. (2004). Hippocampal long-term potentiation suppressed by increased inhibition in the Ts65Dn mouse, a genetic model of Down syndrome. *The Journal of Neuroscience, 24*, 8153–8160.
- Kohonen, T. (2001). *Self-Organizing Maps*. Berlin, Springer, 3rd ed.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.
- Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.). *The Cambridge Handbook of Computational Psychology*. (pp. 267-301). New York: Cambridge University Press.
- Leader, G., Barnes, D., & Smeets, P. M. (1996). Establishing equivalence relations using a respondent-type procedure. *The Psychological Record, 46*, 685-706.
- LeBlanc, L. A., Miguel, C. F., Cummings, A. R., Goldsmith, T. R., & Carr, J. E. (2003). The effects of three stimulus-equivalence testing conditions on emergent US geography relations of children diagnosed with autism. *Behavioral Interventions, 18*: 279–289.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309-332.
- Luciano, C., Gómez-Becerra, I., & Rodríguez-Valverde, M. (2007). The role of multiple-exemplar training and naming in establishing derived equivalence in an infant. *Journal of the Experimental Analysis of Behavior, 87*, 349-365.
- Lyddy, F. & Barnes-Holmes, D. (2007). Stimulus equivalence as a function of training protocol in a connectionist network. *The Journal of Speech Language Pathology and Applied Behavior Analysis, compiled issue 1.4 & 2.1*, 14-24.
- Lyddy, F., Barnes-Holmes, D. & Hampson, P.J. (2001). A transfer of sequence function via equivalence in a connectionist network. *The Psychological Record, 51*, 409-428.
- Lynch, D. & Cuvo, A. (1995). Stimulus equivalence instruction of fraction-decimal relations. *Journal of Applied Behavior Analysis, 28*, 115-126.
- Mackay H. A., Wilkinson K. M., Farrell C., & Serna R. W. (2011). Evaluating merger

- and intersection of equivalence classes with one member in common. *Journal of the Experimental Analysis of Behavior*, 96, 87–105.
- Malenka, R. C., & Bear, M. F. (2004) LTP and LTD: An embarrassment of riches. *Neuron*, 44, 5-21.
- Markham, M. R., & Dougher, M. J. (1993). Compound stimuli in emergent stimulus relations: Extending the scope of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, 60, 529-542.
- Matos, M. A. & Hübner-D'Oliveira, M. M. (1992) Equivalence relations and reading. En S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 83-94). Reno: Context Press.
- Mayor, J., & Plunkett, K. (2010). A neuro-computational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, 117, 1-31.
- McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, 1(1), 11-38.
- McCulloch, W. S., & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- McLeod, P., Plunkett, K. & Rolls, E. T. (1998). *Introduction to Connectionist Modelling of Cognitive Processes*. New York: Oxford University Press.
- Meredith, R. M., Holmgren, C. D., Weidum, M., Burnashev, N., & Mansvelder, H. D. (2007). Increased threshold for spike-timing-dependent plasticity is caused by unreliable calcium signaling in mice lacking fragile X gene Fmr1. *Neuron*. 54(4): 627-38.
- Miguel, C. F., Petursdottir, A. I., Carr, J. E., & Michael, J. (2008). The role of naming in stimulus categorization by preschool children. *Journal of the Experimental Analysis of Behavior*, 89, 383-405.
- Moss-Lourenco, P., & Fields, L. (2011). Nodal structure and stimulus relatedness in equivalence classes: post-class formation preference tests. *Journal of the Experimental Analysis of Behavior*, 95 (3), 343-368.
- Murre, J. M., (1996). TraceLink: A model of amnesia and consolidation of memory. *Hippocampus*, 6, 675-684.
- O'Donnell, J., & Saunders, K. J. (2003). Equivalence relations in individuals with

- language limitations and mental retardation. *Journal of the Experimental Analysis of Behavior*, 80, 131–157.
- O'Reilly, R. C. & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience. Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.
- O'Reilly, R.C. (2005). The division of labor between the neocortex and hippocampus. En G. Houghton (Ed). *Connectionist Models in Cognitive Psychology*, London. Psychology Press.
- O'Leary, C. A., & Bush, K. M. (1996). Stimulus equivalence in the tactile modality. *The Psychological Record*, 46, 509-518.
- Plunkett, K., & Elman, J. L. (1997). *Exercises in Rethinking Innateness: A handbook for connectionist simulations*. Cambridge MA: MIT Press.
- Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34, 909-957.
- Rehfeldt, R., & Barnes-Holmes, Y. (Ed's.) (2009). *Derived Relational Responding: Applications for learners with autism and other developmental disabilities*. New Harbinger/Context Press.
- Rehfeldt, R., A. (2011). Toward a technology of derived stimulus relations: An Analysis of articles published in the Journal of Applied Behavior Analysis, 1992–2009. *Journal of Applied Behavior Analysis*, 44, 109-119.
- Rescorla, R. A., & Wagner, A. R. (1972) *A Theory of Pavlovian Conditioning: Variations in the effectiveness of reinforcement and nonreinforcement*, Classical Conditioning II, A.H. Black & W.F. Prokasy, Eds., pp. 64–99. Appleton-Century-Crofts.
- Rolls, E. T. (2008). *Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach*. Oxford University Press: Oxford.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Erlbaum.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. En D. E. Rumelhart, J. L. McClelland, & the PDP Research Group

- (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. I, pp. 318–362). Cambridge, MA: MIT Press.
- Rumelhart, D. E., McClelland, J. L., & the PDP research group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Saunders, K. J., O'Donnell, J., Vaidya, M., & Williams, D. C., (2003). Recombinative generalization of within-syllable units in nonreading adults with mental retardation. *Journal of Applied Behavior Analysis, 36*, 95-99.
- Schlund, M. W., Cataldo, M. F., Hoehn-Saric, R. (2008). Neural correlates of derived relational responding on tests of stimulus equivalence. *Behavioral and Brain Functions, 4*, 6.
- Schlund, M. W., Hoehn-Saric, R., Cataldo, M. F. (2007). New knowledge derived from learned knowledge: Functional-anatomic correlates of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 87*, 287-307.
- Schusterman R. J., & Kastak, D. (1998). Functional equivalence in a California sea lion: Relevance to animal social and communicative interactions. *Animal Behavior, 55*, 1087–1095.
- Scott-McKean, J. J., & Costa, A. C. S. (2011). Exaggerated NMDA mediated LTD in a mouse model of Down syndrome and pharmacological rescuing by memantine. *Learning and Memory, 18*, 774-778.
- Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 4*, 5-13.
- Sidman, M. (1990). Equivalence relations: Where do they come from?. En D. E. Blackman & H. Lejeune (Eds.). *Behavior Analysis in Theory and Practice: Contributions and controversies*. (pp. 93-114). Hillsdale, NJ: Erlbaum.
- Sidman, M. (1992) Equivalence relations: Some basic considerations. En S. C. Hayes & L. J. Hayes (Eds.) *Understanding Verbal Relations. The second and third international institute on verbal relations*. (pp. 15-27). Reno, NV: Context Press.
- Sidman, M. (1994). *Equivalence relations and behavior: A Research Story*. Boston: Authors Cooperative.
- Sidman, M. (1997). Equivalence: A theoretical or a descriptive model?. *Revista Mexicana de Análisis de la Conducta, 23*, 125-145.

- Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, 74, 127-146.
- Sidman, M., & Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency*, 77, 515-523.
- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37, 5-22.
- Sidman, M., Cresson, O., & Willson-Morris, M. (1974). Acquisition of matching to sample via mediated transfer. *Journal of the Experimental Analysis of Behavior*, 22, 261-273.
- Spencer, T. J., & Chase, P. N. (1996). Speed analysis of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, 65, 643-659.
- Spitzer, M., (1998). The history of neural network research in psychopathology. En D. J. Stein, & J. Ludik. (eds) (1998). *Neural networks and psychopathology: Connectionist models in practice and research*. Cambridge University Press. (pp. 14-33).
- Stromer, R., & Mackay, H. A. (1992). Spelling and emergent picture-printed word relations established with delayed identity matching to complex samples. *Journal of Applied Behavior Analysis*, 25, 893-904.
- Stromer, R., Mackay, H., McVay, A. & Fowler, T. (1998). Written lists as mediating stimuli in the matching-to-sample performances of individuals with mental retardation. *Journal of Applied Behavior Analysis*, 31, 1-19.
- Sun, R. (Ed). (2008). *The Cambridge handbook of computational psychology*. Cambridge University Press.
- Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88, 135-171.
- Thomas, M. S. C. & McClelland, J. L. (2008). Connectionist models of cognition. En R. Sun (Ed). *Cambridge handbook of computational psychology*. Cambridge University Press. 23-58.
- Turrigiano, G. G., & Nelson, S. B. (2000). Hebb and homeostasis in neuronal plasticity. *Current Opinion in Neurobiology*, 10(3), 358-364.

- Wilkinson, K. M., & McIlvane, W. J. (2001). Methods for studying symbolic behavior and category formation: Contributions of stimulus equivalence research. *Developmental Review, 21*, 355-374.
- Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18*, 221-281.
- Wulfert, E., & Hayes, S. C. (1988). Transfer of conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50*, 125-144.
- Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: an introduction. *Journal of the Experimental Analysis of Behavior, 78*, 237-248.

APÉNDICE A

Secuencia de entrenamiento y prueba con el número de ensayos y criterios para el estudio original de Sidman y Tailby (1982) y para la Simulación 2 del Modelo Neurocomputacional 2.

Sidman & Tailby (1982)	Simulación 2	Criterios
Fase A. Pre-entrenamiento:		
1. Colores de muestra y colores de comparación	-----	
2. Nombre de los colores de muestra y colores de comparación		
Fase B. Entrenamiento	Entrenamiento	Correctos/Total de Ensayos
1. Entrenamiento de relaciones AB	1. Entrenamiento de relaciones AB	
A1B1, A2B2	A1B1, A2B2	19/20
A1B1, A3B3	A1B1, A3B3	19/20
A2B2, A3B3	A2B2, A3B3	19/20
A1B1, A2B2, A3B3	A1B1, A2B2, A3B3	29/30
2. Entrenamiento de relaciones AC	2. Entrenamiento de relaciones AC	
A1C1, A2C2	A1C1, A2C2	19/20
A1C1, A3C3	A1C1, A3C3	19/20
A2C2, A3C3	A2C2, A3C3	19/20
A1C1, A2C2, A3C3	A1C1, A2C2, A3C3	29/30
3. Entrenamiento de AB y AC	3. Entrenamiento de AB y AC	
A1B1, A2B2, A3B3, A1C1, A2C2, A3C3	A1B1, A2B2, A3B3, A1C1, A2C2, A3C3	29/30
4. Entrenamiento de relaciones DC	4. Entrenamiento de relaciones DC	
D1C1, D2C2	D1C1, D2C2	19/20
D1C1, D3C3	D1C1, D3C3	19/20
D2C2, D3C3	D2C2, D3C3	19/20
D1C1, D2C2, D3C3	D1C1, D2C2, D3C3	29/30
5. Entrenamiento de AB, AC, y DC	5. Entrenamiento de AB, AC, y DC	
A1B1, A2B2, A3B3, A1C1, A2C2, A3C3, D1C1, D2C2, D3C3	A1B1, A2B2, A3B3, A1C1, A2C2, A3C3, D1C1, D2C2, D3C3	44/45
6. Disminución gradual de la probabilidad de reforzamiento	-----	
Fase C. Pruebas *		
1. Pruebas de DB con ensayos de línea base AB, AC, y DC		Total de ensayos en el estudio de Sidman & Tailby
2. Pruebas de BD con ensayos de línea base AB, AC, y DC		120
3. Pruebas de AD con ensayos de línea base AC y DC		120
4. Pruebas de BC con ensayos de línea base AB y AC	Un bloque de prueba que contuvo todos los ensayos de línea base y las pruebas de las relaciones DB, BD, AD, BC, CB, y CD	90
5. Pruebas de CB con ensayos de línea base AB y AC		60
6. Pruebas de CD con ensayos de línea base DC		60
7. Pruebas de nombramiento	-----	60

Nota *Las etapas 1 a 7 de la Fase C fueron presentadas en diferente secuencia, sólo los participantes A.D. y D.W. recibieron las pruebas en la secuencia de 1 a 7

APÉNDICE B

Secuencia de entrenamiento y prueba con el número de ensayos y criterios para el estudio original de Devany, Hayes y Nelson (1986) y para la Simulación 3 del Modelo Neurocomputacional 2.

Devany Hayes & Nelson (1986)	Simulación 3	Criterios
Entrenamiento	Entrenamiento	Correctos/Total de Ensayos
A1B1	A1B1	9 correctos consecutivos/10
A2B2	A2B2	9 correctos consecutivos/10
A1B1 y A2B2	A1B1 y A2B2	9 correctos consecutivos/10
A1C1	A1C1	9 correctos consecutivos/10
A2C2	A2C2	9 correctos consecutivos/10
A1C1 y A2C2	A1C1 y A2C2	9 correctos consecutivos/10
A1B1, A2B2, A1C1, y A2C2	A1B1, A2B2, A1C1, y A2C2	9 correctos consecutivos/10 en el estudio original. 7/8 en el modelo con cada ensayo apareciendo dos veces
A1B1, A2B2, A1C1, y A2C2 con disminución gradual de la probabilidad de reforzamiento hasta llegar a ≈ 0.25	A1B1, A2B2, A1C1, y A2C2 presentados sin reforzamiento	No se especifica para el estudio original. En el modelo cada ensayo se presentó dos veces
Pruebas	Pruebas	
Un bloque de ensayos para evaluar B1C1, C1B1, B2C2, y C2B2. Cada ensayo se presentó 10 veces.	Un bloque de ensayos para evaluar B1C1, C1B1, B2C2, y C2B2. Cada ensayo se presentó una vez.	-----

APÉNDICE C

Secuencia de entrenamiento y prueba con el número de ensayos y criterios para el estudio original de Spencer y Chase (1996) y para la Simulación 4 del Modelo Neurocomputacional 2.

Spencer & Chase (1996)							Simulación 4						Criterios	
Etapas de Entrenamiento	Relaciones Entrenadas						Etapas de Entrenamiento	Relaciones Entrenadas						
	AB	BC	CD	DE	EF	FG		AB	BC	CD	DE	EF		FG
	Número de Ensayos por Bloque							Número de Ensayos por Bloque						
1. AB	48						1. AB	48						90% correctos
2. BC	24	24					2. BC	24	24					90% correctos
3. CD	12	12	24				3. CD	12	12	24				90% correctos
4. DE	8	8	8	24			4. DE	8	8	8	24			90% correctos
5. EF	6	6	6	6	24		5. EF	6	6	6	6	24		90% correctos
6. FG	3	3	3	6	9	24	6. FG	3	3	3	6	9	24	90% correctos
7. Mantenimiento de Línea Base/ Sin reforzamiento	3	3	3	3	3	3	7. Mantenimiento de Línea Base/ Sin reforzamiento	3	3	3	3	3	3	90% correctos en 5 bloques consecutivos en el estudio original. 90% correctos en un bloque en el modelo.
Etapas de Prueba							Pruebas							
1. Pruebas Combinadas	36 ensayos de línea base 1 ensayo de cada uno de los 45 ensayos combinados						Pruebas Combinadas, de Transitividad y de Simetría	18 ensayos de línea base 18 ensayos de simetría 45 ensayos combinados 45 ensayos de transitividad						90% correctos/ Sin criterio para el modelo
2. Pruebas de Transitividad	36 ensayos de línea base 1 ensayo por cada uno de los 45 ensayos de transitividad													90% correctos/ Sin criterio para el modelo
3. Pruebas de Simetría	15 ensayos de pruebas combinadas 36 ensayos de línea base 1 ensayo por cada uno de los 18 ensayos de simetría													90% correctos/ Sin criterio para el modelo
	15 ensayos de pruebas combinadas 15 ensayos de pruebas de transitividad													
Tipos de Ensayo	Relación entre Estímulos Evaluada para las Clases 1, 2, y 3													
Línea Base	AB, BC, CD, DE, EF, FG													
Simetría	BA, CB, DC, ED, FE, GF													
Transitividad	AC, AD, AE, AF, AG, BD, BE, BF, BG, CE, CF, CG, DF, DG, EG													
Combinadas	CA, DA, EA, FA, GA, DB, EB, FB, GB, EC, FC, GC, FD, GD, GE													

A CONNECTIONIST MODEL OF STIMULUS CLASS FORMATION WITH A YES/NO PROCEDURE AND COMPOUND STIMULI

Angel E. Tovar and Alvaro Torres Chávez

Universidad Nacional Autónoma de México

We analyzed stimulus class formation in a human study and in a connectionist model (CM) with a yes/no procedure, using compound stimuli. In the human study, the participants were six female undergraduate students; the CM was a feed-forward back-propagation network. Two 3-member stimulus classes were trained with a similar procedure in both the human participants and the model. During training, the compound stimuli A1B1, A2B2, B1C1, B2C2 were associated with the yes response, and A1B2, A2B1, B1C2, B2C1 were associated with the no response. During testing, we probe the emergence of the yes and no responses in the presence of new compound stimuli (A1C1, A2C2, A1C2, A2C1). Four participants and five out of six runs of the CM exhibited stimulus class formation. These results show the usefulness of the yes/no procedure for simulating stimulus class formation in a CM in contrast with the models inspired in matching-to-sample procedures.

Key words: connectionist modeling, stimulus class formation, compound stimuli, yes/no procedure, adult humans

Stimulus classes or categories have been defined as stimulus sets that occasion common responses in a specific context; such sets include stimuli involved in an explicit learning history and new stimuli to which behavioral control functions acquired by such history can be transferred (Zentall, Galizio, & Critchfield, 2002).

Research on stimulus equivalence classes in humans—as a case of category formation—has been done principally with the matching-to-sample (MTS) procedure, in which conditional relations between some stimuli of a potential class can be established. For example, in order to establish a class formed by stimuli A1, B1, and C1, relations A1rB1 and B1rC1 are trained. In order to evaluate class formation, the relations that were not directly trained, known as emergent relations, are tested through symmetry (i.e., B1rA1 and C1rB1), transitivity (i.e., A1rC1), and equivalence trials (i.e., C1rA1; see Sidman, 1992, 1994, 2000; Sidman & Tailby, 1982).

Research on equivalence classes through the use of connectionist models (CMs) has been done. These models, also known as artificial neural networks, can simulate complex

This study was conducted in partial fulfillment of the requirements of the PhD degree by the first author at Universidad Nacional Autónoma de México. The first author was supported by a fellowship from the Mexican National Council for Science and Technology (CONACYT). We thank Dr. Florente Lopez for his thoughtful suggestions on the manuscript.

Correspondence concerning this article should be addressed to Angel E. Tovar or Alvaro Torres Chávez, both at Departamento de Psicología Experimental, Universidad Nacional Autónoma de México, Av. Universidad #3004, México, Distrito Federal, Coyoacán, Código postal 04510. E-mail: eugeniot@comunidad.unam.mx or alvarot@unam.mx

behavior by learning to produce response patterns that are similar to the responses observed in humans when a similar stimulation is presented.

CMs are considered research tools that enable behavioral scientists to explore, analyze, and question data obtained by other strategies (for a review of connectionism, see McClelland, 2009; Plunkett & Elman, 1997). It has been proposed that CMs offer advantages such as the complete control of stimuli introduced to the network; therefore, phenomena are explored without the complication of an unspecified preexperimental learning history (Lyddy & Barnes-Holmes, 2007).

CMs consist of a number of interconnected processing units. Each unit receives, computes, and propagates information, which is represented by different activation values. The connections between the units are defined by a weight value. Units are organized in different layers, simulating either specific brain structures or behavioral processes, by a system with input/stimulation, hidden/representation, and output/response layers.

The activation values of the input layer are specified directly by the experimenter and represent different stimulation environments; the activation values of the hidden and output layers are generated by the functioning of the model. CMs are able to learn complex tasks by modifying their connection weights by means of a learning algorithm (see, e.g., McClelland, 2009; O'Reilly & Munakata, 2000; Thomas & McClelland, 2008).

These models have been used in the study of diverse psychological phenomena, such as pattern recognition (Rumelhart, McClelland, & PDP Research Group, 1986), conditional discriminations (Maki & Abunawass, 1991), generalization (Gluck, 1991), and early word learning (Mayor & Plunkett, 2010). CMs have been used also for the study of equivalence class formation. Barnes and Hampson (1993) proposed a model called RELNET—there are other RELNET-derived models (e.g., Cullinan, Barnes, Hampson & Lyddy, 1994; Lyddy & Barnes-Holmes, 2007; Lyddy, Barnes-Holmes, & Hampson, 2001). These authors pointed out that one contribution of RELNET models has been to provide evidence showing that CMs are able to respond correctly to relations between stimuli that have not been directly trained, just like human participants do.

The network developed by Barnes and Hampson (1993) has three principal elements in the input pattern: The first element represents the actual stimuli that function as samples and comparisons; the second element (the “sample-marking duplicator”) marks the stimulus that functions as the sample; and the third element represents the contextual stimulus (same, different, and opposite). Barnes and Hampson (1993) pointed out that

the second element of the input . . . basically copies, or mirrors, the activation from each task . . . and marks one of the stimuli as a sample from the task (e.g., when A1 is activated as a sample with B1 and B2 as comparisons, then Z1, Z1/s, Z2, and Z3 are activated . . .). The sample-marking duplicator mirrors activation in exactly the same way for each individual task across each of the eight stimulus sets (e.g., when W1 is activated as a sample with X1 and X2 as comparisons then Z1, Z1/s, Z2, and Z3 are activated in the sample marking duplicator). (p. 625)

During probes for emergent relations, the stimulation patterns are identical regarding the second element of the input (sample-marking duplicator) to some of the training trials (look at Z input units in the previous paragraph). Thus, even when the first element changes depending on the particular trial, the marking element of the sample stimulus remains constant. This has a determinant influence on the activation values obtained in the output layer; as a consequence, RELNET models do not give an “emergent” response in the test trials. The outputs generated by these networks result from previously learned information about the sample marking.

In order to avoid stimuli marking problems, one alternative is to design a simulation that will not require the representation of sample and comparison stimuli functions. In this context, the go/no-go and yes/no procedures with compound stimuli have been suggested as alternatives to the MTS procedure for the study of stimulus class formation (Debert,

Huziwara, Faggiani, Simões de Mathis, & McIlvane, 2009; Debert, Matos & McIlvane, 2007; Fields, Doran, & Marroquin, 2009). In these procedures, compound stimuli are formed by two components (e.g., abstract figures) that belong either to the same class or to different classes, arbitrarily defined by the researcher.

In the *go/no-go* procedure, participants are required to emit a response (*go*) in the presence of compound stimuli with components that belong to the same class and not emit the response (*no-go*) in the presence of compound stimuli with components that belong to different classes (Debert et al., 2007).

In the *yes/no* procedure, participants emit two different responses; when the two components of a compound stimulus belong to the same class, participants should respond with the *yes* option; when the two components do not belong to the same class, participants should respond with the *no* option (Fields et al., 2009).

Using *go/no-go* responses, Debert et al. (2007) established three 3-member stimulus classes. The participants were taught to respond (*go*) in the presence of compound stimuli formed by the pairs A1B1, B1C1, A2B2, B2C2, A3B3, and B3C3 and not to respond (*no-go*) when the stimulus pairs A1B2, A1B3, B1C2, B1C3, A2B1, A2B3, B2C1, B2C3, A3B1, A3B2, B3C1, and B3C2 are presented. During emergent relation probes, four out of six participants responded (*go*) when facing new stimulus pairs whose components belonged to the same class (i.e., A1C1, A2C2, and A3C3); and they did not respond (*no-go*) in the presence of stimulus pairs whose components belonged to different classes (i.e., A1C2, A1C3, A2C1, A2C3, A3C1, and A3C2).

Fields and colleagues (2009) showed the formation of stimulus classes through compound stimuli with *yes/no* and *same/diff* responses. In their second experiment, they found that 100% of the participants in the preliminary training condition (i.e., generalized transitivity repertoire) formed all equivalence classes; such preliminary training consisted of teaching all the stimulus relations in different equivalence classes. Transitivity trials were presented with direct reinforcement instead of testing under extinction.

The studies with compound stimulus procedures showed that it is not essential to consider different functions for stimuli, which, in an MTS arrangement, work as conditional/sample or discriminative/comparison (Debert et al., 2007; Debert et al., 2009). We consider that these procedures are preferable when working with computational simulations because they allow the design of parsimonious CMs; these models can be centered on the analysis of stimulus relations without considering special stimulus functions.

Therefore, our main purpose was to develop a CM to evaluate stimulus class formation, avoiding the problems generated by the marking of stimulus functions and allowing an appropriate evaluation of CMs capability to respond correctly to emergent relations between stimuli.

In this context, we carried out a stimulus class formation study with human participants using compound stimuli and *yes/no* responses. This was made in order to obtain data to contrast the performance of the CM that simulates stimulus class formation.

Method

Participants

Six female undergraduate students, ages 19 to 22 years, participated in this study. None of them had previous experience with class formation experimental tasks. They received partial course credit for their participation, regardless of performance. Informed consent was obtained from all participants.

Apparatus

Sessions were conducted in a 4 m × 6 m room; each participant was seated facing a touch-screen monitor connected to a Dell computer. A custom software developed in

Visual Basic controlled all stimulus presentations and recorded all responses. The stimuli were six abstract figures such as the ones used by Debert et al. (2007), originally developed by Markham and Dougher (1993); they were designated as A1, B1, C1, A2, B2, and C2 (see Figure 1).

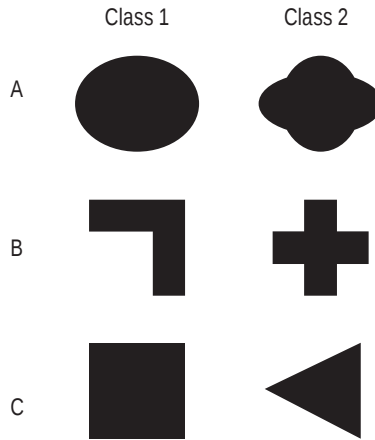


Figure 1. Abstract figures used as the components of compound stimuli and their designations.

Procedure

The procedure was similar to the one used by Debert et al. (2007), but instead of working with go/go-no we used yes/no responses, the training phase was divided into six stages, and we worked with two experimental classes instead of three. The experiment was carried out in two sessions conducted in two consecutive days.

Session 1, Phase 1: Training of baseline relations. During this phase, the compound stimuli were introduced and participants chose between the *yes* and *no* responses for each stimulus pair. The *yes* response was reinforced in the presence of the pairs A1B1, B1C1, A2B2, and B2C2; the *no* response was reinforced in the presence of the pairs A1B2, B1C2, A2B1, and B2C1. The designation of each stimulus pair makes reference to the spatial location of the abstract figures: When A1B1 is read, A1 appears on the left side and B1 on the right side (see Figure 2).

At the beginning of Session 1, participants were asked to read the following instructions (translated from the Spanish).

Thank you for your participation.

The following study is not an intelligence test and will not evaluate any aspect of your intellectual abilities. At the end of the study you will receive a complete explanation. A researcher will remain inside the room in order to help you if a technical problem arises, but he will not provide you with assistance to answer the test.

Your goal is to obtain as many points as possible; these points will be shown on a counter placed in the upper left corner of the screen.

At the center of the screen, two figures will be displayed. Your task is to choose *yes* by touching the screen on the *yes* option whenever you consider that figures are related, and you will touch the *no* option whenever you consider figures are not related.

You will be able to learn which figures are related and which are not by checking the points on the counter.

The task seems simple, but it will get more difficult as it goes on, so pay attention.

Please repeat these instructions to the experimenter, and when he approves, touch in the “Go on” box.

After touching the “Go on” box, the trials began. On each trial, a compound stimulus appeared at the center of the screen, and the *yes* and *no* responses were at the bottom. There was a counter on the upper left corner of the screen with the total points displayed in black (see Figure 2).

When participants made a correct response, the counter increased its value by 5 points and flickered in green for 2 seconds; at the same time, the participant heard an “approval” sound. When the responses were wrong, the counter decreased its value by 5 points and flickered in red for 2 seconds, accompanied by an “error” sound.

During the 2 seconds of the flickering, the compound stimulus and the response options remained visible on the screen; however, no consequences were programmed for responding during that period. The trials were separated by 1 s intertrial intervals. The next trial started with the value on the counter in black.

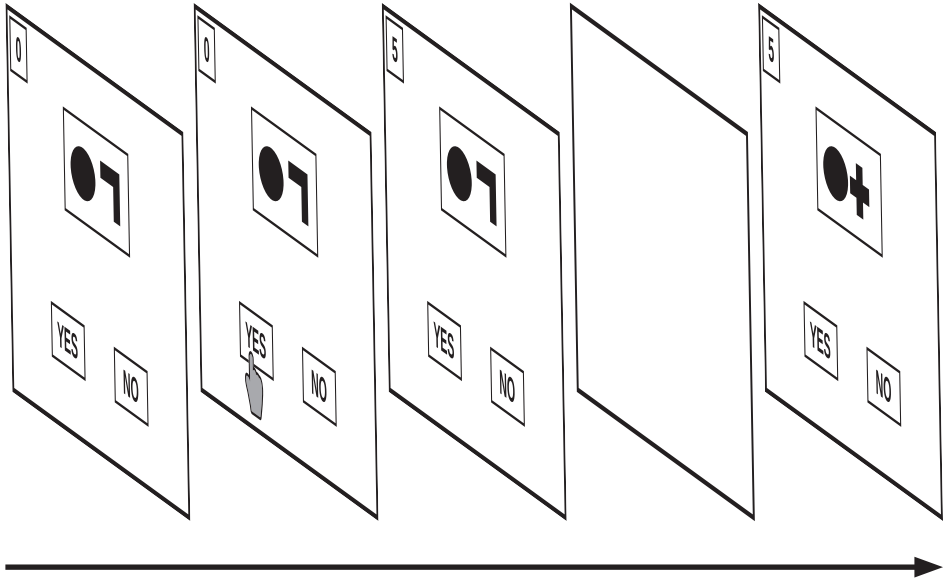


Figure 2. An example of the events presented during training. From left to right, the first frame shows the trial A1B1, in the second frame the *yes* response is selected, the third frame shows the addition of 5 points on the counter, the intertrial interval appears in the fourth frame, and the next trial is shown in the last frame.

Phase 1 was divided into six stages. Table 1 shows the sequence of stages and learning criteria used during this phase. The trained relations in each stage were presented in a random order; there were no more than three consecutive trials of the same type. Each stage was repeated until participants reached the mastery criterion.

During Stages 1 to 5, all responses generated the programmed consequences. At the beginning of Stage 6 the participants were informed about the lack of feedback, and they read the next instructions in the computer screen: “In the next trials, the counter and the sounds will be removed. However, your correct and wrong responses will be considered. Keep doing your best in accordance with what you have learned.”

Session 1 was extended until participants met the learning criteria of Phase 1; no more than 30 minutes were required for each participant.

Table 1
Sequence of Stages During Phases 1 and 2

Stage	Compound Stimuli	Correct Response	Reinforcement Percentage	Phase 1 Training		Phase 2 Retraining	
				Number of Trials	Criteria	Number of Trials	Criteria
1	A1B1	YES	100%	12	21/24	6	11/12
	A1B2	NO	100%	12		6	
2	B1C1	YES	100%	12	21/24	6	11/12
	B1C2	NO	100%	12		6	
3	A2B2	YES	100%	12	21/24	6	11/12
	A2B1	NO	100%	12		6	
4	B2C2	YES	100%	12	21/24	6	11/12
	B2C1	NO	100%	12		6	
5	A1B1	YES	100%	3	23/24	2	15/16
	B1C1	YES	100%	3		2	
	A2B2	YES	100%	3		2	
	B2C2	YES	100%	3		2	
	A1B2	NO	100%	3		2	
	B1C2	NO	100%	3		2	
	A2B1	NO	100%	3		2	
	B2C1	NO	100%	3		2	
6	A1B1	YES	0%	3	23/24	2	15/16
	B1C1	YES	0%	3		2	
	A2B2	YES	0%	3		2	
	B2C2	YES	0%	3		2	
	A1B2	NO	0%	3		2	
	B1C2	NO	0%	3		2	
	A2B1	NO	0%	3		2	
	B2C1	NO	0%	3		2	

Note. The compound stimuli, percentage of reinforcement, total number of trials, and criteria are shown for each stage.

Session 2, Phase 2: Retraining of baseline relations. During this phase, the baseline relations were retrained. The six stages of Phase 1 were presented with half of the trials. Table 1 shows information about the total number of trials and the learning criteria for each stage in Phase 2.

After completion of Phase 2, participants moved to emergent relations tests. Phases 3 and 4 tested for symmetry-like and transitivity-like relations, respectively. We use these terms because, traditionally, *symmetry* and *transitivity* are used in MTS procedures.

Phase 3: Symmetry-like tests. This phase tested for symmetry-like relations; components of the stimulus pairs were presented in a reversed spatial location. For example, if A1 was located at the left side of B1 during training (A1B1), in the symmetry-like test, B1 was located at the left side of A1 (B1A1). There were eight symmetry-like

trials, one for each trained relation (see Table 2). Each relation appeared four times in a random order; thus, a total of 32 trials were presented. At the beginning of this phase, participants were informed about the lack of feedback. At the end of the 32 trials, participants moved to Phase 4.

Phase 4: Transitivity-like tests. During this phase we tested the emergence of the *yes* and *no* responses in the presence of new compound stimuli. A total of 96 trials were presented, 48 of transitivity-like relations and 48 of equivalence-like relations. Each relation was presented 12 times in a random order (see Table 2). At the beginning of this phase, participants were informed again about the lack of feedback. Session 2 lasted no more than 25 minutes.

Table 2
Compound Stimuli Presented During Phases 3 and 4

Compound Stimuli	Phase 3		Phase 4		
	Number of Trials	Correct Response	Compound Stimuli	Number of Trials	Correct Response
B1A1	4	YES	A1C1	12	YES
B2A2	4	YES	A2C2	12	YES
B1A2	4	NO	A1C2	12	NO
B2A1	4	NO	A2C1	12	NO
C1B1	4	YES	C1A1	12	YES
C2B2	4	YES	C2A2	12	YES
C1B2	4	NO	C1A2	12	NO
C2B1	4	NO	C2A1	12	NO

Results and Discussion

All six participants completed the two experimental sessions. Four out of six participants formed the two stimulus classes (A1B1C1 and A2B2C2); in Phases 3 and 4, they responded in accordance with the stimulus class memberships established during training with accuracy levels of 85% or higher.

Table 3 presents the number of stage repetitions and number of trials required to complete the training and retraining phases as well as the number and percentage of correct trials during symmetry-like and transitivity-like tests. In most cases, participants required only one stage repetition in order to achieve the criterion. An outstanding exception was Participant 101, who required five repetitions of Stage 5 to meet the criterion. Her errors were not made in a particular relation; however, during the retraining phase, she made only one error in Stage 4 and four errors in Stage 5. The other participants showed a minimal number of errors during retraining stages. Participant 106, for example, showed an errorless performance in all retraining stages.

During tests for emergent relations (Phases 3 and 4), Participants 101, 105, and 106 showed the highest performance by responding correctly in at least 96% of the trials. Participant 103 achieved accuracy levels of 93.7% correct and 96.8% correct in Phases 3 and 4, respectively. These four participants were the ones who formed classes.

Participant 102 responded correctly to all the symmetry-like trials; however, she achieved 60% correct in the transitivity-like trials. She made 36 errors. The trials in which she made most errors were A1C1 (12 errors) and C1A1 (9 errors). For these two stimulus pairs, the *yes* response was expected and she responded with *no*. The 15 remaining errors were made in the presence of the other compound stimuli.

Participant 104 responded correctly to all symmetry-like trials, but in the transitivity-like tests she achieved 56% correct. At the beginning of Phase 4, her errors were distributed across all compound stimuli—the ones whose components were related and the ones whose components were not related—but throughout the course of the trials she began to show more errors in the presence of compounds whose components were related while she

Table 3
Performance of the six participants during Sessions 1 and 2

Participant	Session	Training and Retraining (Phases 1 and 2)						Tests (Phases 3 and 4)		
		Stage 1	Stage 2	Stage 3	Stage 4	Stage 5 All relations reinf. 100%	Stage 6 All relations reinf. 0%	Symmetry-like trials	Transitivity-like trials	
101	1	1 (21/24)	1 (23/24)	1 (24/24)	1 (24/24)	5 (86/120)	3 (59/72)			
	2	1 (12/12)	1 (12/12)	1 (12/12)	1 (11/12)	2 (28/32)	1 (16/16)	(32/32) 100%	(94/96) 97.9%	
102	1	2 (39/48)	1 (23/24)	1 (21/24)	1 (22/24)	2 (44/48)	1 (23/24)			
	2	1 (11/12)	1 (12/12)	1 (12/12)	1 (12/12)	1 (15/16)	2 (30/32)	(32/32) 100%	(60/96) 62.0%	
103	1	1 (22/24)	1 (23/24)	1 (24/24)	1 (23/24)	1 (22/24)	1 (24/24)			
	2	1 (12/12)	1 (12/12)	1 (12/12)	1 (11/12)	1 (16/16)	1 (16/16)	(30/32) 93.7%	(93/96) 96.8%	
104	1	1 (22/24)	1 (22/24)	1 (23/24)	1 (24/24)	1 (23/24)	1 (24/24)			
	2	1 (12/12)	1 (11/12)	1 (12/12)	1 (12/12)	1 (15/16)	1 (16/16)	(32/32) 100%	(54/96) 56.0%	
105	1	2 (44/48)	1 (22/24)	1 (23/24)	1 (24/24)	1 (23/24)	1 (24/24)			
	2	1 (12/12)	1 (12/12)	1 (12/12)	1 (12/12)	1 (15/16)	1 (16/16)	(31/32) 96.8%	(96/96) 100%	
106	1	1 (21/24)	1 (23/24)	1 (24/24)	1 (24/24)	1 (24/24)	1 (24/24)			
	2	1 (12/12)	1 (12/12)	1 (12/12)	1 (12/12)	1 (16/16)	1 (16/16)	(32/32) 100%	(96/96) 100%	

Note. The number of stage repetitions is indicated in bold type. The number of correct responses and total of trials are in parentheses. Percentage of correct performance during testing is also shown.

showed correct responses in the presence of compounds whose components were not related; this was because she responded *no* in the last 25 trials. This suggests that Participant 104 could use the *no* response by default; this type of performance has been described previously by Fields et al. (2009).

As suggested by Debert et al. (2009), components of the compound stimuli do not have a specific function as conditional or discriminative stimuli. However, we can consider a conditionality between the compound stimuli and the response options because the display used in this study can also be considered as a MTS format, where the compound stimulus could be viewed as a conditional or sample stimulus and the *yes* and *no* responses could be viewed as discriminative or comparison stimuli. Therefore, the compound stimuli A1B1, B1C1, A2B2, B2C2 and the *yes* response can be integrated in a class, and the compound stimuli A1B2, B1C2, A2B1, B2C1 and the *no* response can be integrated in another class. However, we did not administer the necessary probes in order to evaluate this hypothesis.

Another aspect of the present results that requires discussion is the equal likelihood of the *yes* and *no* responses to emerge in the presence of new compound stimuli. As Fields et al. (2009) pointed out, during baseline training the stimulus pairs A1B1 and B1C1, associated with the *yes* response, established the basis to evoke the *yes* response in the presence of A1C1. Similarly, the stimulus pairs A1B2 and B2C1, associated with the *no* response, also established the basis to evoke the *no* response in the presence of A1C1. Likewise, the stimulus pair A2C2 could evoke the *yes* and *no* responses with equal likelihood. However, in the present study, four out of six participants used the *yes* response when the pairs A1C1 and A2C2 were presented. We think that their performance could be biased by the semantic-inclusion role that the *yes* response has acquired extraexperimentally, probably enhanced by the instructions on the use of the *yes/no* responses (see Method section). Future research should examine this hypothesis by using arbitrary figures as response options.

In the next section, we show the architecture of the CM, the experimental task used, and the obtained results.

Connectionist Model

We developed a three-layer feed-forward artificial neuronal network to simulate the human study (see Figure 3). The input layer was composed of nine units. The activation (i.e., 1) or no activation (i.e., 0) of input units represented the presence or absence of each component of the compound stimuli (e.g., the compound stimulus A1B1 was represented with the input pattern 100000010). Table 4 shows all input patterns used during training. The position of the components of the compound stimulus in the input layer was determined in a way that did not allow two components of the same class to be adjacent. This was made in order to avoid teaching spatial or sequential relations between stimuli.

The hidden layer consisted of four units. The activation values on this layer were considered as the representation developed by the network for each trial presented in the input layer; the hidden layer was fully connected to the input and the output layers. The output layer consisted of two units that represent the *yes* and *no* responses, respectively (see Figure 3).

Each input pattern activated the units in the input layer, and then the activation passed on to the hidden layer and from this to the output layer. This kind of propagation is known as *feed-forward* (see Thomas & McClelland, 2008). The spread of activation levels was determined by the connection weights, initially randomized.

In the training phase, baseline relations were established. The network learned to activate the *yes* response when the compound stimuli A1B1, B1C1, A2B2, and B2C2 were presented as inputs. It also learned to respond to the compound stimuli A1B2, B1C2, A2B1, and B2C1 by activating the *no* response (see Table 4). Training used with humans was simulated in this way.

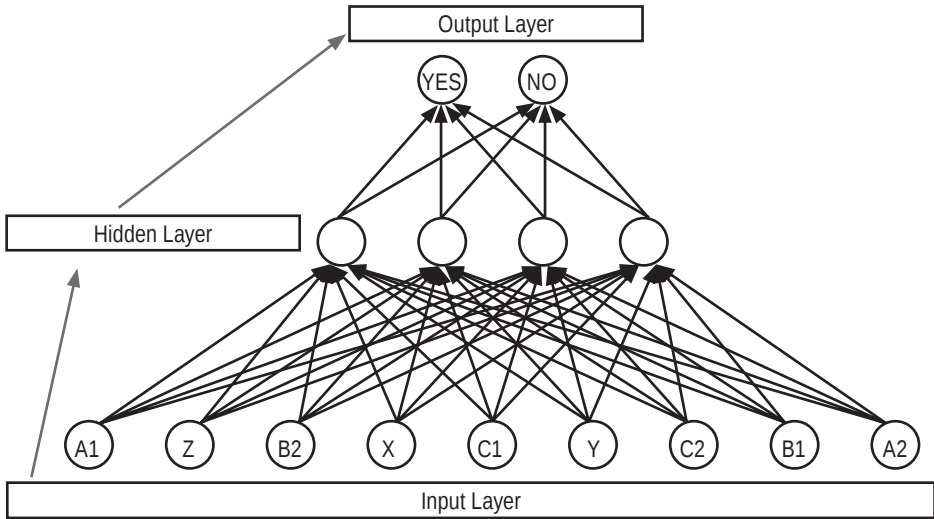


Figure 3. The connectionist model architecture.

During this phase, the model also learned all the relations between the components of a third class designated as XYZ; that is to say, the model learned to activate the yes response when the pairs XY, YZ, and XZ were presented as inputs. This constitutes a generalized transitivity repertoire (Fields et al., 2009). This was made with the purpose of simulating the previous knowledge acquired extraexperimentally by human participants about class formation.

Other connectionist models have suggested that the complete training of additional classes facilitates the formation of new equivalence classes (Barnes & Hampson, 1993; Lyddy & Barnes-Holmes, 2007). The complete XYZ class training also allows the model to learn to use differentially the units of the output layer as *yes* and *no* responses. As the units are functionally similar, the model must learn that the activation of the *yes* response unit in the output layer follows the presentation of any stimulus pair whose components belong to the same class.

Table 4
Input Patterns Presented During Training and Correct Output Patterns for Each Compound Stimulus

Compound Stimuli	Components of the Compound Stimuli in the Input Layer									Response Units in the Output Layer	
	A1	Z	B2	X	C1	Y	C2	B1	A2	YES	NO
	Input Patterns									Correct Output Patterns (trained)	
A1B1	1	0	0	0	0	0	0	1	0	1	0
A1B2	1	0	1	0	0	0	0	0	0	0	1
B1C1	0	0	0	0	1	0	0	1	0	1	0
B1C2	0	0	0	0	0	0	1	1	0	0	1
A2B2	0	0	1	0	0	0	0	0	1	1	0
A2B1	0	0	0	0	0	0	0	1	1	0	1
B2C2	0	0	1	0	0	0	1	0	0	1	0
B2C1	0	0	1	0	1	0	0	0	0	0	1
XY	0	0	0	1	0	1	0	0	0	1	0
YZ	0	1	0	0	0	1	0	0	0	1	0
XZ	0	1	0	1	0	0	0	0	0	1	0

In each trial, the connectionist model learned the correct response by comparing the activation values generated in the output layer with the ones expected in such layer (see right side of Table 4). In this way, the network found an error value to which the

connection weights were adjusted. As a consequence, in the following trials, error decreased. The model learned by error correction, modifying the connection weights through the back-propagation algorithm. For the adjustment of the weights, a learning rate of 0.3 was used. The learning rate is a constant value between 0 and 1 that determines the magnitude of change in the weights (McLeod, Plunkett, & Rolls, 1998; Rumelhart, McClelland, et al., 1986).

Training phase ended when the root mean square error (RMS) reached a value ≤ 0.05 , then we proceeded with the tests for emergent relations.

During testing, the learning algorithm was deactivated so that there were no more changes in the network due to learning processes, and the connection weights achieved at the end of the training phase were fixed. In this phase, the compound stimuli A1C1, A2C2, A1C2, and A2C1 were presented to the network. Each compound stimulus was presented only once (see Table 5). Since this model is inspired in a compound stimulus arrangement that do not consider spatial location, no symmetry-like trials were tested.

The training and testing phases were repeated in six runs of the CM, which were equivalent to the six participants of the human study; each of the six runs started at different initial random weights (between 0 and 1).

Table 5
Input Patterns Presented During Testing

Compound Stimuli	Components of the compound stimuli								
	A1	Z	B2	X	C1	Y	C2	B1	A2
	Activation patterns in the input layer								
A1C1	1	0	0	0	1	0	0	0	0
A1C2	1	0	0	0	0	0	1	0	0
A2C2	0	0	0	0	0	0	1	0	1
A2C1	0	0	0	0	1	0	0	0	1

Results and Discussion

During training, the six runs achieved an RMS error value ≤ 0.05 in an average of 13,898 iterations. This indicates that the model acquired the baseline relations properly. During testing, we considered the model execution as indicative of class formation if the *yes* unit reached an activation value ≥ 0.85 when the pairs A1C1 and A2C2 were presented and also if the *no* unit reached that value when the pairs A1C2 and A2C1 were presented in the input layer.

Based on the above criteria, Runs 1 to 5 showed class formation, in the sixth run activation values were 0.788 in the *yes* unit for the stimulus pair A1C1 and 0.841 in the *no* unit for the stimulus pair A1C2. For stimulus pairs A2C2 and A2C1, the correct output unit exceeded 0.85 activation value. However, when considering all trials, this run did not form the two stimulus classes (see Figure 4).

Cluster Analysis of Hidden Layer

In the hidden layer, the first transformation of the input patterns is done so that an intermediate representation between the input and the output patterns is developed (see Thomas & McClelland, 2008). With a cluster analysis we can observe which stimulation patterns were transformed to generate similar representations in the hidden layer. This analysis grouped similar patterns by computing the Euclidean distance between them. Figure 5 shows the graphical representation of the cluster analysis of the second run. Cluster analyses conducted on the other five runs generated similar groups; these are not presented here but can be obtained from the authors.

Three main clusters can be observed in Figure 5: the first with the stimulus pairs A1B1, B1C1, and A1C1; the second cluster with the pairs A2B2, B2C2, and A2C2; and the

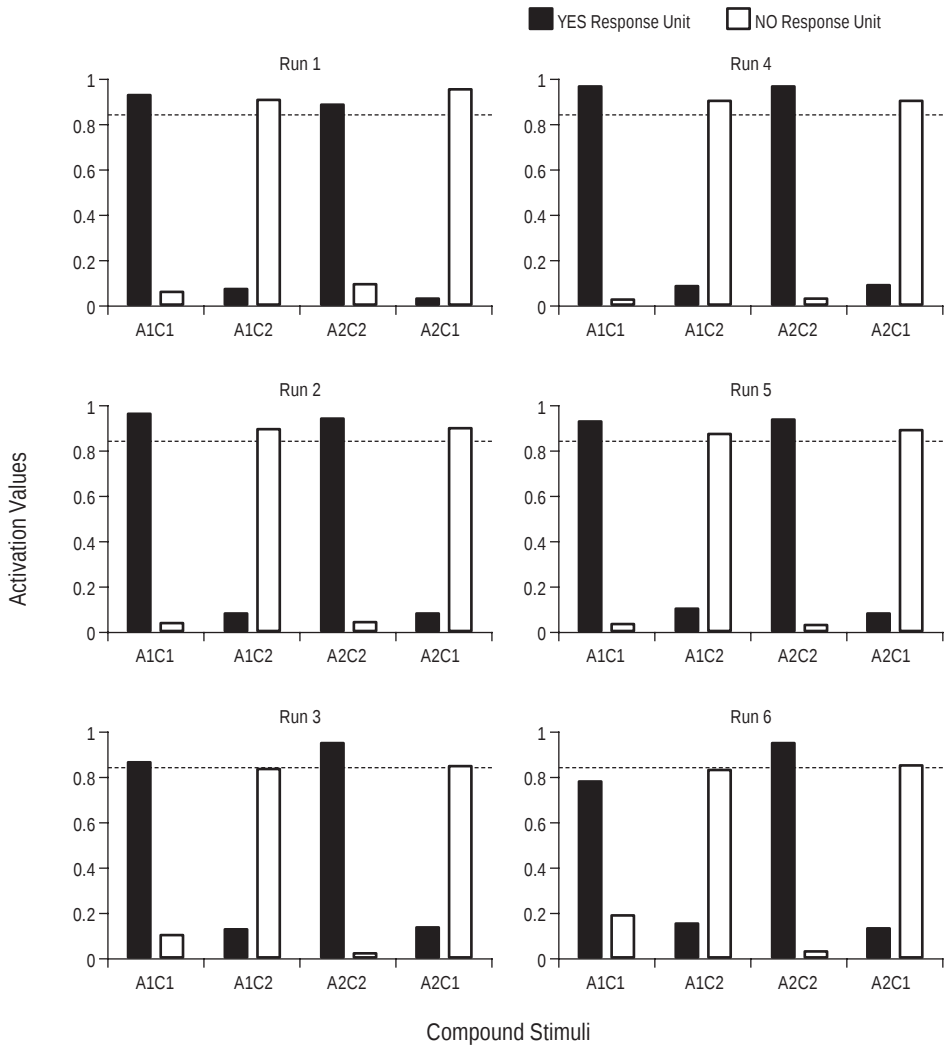


Figure 4. Activation values of the output units during testing. The compound stimuli are shown along the abscissa. Black bars correspond to the yes unit and white bars correspond to the no unit. Dotted lines show the criterion value.

third with the pairs A2C1, B1C2, A2B1, A1B2, B2C1, and A1C2. The stimulus pairs of the first two clusters were associated with the yes response; the pairs of the third one were associated with the no response.

This clustering indicates that during training in the hidden layer the network transformed the input patterns into similar activation values, depending on their class membership. For example, stimulus pair A2B2 was represented in the input layer with the pattern 001000001, and stimulus pair B2C2 was represented with the pattern 001000100. However, both pairs had an activation value of 0.9, 0.0, 0.0, and 0.9 in the four units of the hidden layer, respectively. The network established the connection weights that allowed the stimulus pair A2C2 (represented with the input pattern 000000101) to produce the values 0.9, 0.0, 0.0, and 0.9 in the four units of the hidden layer during testing.

This could be one strategy for responding correctly during arbitrary class formation tests that humans can also use; no matter how different the stimulus pairs are, all of them can evoke a similar representation.

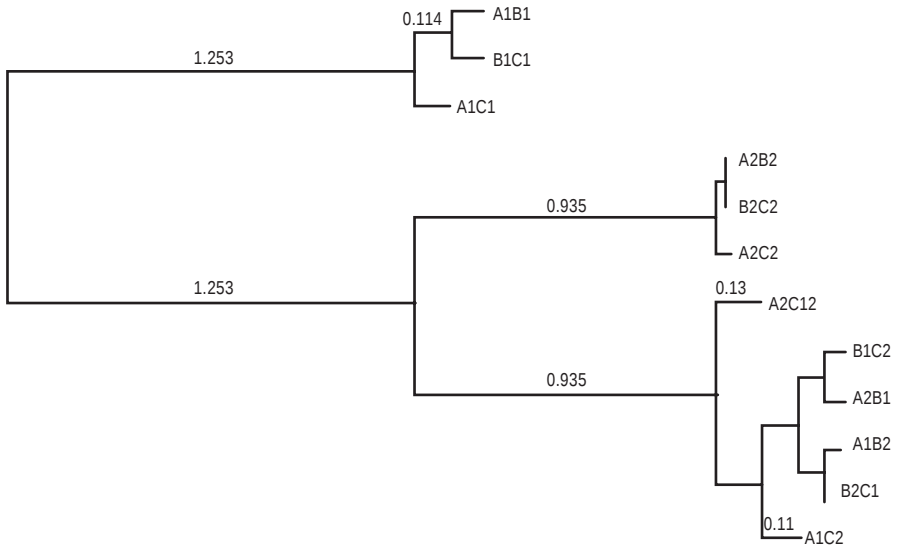


Figure 5. Cluster diagram of the hidden layer activation values during Run 2. Some Euclidian distance values are indicated between clusters. The closer the stimulus pairs, the more similar activation values they produce in the hidden units.

Cluster analysis revealed that the model classifies the compound stimuli before producing the “final” *yes* or *no* response. This final response can be considered as a general classification; multiple class memberships can be seen for the different compound stimuli. For example, pairs A1B1, B1C1, and A1C1 formed Class 1; pairs A2B2, B2C2, and A2C2 formed Class 2. These two classes are included in the *yes* class. The model is able to work with multiple class memberships. It can generate the same response for different classes and keep the integrity of each class in the hidden layer representation.

A similar process could occur with the human participants but in a behavioral level that is difficult to analyze because we can just observe the general classification response. However, the data of a postexperimental interview applied to the human participants point out that they can also distinguish between Classes 1 and 2; thus, the data obtained from the model and from the postexperimental interview seem to match.

Further Implications of the Model

In order to test the semantic bias supposition (see Results and Discussion section of the human study), we performed six new runs (7–12) of the CM. We removed the training of the XYZ class to suppress the complementary learning history associated with the response options so that they can be considered just “Response-1” and “Response-2”; this is similar to using arbitrary response options with human participants. The CM only learned to respond to the pairs A1B1, B1C1, A2B2, and B2C2 with Response-1 and to respond to the pairs A1B2, B1C2, A2B1, and B2C1 with Response-2.

Probes for emergent relations under this condition revealed as expected: that the model cannot respond consistently along all runs. Only in Runs 7 and 8, Response-1 unit was activated to the test pairs A1C1 and A2C2 and Response-2 unit was activated to the test pairs A1C2 and A2C1; in Runs 9, 11, and 12, the model inverted this activation pattern; finally, in the Run 10 the output, activation was undefined.

These results support our hypothesis and suggest that when working with two stimulus classes in a compound stimuli procedure, the use of arbitrary response options will generate a heterogeneous pattern of responses during testing.

General Discussion

The characteristics of the yes/no procedure with compound stimuli enabled the simulation of stimulus class formation in a CM. The model was capable of acquiring all baseline relations and responding correctly to the emergent relations. This model focused on the relations between the components of the compound stimuli. The simulation did not require the inclusion of sample and comparison functions. This represents a different approach to class formation connectionist modeling (see Barnes & Hampson, 1993; Cullinan et al., 1994; Lyddy & Barnes-Holmes, 2007; and Lyddy et al., 2001, for other approaches). These results showed that CMs are capable of responding correctly to nondirectly trained relations between stimuli.

The present results also extend the previous findings on arbitrary stimulus class formation in humans using yes/no procedures. This supports the idea that class formation can be analyzed with procedures other than MTS (e.g., Delbert et al., 2007; Debert et al. 2009; Fields et al., 2009). Future research should focus on detailing the role of the semantic bias of the response options and the instructions.

An advantage of the CM is that it allows us to analyze the behavioral process in an observable system. In the next paragraphs, we summarize what we consider the main contributions of the present model.

By working with compound stimuli and yes/no responses, we were able to use simple stimulation patterns. Using cluster analysis, we could observe how the model classifies the compound stimuli in the hidden layer. With this classification process, the model is able to produce the correct response; the compound stimuli that belong to the same class generated the same activation values in the hidden layer. We hypothesized that this could be one strategy that humans can also use, no matter how different the stimuli are. Once the class is already formed, all stimuli evoke a similar representation.

This model allowed us to manipulate the effect of the complete training of a third class (XYZ). When it was trained, five out of six runs exhibited stimulus class formation, and when training of XYZ was removed, the model did not form the stimulus classes along six new runs consistently. This indicates that the XYZ training established the necessary experience in order to acquire a “generalized transitivity repertoire” (e.g., Fields et al., 2009).

As mentioned in the Results and Discussion section of the human study, the experimental procedure established the basis to evoke the yes and the no responses with equal likelihood when A1C1 and A2C2 trials were presented. Nevertheless, the preferred election of the yes response by both the human participants and the model leads us to think that there is a functional similarity between the semantic bias and the generalized transitivity repertoire trained to the model. Future research should analyze stimulus class formation with compound stimuli and arbitrary responses. The present model indicates that class formation will be difficult to demonstrate under these conditions.

In general, both the human study and the CM showed the relevance of analyzing arbitrary stimulus class formation with different research strategies. Both strategies give each other feedback and generate a reciprocal heuristic value. The use of different experimental and simulation procedures extends our knowledge on the determinant factors involved in arbitrary categorization and symbolic processing.

References

- BARNES, D., & HAMPSON, P. J. (1993). Stimulus equivalence and connectionism: Implications for behavior analysis and cognitive science. *The Psychological Record*, 43, 617–638.
- CULLINAN, V., BARNES, D., HAMPSON, P. J., & LYDDY, F. (1994). A transfer of explicitly and non-explicitly trained sequence responses through equivalence relations: An experimental demonstration and connectionist model. *The Psychological Record*, 44, 559–586.

- DEBERT, P., HUZIWARA, E., FAGGIANI, R., SIMÕES DE MATHIS, M., & MCILVANE, W. (2009). Emergent conditional relations in a go/no-go procedure: Figure-ground and stimulus-position compound relations. *Journal of the Experimental Analysis of Behavior*, 92, 233–243.
- DEBERT, P., MATOS, M. A., & MCILVANE, W. J. (2007). Conditional relations with compound abstract stimuli using a go/no-go procedure. *Journal of the Experimental Analysis of Behavior*, 87, 89–96.
- FIELDS, L., DORAN, E., & MARROQUIN, M. (2009). Equivalence class formation in a trace stimulus pairing two-response format: Effects of response labels and prior programmed transitivity induction. *Journal of the Experimental Analysis of Behavior*, 92, 57–84.
- GLUCK, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science*, 2(1), 50–55.
- LYDDY, F., & BARNES-HOLMES, D. (2007). Stimulus equivalence as a function of training protocol in a connectionist network. *The Journal of Speech Language Pathology and Applied Behavior Analysis, compiled issue 1.4 & 2.1*, 14–24.
- LYDDY, F., BARNES-HOLMES, D., & HAMPSON, P. J. (2001). A transfer of sequence function via equivalence in a connectionist network. *The Psychological Record*, 51, 409–428.
- MAKI, W. S., & ABUNAWAS, A. M. (1991). A connectionist approach to conditional discriminations: Learning, short-term memory, and attention. In M. L. Commons, S. Grossberg, & J. E. R. Staddon (Eds.), *Neural networks models of conditioning and action* (pp. 241–278). Hillsdale, NJ: Erlbaum.
- MARKHAM, M. R., & DOUGHER, M. J. (1993). Compound stimuli in emergent stimulus relations: Extending the scope of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, 60, 529–542.
- MAYOR, J., & PLUNKETT, K. (2010). A neuro-computational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, 117, 1–31.
- MCCLELLAND, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, 1(1), 11–38.
- MCLEOD, P., PLUNKETT, K., & ROLLS, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. Oxford, England: Oxford University Press.
- O'REILLY, R. C., & MUNAKATA, Y. (2000). *Computational explorations in cognitive neuroscience. Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.
- PLUNKETT, K., & ELMAN, J. (1997). *Exercises in rethinking innateness. A handbook for connectionist simulations*. Cambridge, MA: MIT Press.
- RUMELHART, D. E., MCCLELLAND, J. L., & THE PDP RESEARCH GROUP (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- SIDMAN, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations. The second and third international institute on verbal relations* (pp. 15–27). Reno, NV: Context Press.
- SIDMAN, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA: Authors Cooperative.
- SIDMAN, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, 74, 127–146.
- SIDMAN, M., & TAILBY, W. (1982). Conditional discrimination vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37, 5–22.

- THOMAS, M. S. C., & MCCLELLAND, J. L. (2008). Connectionist models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 23–58). Cambridge, MA: Cambridge University Press.
- ZENTALL, T. R., GALIZIO, M., & CRITCHFIELD, T. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior*, *78*, 237–248.