



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

MODELACIÓN DE PROCESOS COGNITIVOS CON APLICACIONES EN
EDUCACIÓN MATEMÁTICA.

TESIS PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

ABELINO ENRIQUE JIMÉNEZ GAJARDO

PROFESOR GUÍA:
PABLO DARTNELL ROY

MIEMBROS DE LA COMISIÓN:
ROBERTO ARAYA SCHULZ
RAFAEL CORREA FONTECILLA

SANTIAGO DE CHILE
OCTUBRE 2012

Resumen

Esta memoria fue concebida con el intento de ser un aporte a la Educación desde la Modelación Matemática de Procesos Cognitivos. En el Capítulo 1, se hace un recorrido por los principales aspectos de la Cognición Matemática y el Sentido Numérico, analizando el porqué de su importancia y los distintos enfoques para su estudio.

El Capítulo 2 trata del problema de Representación de Estímulos y cómo medir su grado de similaridad. Junto con mostrar resultados clásicos de Psicofísica, se analizan los dos modelos de representación numérica más conocidos; el modelo de escala comprimida y el modelo lineal con variabilidad escalar. Se demuestra que ambos modelos pertenecen a una familia más general de representaciones que cumplen con las mismas propiedades fundamentales. Además, se describe una situación experimental en donde los modelos hacen predicciones distintas.

En el Capítulo 3 se trabaja el problema de estimación de cantidades, esto es, frente a un determinado estímulo numérico no simbólico, por ejemplo, nubes de puntos, secuencias de tonos, etc. se pide estimar la numerosidad del estímulo. Para ello se propone y estudian dos modelos para explicar dicho fenómeno. Para uno de estos modelos, se obtiene predicciones acordes con los principales resultados experimentales conocidos.

Finalmente, en el Capítulo 4, se aborda el tema del Aprendizaje de Fracciones. Para ello, se realizó un estudio con más de 200 niñas y niños de cuarto básico, con el fin de comparar tres métodos de enseñanza. Se comparó la representación Geométrica de fracción (por ejemplo, $1/2$ es equivalente a la mitad de un cuadrado), con una representación Temporal ($1/2$ equivale a recibir un dulce cada dos días) y con una representación de Intercambio ($1/2$ equivale a recibir un objeto por cada dos monedas que doy). La primera es una de las más utilizadas en textos escolares y en el aula, mientras que las otras son poco conocidas o se suelen presentar en otros contextos. Se encontraron diferencias en desempeño a favor de las representaciones Temporal y de Intercambio en la tarea de comparación de fracciones, siendo un hallazgo promisorio para la educación. Junto con esto, aprovechando la información se proponen tres metodologías para representar estímulos en una escala numérica y se construyó un modelo que permite estimar el impacto de cada método de enseñanza en el cambio de estrategias que los estudiantes utilizan para comparar fracciones. Se concluye que con períodos de entrenamiento breves, los estudiantes son capaces de cambiar la manera de contestar preguntas de fracciones fuera de un contexto metafórico, siendo esto una extensión de las investigaciones en torno a este tema.

A mis padres por su apoyo incondicional.
A Nicole por su compañía, paciencia y amor.

Agradecimientos

En primer lugar, me gustaría agradecer a mi profesor guía, el señor Pablo Dartnell, quien con su infinita confianza y paciencia me orientó en el desarrollo de este trabajo. Siempre escuchó con atención y entusiasmo mis ideas, ayudándome a concretarlas y perfeccionarlas.

Quisiera también agradecer al profesor Roberto Araya, por estar permanentemente disponible para discutir y orientar mis inquietudes entorno a la educación y la enseñanza de la matemática.

También me gustaría dar gracias a David Gómez, quien a pesar de la distancia, estuvo muy interesado en el progreso de esta memoria y siempre dispuesto a colaborar y aclarar mis dudas sobre el apasionante mundo de la neurociencia.

Agradezco además al profesor Jorge Soto-Andrade, quien siempre se mostró muy interesado en el estudio de fracciones aquí desarrollado, y colaboró en la confección del diseño experimental de éste.

Doy también las gracias a mis compañeros del Departamento de Ingeniería Matemática, en especial a Johan Van der Molen, Valentina Giaconi, Emilio Vilches, Mónica Carvajal, Sebastián Reyes Ríffo y Francisco Plana. Siempre estuvieron disponibles para conversar, escuchar y discutir nuevas ideas que podrían ser útiles para abordar los problemas que les presentaba.

No puedo olvidar a mis amigos de siempre Laura, Gonzalo y Tomás. Por acompañarme durante estos años y crecer juntos como profesionales y personas.

Quiero agradecer de forma especial a mis padres, Abelino y Evangelina, quienes me dieron educación, y juntos a mi hermana Eliana, confiaron siempre en mí incondicionalmente.

Finalmente, tengo que dar las gracias a mi novia Nicole. Por soportar mi mal humor, por darme fuerzas para continuar y por acompañarme en este y futuros caminos venideros.

Índice general

Introducción	1
1. Del Sentido Numérico	3
1.1. Alcances del sentido numérico, ¿Por qué es un tema que debe ser tratado?	4
1.2. Sentido Numérico en:	5
1.2.1. Animales	5
1.2.2. Niños y bebés	6
1.3. Resultados Clásicos de la Psicología Conductual	8
1.4. Aspectos Neurobiológicos Involucrados	9
1.5. Algunos Modelos del Sentido Numérico	12
1.6. Primeras Conclusiones	16
2. Representaciones y Medidas de Similaridad	17
2.1. Representación de Estímulos	18
2.1.1. Modelos de Representación Aleatoria para Numerosidades	23
2.2. Construcción de Métricas Fechnerianas	28
2.2.1. Aplicaciones a Modelos de Numerosidad	30
2.3. Método de Escalamiento Multidimensional	40
2.3.1. Un Ejemplo Aplicado	43
2.4. Conclusiones	45
3. Estimación y Modelo de Acumuladores	47
3.1. Modelo de Acumuladores Difusos	49
3.1.1. Dinámica de Acumuladores	50
3.1.2. Solución	51
3.1.3. Evaluación para caso particular	52
3.2. Modelo de Competencia	58
3.2.1. Esperanza de la Estimación	59
3.2.2. Evaluación de Caso Particular	59
3.2.3. Variabilidad Escalar	63
3.3. Conclusiones y Preguntas Abiertas	64
4. Un Estudio sobre la Enseñanza de Fracciones	65
4.1. Marco Teórico	66
4.1.1. La dificultad en la enseñanza de fracciones	66
4.1.2. La importancia del buen aprendizaje de las fracciones	67
4.1.3. Representaciones Privilegiadas y Mecanismos Ecológicamente Válidos	68

4.1.4. Otros estudios sobre fracciones	69
4.2. Presentación del Experimento	71
4.2.1. Objetivo	71
4.2.2. Participantes	72
4.2.3. Materiales	72
4.2.4. Descripción del Test	72
4.3. Principales Resultados	76
4.3.1. Primeras Conclusiones del Estudio	78
4.4. Escalamiento de Fracciones	80
4.4.1. Método de Escalamiento Multidimensional	81
4.4.2. Método de Máxima Verosimilitud	82
4.4.3. Método de Camino Máximo	85
4.4.4. Comparación de los Métodos	88
4.5. Estimación de Estrategias	91
4.5.1. Resultados	92
4.6. Conclusiones del Estudio	93
Conclusión	95
Bibliografía	96

Índice de tablas

4.1. Tiempo y Contenido de Videos de la Etapa de Entrenamiento	75
4.2. Clasificación de Preguntas	76

Índice de figuras

1.1. Animales en Estudios de Cognición Numérica	5
1.2. Niña realizando el Test de Conservación Numérica	7
1.3. Promedio de Estimaciones en Función de los Estímulos	10
1.4. Vista Lateral del Cerebro de un Mono	10
1.5. Actividad de distintas Neuronas Selectivas de Numerosidad	11
1.6. Actividad de las Neuronas monótonas en función de la numerosidad expuesta.	12
1.7. Esquema del Modelo Modo-Control de Meck y Church	13
1.8. Esquema del Modelo de Filtro Neuronal de Dehaene y Changeux	13
1.9. Modelos de Escala Comprimida y Lineal con Variabilidad Escalar	14
2.1. No cumplimiento de la Minimalidad Regular en el Modelo Lineal	38
2.2. Esquema de cada trial del Experimento	43
2.3. Representación en dos dimensiones del Primer Conglomerado de Estímulos .	45
2.4. Representación en dos dimensiones del Segundo Conglomerado de Estímulos	45
3.1. Esquema del Modelo de Estimación Numérica de Dehaene	48
3.2. Ejemplo de Partición de la Recta Numérica Mental.	49
3.3. Comparación de funciones de asociación a números positivos	53
3.4. Estimaciones según Modelo de Competencias	62
4.1. Error común al ubicar fracciones en la recta numérica	66
4.2. Problema en la Representación Geométrica de Fracción	71
4.3. Esquema del Diseño Experimental	73
4.4. Ejemplo de Preguntas del Pre Test	74
4.5. Resultados Pre Test	78
4.6. Resultados PostTest	79
4.7. Nivel de Aprendizaje entre PreTest y PostTest	80
4.8. Problema del Método de Escalamiento Multidimensional	82
4.9. Consistencia del Método de Máxima Verosimilitud	85
4.10. Situaciones simples de existencia de camino Hamiltoniano	86
4.11. Situación General de existencia de camino Hamiltoniano	87
4.12. Contraejemplo sobre condición suficiente de Camino Hamiltoniano.	87
4.13. Consistencia del Método de Camino Máximo	89
4.14. Escalamiento con Máxima Verosimilitud según Tipo de Entrenamiento . . .	89
4.15. Camino Máximo según Tipo de Entrenamiento	90
4.16. Variación de valores promedios de α según tipo de entrenamiento	93

Introducción

En 2009, Chile participó como miembro de la OCDE en la prueba PISA (Programme for International Student Assessment). Los resultados obtenidos por los estudiantes chilenos fueron más que insatisfactorios. De un total de 65 países, Chile quedó en la posición 49 en la prueba de Matemáticas. Países como Azerbaiyán o Rumania tuvieron mejores desempeños que nuestro país. [10]. El problema que tiene nuestra nación en educación es serio, por lo que deben estar puestas todas las energías en tratar de solucionarlo.

Respecto a la educación matemática, una de las piezas claves es entender cómo los niños y niñas construyen sus representaciones mentales de los distintos conceptos matemáticos. Al entender cómo los alumnos van comprendiendo y asimilando las ideas matemáticas, la tarea educativa podría llegar a ser mucho más fructífera.

Este trabajo intenta hacer un aporte a la Educación desde la Modelación Matemática de fenómenos de la Psicología Cognitiva y la Percepción Numérica. Para ello, bajo el marco de la Cognición Matemática, se pondrá atención a uno de los conceptos más fundamentales de la matemática, el concepto de número.

La Cognición Matemática estudia el modo en que se entienden los diversos conceptos matemáticos, haciendo un fuerte uso de todas las herramientas de las ciencias cognitivas. Una de las más importantes líneas de investigación es el estudio del Sentido Numérico, esto es, el conjunto de mecanismos innatos que participan en la confección del concepto de número. Como se dará cuenta en este trabajo, se sabe que dicha percepción en niños tiene una relación directa con el desempeño que lograrán en la educación matemática formal. De ahí la importancia que podría llegar a tener conocer en profundidad los mecanismos asociados al Sentido Numérico.

En el primer capítulo de este trabajo, se hace un recorrido por los principales aspectos de la Cognición Matemática y el Sentido Numérico, analizando el porqué de su importancia, revisando los principales resultados experimentales tanto en estudios conductuales como neurobilógicos, y estudiando algunos modelos que intentan describir estos fenómenos.

El capítulo 2 trata del problema de Representación de Estímulos y cómo medir su grado de similaridad. Junto con mostrar resultados clásicos de Psicofísica, se analizan en detalle los dos modelos más conocidos de representación numérica; el modelo de escala comprimida y el linal con variabilidad escalar. Para esto, se estudia el problema de discriminación de cantidades bajo dos enfoques. El primero, se sitúa en la condición que frente a dos estímulos numéricos no simbólicos, se pide al sujeto experimental identificar el mayor de los dos. El

otro enfoque consiste en determinar si los dos estímulos numéricos no simbólicos expuestos son iguales o no. Analizando las predicciones que hacen los dos modelos en las situaciones descritas se encuentran diferencias entre ellos. Además, se prueba que ambos modelos pertenecen a familias de modelos más generales que siguen cumpliendo las principales propiedades estudiadas.

En el tercer capítulo se trabaja el problema de estimación de cantidades, esto es, frente a un determinado estímulo numérico, por ejemplo, nubes de puntos, secuencias de tonos, etc. se pide estimar la cantidad asociada. Para ello se propone y estudia dos modelos para explicar dicho fenómeno. Se contrastan las predicciones hechas por los modelos con los resultados experimentales conocidos y se dejan algunas preguntas abiertas para futuras investigaciones.

Finalmente, en el Capítulo 4, se aborda el tema del Aprendizaje de Fracciones. Para ello, se realizó un estudio con más de 200 niñas y niños de cuarto básico, con el fin de comparar tres métodos de aprendizaje. Se comparó la representación Geométrica de fracción (por ejemplo, $1/2$ es equivalente a la mitad de un cuadrado), con una representación Temporal ($1/2$ equivale a recibir un dulce cada dos días) y con una representación de Intercambio ($1/2$ equivale a recibir un objeto por cada dos monedas que doy). La primera es una de las más utilizadas en textos escolares y en el aula, mientras que las otras son poco conocidas o se suelen presentar en otros contextos.

Se encontraron diferencias en desempeño a favor de las representaciones Temporal y de Intercambio en la tarea de comparación de fracciones. Junto con esto, aprovechando la información recopilada, se propone tres métodos para representar estímulos en una escala numérica y se construyó un modelo que permite estimar el impacto de cada método de enseñanza en el cambio de estrategias que los estudiantes utilizan para comparar fracciones.

Capítulo 1

Del Sentido Numérico

“También contiene la memoria las razones y leyes infinitas de los números y dimensiones, ninguna de las cuales ha sido impresa en ella por los sentidos del cuerpo, por no ser coloradas, ni tener sonido ni olor, ni haber sido gustadas ni tocadas.”

Agustín de Hipona

Confesiones

“Un hombre sabe que dieciocho y diecinueve suman treinta y siete con la misma seguridad que sabe que uno y dos suman tres; sin embargo, un niño no sabe esto tan pronto como aquél, no por falta de uso de razón, sino porque las ideas que expresan las palabras dieciocho, diecinueve y treinta y siete no las alcanza tan pronto como las que significan uno, dos y tres.”

John Locke

Ensayo sobre el Entendimiento Humano

Parece tener un consenso enorme considerar que uno de los conceptos fundamentales de las matemáticas es el de número. El rol central que ocupa en la vida cotidiana del hombre y en el desarrollo de ideas matemáticas más profundas son aspectos más que suficientes para hacerlo materia de este estudio. Sin embargo, este no es un trabajo que aborda al número desde su axiomatización o su generalización; así como lo hizo Dedekind, Peano o Galois. Más bien, este tema será tratado desde el punto de vista de la Cognición Matemática.

La cognición matemática se basa en el estudio de los conceptos matemáticos haciendo uso de las herramientas de las ciencias cognitivas. Su objetivo central es entender los mecanismos que le permite al cerebro humano crear el universo de las matemáticas. El objeto de estudio será en este caso el “Sentido Numérico”; esto es, el conjunto de facultades cognitivas que permiten comparar, estimar e identificar magnitudes numéricas [11] [6]. Todas las personas, incluso en sus primeros años de vida, poseen un sentido numérico, es decir, una buena intuición de numerosidad.

En este capítulo se encontrarán los principales hallazgos relacionados con el sentido numérico. Se revisará el vínculo que posee con las matemáticas superiores, su presencia en animales de los más diversos tipos, así como también en niños y bebés humanos, además de los resultados sistemáticos observados en experimentos conductuales y las investigaciones neurobiológicas recientes que asocian neuronas específicas a los números.

1.1. Alcances del sentido numérico, ¿Por qué es un tema que debe ser tratado?

Dado lo crucial que son las habilidades numéricas para un buen desarrollo de la vida en la sociedades occidental, el comprender por qué se producen las diferencias en desempeño en matemáticas es un tema que actualmente tiene relevancia en la investigación educativa. Inicialmente, estas investigaciones intentaron relacionar las diferencias individuales en el desempeño académico con factores como la memoria de trabajo y la velocidad de procesamiento de la información por parte de los alumnos [1] [5] [23] . Sin embargo, otro enfoque teórico propone que el sentido numérico juega un rol predominante para el desarrollo del aprendizaje de las matemáticas formales, es decir, la habilidad de comprender y representar magnitudes numéricas condiciona el desarrollo de las habilidades matemáticas superiores, como por ejemplo, la aritmética.

Son varios los argumentos a favor de la relación entre el sentido numérico y el aprendizaje de las matemáticas. El primero establece que una buena comprensión de las magnitudes numéricas reduce el rango de respuestas posibles en problemas de aritmética, logrando así una mejora en el desempeño [2]. Otra línea argumentativa está ligada a estudios con individuos que presentan discapacidades en matemáticas; éstos tienen un déficit particular con la comprensión y procesamiento de magnitudes numéricas [24].

En general, las tareas que usualmente se ocupan para estudiar el sentido numérico y que tratan de establecer una correlación con las habilidades matemáticas superiores son dos. Una es la que hace uso de la recta numérica, en donde a los niños se les pide ubicar un número específico en ella. La otra tarea es la de comparación de magnitudes, en la cual a los niños se les pide identificar cuál de dos cantidades es la mayor. Sorprendentemente, ambas tareas parecen relacionarse con el desempeño en el primer año de escolaridad [37] [50]. Por ejemplo, estudios indican que cambios en la recta numérica estimada se relacionan con las diferencias individuales en el desempeño en matemáticas y la capacidad que tienen los niños a enfrentarse a problemas aritméticos nuevos.

Pese a los hallazgos mencionados, una pregunta que surge es si este sentido numérico y las habilidades matemáticas son manifestaciones de un mismo factor, el cual se podría denominar informalmente “ser bueno para las matemáticas”. Esto en vista que la mayoría de los estudios establecen correlaciones entre estos dos aspectos tomando mediciones en períodos de tiempo muy cercanos. De este modo, a priori no es clara la capacidad predictiva que podría tener un buen desarrollo del sentido numérico sobre el logro académico en matemáticas.

Sin embargo, estudios recientes que consideran un diseño longitudinal, concluyen que el

sentido número predice, en el sentido estricto del término, el desempeño futuro en matemáticas [57]. Contrastando los resultados de un test de comparación de magnitudes a un grupo de niños antes del inicio de la educación primaria, con el desempeño en matemáticas del mismo grupo un año más tarde, se concluye que un desarrollado sentido numérico anterior a la educación formal predice el nivel de aprendizaje que podría adquirir un alumno.

1.2. Sentido Numérico en:

1.2.1. Animales

El hecho que los animales sean capaces de contar es un asunto debatido entre los científicos. En cualquier caso, una gran cantidad de investigaciones concluyen que los animales tienen cierta noción de numerosidad. Estudios indican que ratas, primates, aves, peces e incluso hormigas tienen un sentido numérico. Por ejemplo, se tiene a la hormiga del desierto del Sáhara, la cual recorre enormes distancias de terreno arenoso en busca de alimento y vuelve sin problemas a su hormiguero. Un grupo de científicos alemanes y suizos descubrió que las hormigas “cuentan” sus pasos [63] [64]; manipulando la longitud de sus patas notaron que cuando éstas eran alargadas con una especie de zancos caminaban mayores distancias, mientras que si eran acortadas por medio de una amputación, recorrían una distancia más corta.

Otros estudios revelan que los animales no tan sólo serían capaces de comparar magnitudes numéricas, sino también realizarían operaciones aritméticas sencillas. Por ejemplo, estudios con palomas indicarían que son capaces de realizar restas [3]; las palomas ven dos series de flashes de luz y deben responder picoteando un botón tantas veces como la resta los flashes mostrados, si la paloma acierta, recibe una premio en alimento. Los investigadores concluyen que las palomas serían capaces de realizar correctamente la operación incluso en situaciones en las cuales no fueron entrenadas.

Un buen marco teórico para explicar la existencia de este mecanismo es la teoría de la evolución; tener esta habilidad cognitiva daría ventajas comparativas contra los sujetos de la misma especie que no la posee. Así, distinguir dónde hay mayor cantidad de alimento, qué manada es más grande, a cuál cardumen conviene unirse o cuánta distancia ha recorrido el animal son preguntas que se resuelven de manera eficiente teniendo un mecanismo asociado a la numerosidad.

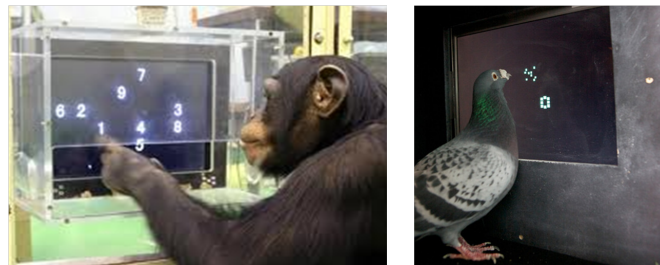


Figura 1.1: Animales en Estudios de Cognición Numérica.

Obviamente, en ningún instante se ha dicho que los animales entienden los números de igual manera que los seres humanos. La noción de sentido numérico no se relaciona a priori con la forma simbólica de éstos. Si bien, hay estudios en donde se muestra que primates pueden aprender la asociación entre un conjunto de puntos y el símbolo que representa su cantidad, o incluso que pueden aprender a ordenar los números expresados de manera simbólica de forma muy precisa, el sentido numérico en animales está asociado principalmente a características no simbólicas de los números.

Con todo esto sobre la mesa, una de las primeras conclusiones que aparece es el hecho que el sentido numérico es en principio independiente del lenguaje, teniendo sus orígenes en aspectos biológicos y adaptativos.

1.2.2. Niños y bebés

Hace más de medio siglo, las teorías de cómo los niños adquieren sus primeras habilidades matemáticas estaban dirigidas principalmente por las ideas de Piaget. Él planteaba que las habilidades lógicas y matemáticas se construían de manera progresiva en la mente de los niños mediante observación, internalización y abstracción de las regularidades del mundo que les rodea. El principal supuesto establecía que los infantes no tienen ningún tipo de conocimiento al nacer.

Respecto a los números, Piaget consideraba que éstos, del mismo modo que cualquier tipo de representación abstracta, debían su origen a interacciones con el medio ambiente haciendo uso de los cinco sentidos; manipulando colecciones de objetos los niños eventualmente descubren que existe una propiedad que permanece invariante cuando los objetos se mueven [48]. Se pensaba entonces, que los niños nacían sin ninguna idea preconcebida sobre los números.

Esta teoría se sustentó en diversos estudios experimentales. Una de las principales líneas argumentativas planteaba que los niños de pocos meses de edad no tenían asociado a los objetos la propiedad de permanencia, es decir, los objetos existen en la medida que el niño los percibe; por ejemplo, si un niño observa que un juguete es escondido detrás de una tela, él no va en busca del objeto. Así, si se instaura la idea de que los niños no perciben que los objetos permanecen en el tiempo, es poco plausible considerar que algún tipo de conocimiento sobre números y aritmética existe de manera innata.

Otro tipo de experimentos se basan en estudiar lo que se denominó como la conservación numérica. A niños de entre dos y cuatro años se les muestra dos filas de objetos, ambas filas contienen la misma cantidad, haciendo corresponder un objeto de la primera fila con otro de la segunda, así se tiene que ambas filas, además de tener la misma cantidad de objetos tienen el mismo largo. Una vez que los niños observan esto, se les pregunta cuál de las dos filas tienen más cantidad de objetos. Es sistemático que responden que ambas tienen la misma cantidad de elementos. Luego de esto, se selecciona una fila y se agrupan sus objetos, manteniendo la cantidad pero reduciendo el largo. Nuevamente se les pregunta a los niños cuál fila tiene más objetos, obteniendo como respuesta sistemática que la fila más larga tiene más elementos. Además si se agregan elementos a la fila más corta, manteniendo su longitud, los niños siguen declarando que la fila más larga tiene más objetos.



Figura 1.2: Niña realizando el Test de Conservación Numérica

Este hecho se ocupó para instaurar la idea que niños menores a cuatro años no son capaces de comprender la noción de número. De esta manera, la enseñanza de la matemática quedó en general ajustada a un programa de acuerdo a la edad de los niños, y en vista que ellos, en teoría, no están preparados para entender a cabalidad lo que es un número, parece primordial que en los primeros años se debe incentivar el pensamiento lógico y las nociones de orden.

Sin embargo, Mehler y Bever mostraron que el resultado del test de conservación de numerosidad dependía fuertemente del contexto y la motivación de los niños [41] [42]. En particular, mostraron que si el experimento se realizaba con dulces, sistemáticamente los niños escogían la fila de mayor cardinalidad, independiente la longitud de ésta. Una de las posibles explicaciones a este fenómeno es bastante simple; los niños no tienen claridad de qué se les está preguntando. Dado que las respuestas podrían ser evidentes, el cambio aparente de escenario podría hacer pensar que se busca una respuesta alternativa, generando un conflicto entre lo que se considera correcto con la supuesta lógica del cambio de escenario; si el respuesta es la misma, el cambio de escenario no tiene sentido, por lo que se debe contestar de manera diferente.

En relación a niños aún más pequeños, distintas investigaciones han llegado a la conclusión que bebés de meses de edad serían capaces de distinguir grupos de dos objetos con grupos de tres [58] [59]. Dado que no se les puede preguntar directamente a los niños, diversos métodos, por lo demás bastante ingeniosos, son ocupados para inferir estos hechos. La principal medida utilizada es el tiempo de observación al estímulo que destina el bebé. Por ejemplo, en algunos experimentos se le muestra al bebé una serie grupos de objetos todos de tamaño dos. Una vez habituado, se le muestra grupos de tres objetos, observando que cuando se produce el cambio de cantidad los niños están más tiempo mirando el estímulo.

Además de lo anterior, otras investigaciones indicarían que bebés, incluyendo recién nacidos, serían capaces de distinguir que $1 + 1$ es igual 2 , o incluso que $2 - 1$ es igual a 1 [67]. Sin embargo, estas capacidades aritméticas serían bastante limitadas, ya que cuando se testean igualdades que involucran números más grandes, los resultados no permiten hacer ningún tipo de inferencia.

Los hallazgos descritos hacen pensar que las ideas Piagetianas no estarían del todo correctas. Sin embargo, sería una falacia afirmar que los niños no desarrollan un razonamiento más profundo sobre el concepto de número en la medida que crecen y avanzan en el sistema

educativo, o que el éxito en las matemáticas puede ser predicho del todo en los primeros meses de vida. Lo cierto es que la evidencia expuesta debería dar un buen punto de partida para la enseñanza; saber que existe un mecanismo incipiente de numerosidad, conociendo sus bondades y deficiencias, podría facilitar de sobre manera la tarea educativa en los primeros años.

1.3. Resultados Clásicos de la Psicología Conductual

Sin duda, si se quisiera estudiar el sentido numérico en adultos, una de las principales ventajas es que los diseños experimentales no deben ser demasiado sofisticados como en los casos anteriores. Trabajar con adultos permite registrar gran cantidad de datos, teniendo respuestas concretas a preguntas específicas. Sin embargo, una nueva dificultad emerge. El nivel de educación matemática que han tenido los sujetos podría ser influyente en los resultados. Naturalmente, si se pregunta a persona mayor cuánto es $1 + 1$, difícilmente responderá algo distinto de 2. Si se quiere estudiar el sentido numérico en adultos, el diseño experimental debe ser tal que el cálculo numérico exacto no sea una estrategia factible.

Científicos aseguran que los adultos humanos poseen una representación numérica aproximada. Utilizando como principales medidas los porcentajes de correctitud y los tiempos de respuestas, se han estudiado en profundidad dos tipos de experimentos: el de comparación de magnitudes y el de estimación numérica.

En el primero, como su nombre lo indica, el sujeto debe señalar cuál de los dos estímulos presentados es el mayor. Los estímulos pueden ser mostrados en distintos formatos, ya sea como nubes de puntos, segmentos de distintos largos o secuencias de tonos. También existen variaciones de este experimento, en donde se presenta una serie de estímulos y se le pide al individuo indicar si lo observado es mayor o menor a una cantidad predefinida. El diseño en este tipo de experiencia debe estar formulado de manera que los sujetos no puedan hacer un conteo exacto de los estímulos, ya que se espera que al momento de responder los sujetos apelen a la representación aproximada que poseen de los números.

En estas y otras modalidades del experimento de comparación aparece un resultado sistemático: es más fácil discriminar números lejanos que cercanos. Así, discriminar entre 2 y 6 es más fácil que entre 5 y 6. A este fenómeno se le denomina Efecto Distancia. Además se observa lo que se conoce como Efecto de Tamaño. Esto indica que si se mantiene constante la distancia numérica, son más fáciles de discriminar cantidades pequeñas. Por ejemplo, distinguir entre 2 y 4 es más fácil que entre 8 y 10; en ambos casos la distancia entre los estímulos es 2. Ajustes estadísticos a resultados de distintos experimentos de este tipo indican que la probabilidad de discriminar correctamente dos estímulos numéricos depende principalmente del cociente entre ellos [7] [12]. Esto se conoce comúnmente como la Ley de Weber, y es concordante con los dos efectos mencionados.

Por otra parte, en la tarea de la estimación numérica, los sujetos se enfrentan a un determinado tipo de estímulo, comúnmente nubes de puntos, y se pide dar una estimación de la cantidad de objetos que se observan. Los estímulos son mostrados en períodos breves, del

orden de los milisegundos, de modo que no sea una estrategia contar. Como es de esperar, las estimaciones entregadas por las personas están lejos de ser exactas, mostrándose que en general, existe una tendencia a la subestimación [30] [34].

Estudios indican que la calidad de las estimaciones estaría afectada por distintos factores, como la cantidad de objetos por unidad de superficie de la imagen; a mayor densidad, mayor la tendencia a subestimar. También señalan que los estímulos asociados a cantidades pequeñas son sobrestimados y de cantidades grandes subestimados, y que si los estímulos están equiespaciados, las cantidades tienden a sobrestimarse.

Al momento de establecer una relación funcional entre estímulo y respuesta del sujeto, los ajustes estadísticos indicarían que en promedio es una relación potencial, es decir

$$R = k \cdot E^\alpha$$

con $k > 0$ y $\alpha \in (0, 1)$, donde E es el estímulo presentado y R la estimación de numerosidad de dicho estímulo [34] [35].

Otro resultado que suele aparecer tiene que ver con el nivel de variabilidad de las respuestas entregadas por los individuos. A mayor el número que se pide estimar, mayor la variabilidad observada. Algunos investigadores han llegado a establecer que si R_ε es la variable aleatoria que representa la respuesta que da un sujeto frente al estímulo ε , entonces la cantidad

$$\frac{\mathbb{E}(R_\varepsilon)}{\sqrt{\text{Var}(R_\varepsilon)}}$$

no depende de ε . A esto se le conoce como la Propiedad de Variabilidad Escalar.

Junto con esto, se sabe que dando algún tipo de información a los sujetos experimentales, disminuye la variabilidad de las respuestas. Por ejemplo, Izard y Dehaene [31] indican que mostrando una nube de puntos al inicio del experimento e indicando su cantidad, las respuestas sucesivas tienen a ser más certeras y de menor varianza. A este fenómeno se le denominó como Calibración de la Recta Numérica Mental y se ilustra en la figura 1.3.

1.4. Aspectos Neurobiológicos Involucrados

Si bien, es sabido que existe una variedad de experimentos conductuales con animales, niños y adultos humanos, los cuales permiten inferir la existencia de un sentido numérico, en los últimos años ha sido de gran impacto el descubrimiento de neuronas específicas que se sintonizan con la numerosidad de conjuntos.

Midiendo la actividad neuronal en monos cuando se les presentaba un estímulo visual, consistente en nubes de puntos, científicos encontraron neuronas específicas de numerosidad, es decir, se encontró que la actividad máxima de un determinado grupo de neuronas se asociaba a una numerosidad específica, hallando además que en la medida que el estímulo se iba alejando del número preferido ocurría una caída progresiva de la actividad neuronal del grupo asociado [45] [14].

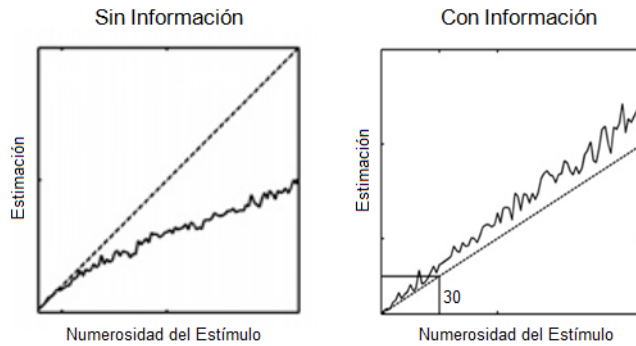


Figura 1.3: Promedio de Estimaciones en Función de los Estímulos. En ambos gráficos el eje horizontal indica la numerosidad real de un estímulo numérico, mientras que el eje vertical indica la estimación que hace el individuo. El gráfico de la izquierda muestra cómo son las estimaciones cuando los sujetos no reciben ningún tipo de información sobre las numerosidades de los estímulos. El gráfico de la derecha muestra el comportamiento de las estimaciones después que a los sujetos se le muestra una nube de 30 puntos y se les dice cuántos puntos hay. Fuente: [31]

Mientras los monos realizaban una tarea en donde se les pedía identificar si dos cantidades eran iguales o no, se identificaron dos zonas principales del cerebro que registraron activación inducida por los estímulos numéricos; la Corteza Prefrontal Lateral (PFC) y la Corteza Parietal Posterior (PPC). La primera también está asociada a las emociones y al control del comportamiento, mientras que la segunda zona se asocia además al área visual y al control sensoriomotor. En la figura 1.4 se observa la proporción de neuronas selectivas de numerosidad de acuerdo a la escala de colores.

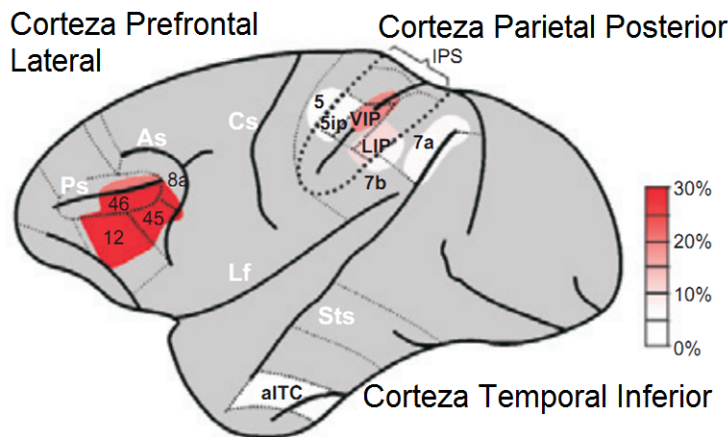


Figura 1.4: Vista Lateral del Cerebro de un Mono. Según experimentos con monos, la Corteza Prefrontal Lateral y la Corteza Parietal Posterior son las zonas del cerebro que registran mayor actividad neuronal al momento de realizar tareas de percepción numérica. Fuente: [14].

Científicos consideran que la Corteza Parietal Posterior podría ser el primer escenario neuronal donde se extrae la información numérica visual y somatosensorial, ya que las neuronas

correspondientes requieren de un tiempo necesario para transmitir información más corto que las neuronas de la Corteza Prefrontal. A su vez, sería en esta última zona en donde se amplificaría y almacenaría la información numérica.

Un descubrimiento sorprendente indica que existe un conjunto de neuronas que codifican numerosidad independiente del formato en que se presenta el estímulo. Así, si se presentan nubes de puntos en una pantalla, o una cantidad secuencial de puntos, habrían neuronas que su actividad depende exclusivamente de la cantidad expuesta. Esto daría pie para plantear que existiría una síntesis de información para luego formar una representación abstracta de cantidades.

Con todo esto sobre la mesa, investigadores consideran que el comportamiento observado en las células nerviosas selectivas de numerosidad podría explicar los fenómenos psicológicos básicos en monos, como los mencionados efectos de distancia numérica y de tamaño. El primero dependería de qué tan superpuestas estarían las funciones neuronales; numerosidades adyacentes tendrían un nivel de superposición muy alto, mientras que numerosidades lejanas apenas se solaparían. Algo similar se tiene para el efecto de tamaño, ya que a mayor numerosidad del estímulo, conservando la distancia entre numerosidades, mayor la superposición de las funciones neuronales. En la figura 1.5 se muestra el comportamiento de distintos tipos de neuronas selectivas frente a estímulos de distintas numerosidades.

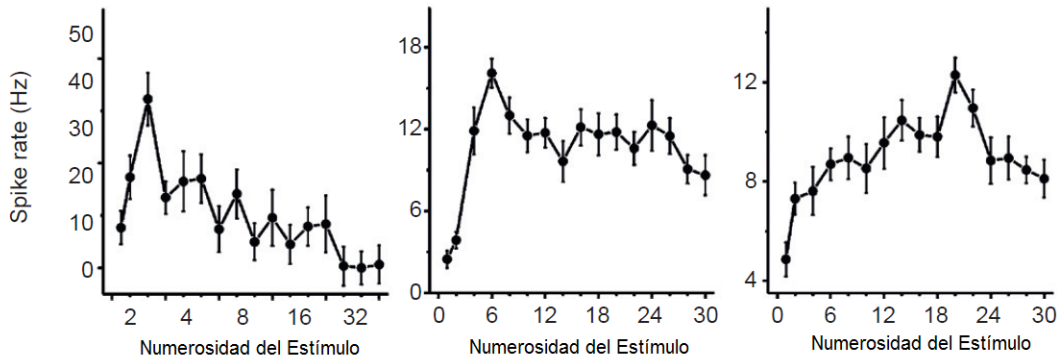


Figura 1.5: Actividad de distintas Neuronas Selectivas de Numerosidad. En los tres gráficos el eje horizontal indica la numerosidad del estímulo expuesto, y el eje vertical la actividad neuronal del grupo de neuronas. En el gráfico de la izquierda se observa que la numerosidad preferida es 3, en el gráfico del centro es 6 y en el de la derecha 22. Fuente: [14]

Por otra parte, además de las neuronas selectivas, se han hallado otro tipo de células que codifican numerosidad. Roitman y su equipo encontraron que la actividad de la mayoría de las neuronas del Lóbulo Intra Parietal (LIP) aumentó o disminuyó de forma sistemática al aumentar el número de elementos durante la presentación del estímulo, esto con independencia de las otras características de los estímulos o demandas cognitivas hacia el sujeto [51]. A este tipo de neuronas se les llamó neuronas monótonas de numerosidad. Los autores terminan sugiriendo que la magnitud del nivel de activación de las neuronas monótonas del LIP puede proporcionar información a las neuronas selectivas de numerosidad del PPC y PFC para calcular cantidades, logrando la sintonización con las numerosidades preferidas de acuerdo al grupo neuronal asociado.

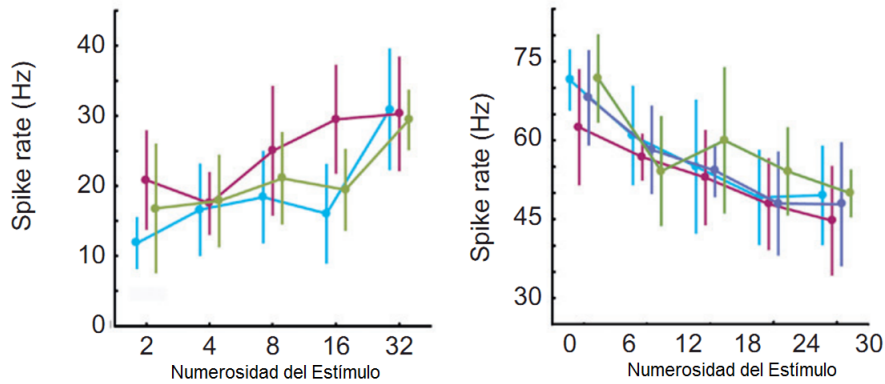


Figura 1.6: Actividad de las Neuronas monótonas en función de la numerosidad expuesta. En ambos gráficos el eje horizontal indica la numerosidad del estímulo expuesto, y el eje vertical la actividad neuronal del grupo de neuronas. El gráfico de la izquierda corresponde a un grupo de neuronas monótonas crecientes, mientras que el de la derecha corresponde a un grupo de neuronas monótonas decrecientes. Fuente: [14]

Otras investigaciones sugieren que la selectividad numérica no sería cosa exclusiva de estímulos discretos, como nubes de puntos, si no que además habrían neuronas selectivas de magnitudes continuas, como largos de segmentos, e que incluso existirían células nerviosas asociadas a proporciones [32]. En relación a estímulos simbólicos, estudios en monos después de un largo período de entrenamiento, concluyen que a pesar de que los monos utilizan el PFC y el PPC para las representaciones de cantidad no-simbólica, sólo la parte prefrontal se dedica a las asociaciones simbólicas de los números.

A esta altura cabría preguntarse si todos estos descubrimientos en monos son extrapolables a seres humanos. Si existen similitudes tanto anatómicas como funcionales entre los cerebros de primates no humanos y los seres humanos, entonces zonas equivalentes en el cerebro también deberían ser activadas en las personas cuando cantidades en formato no verbal son procesadas. Por medio de pruebas de resonancia magnética funcional (fMRI), Piazza y su equipo encontraron una esquema de activación neuronal similar en humanos, mientras los sujetos desarrollaban tareas numéricas no simbólicas [49] [14].

1.5. Algunos Modelos del Sentido Numérico

Explicar las leyes anteriormente expuestas a partir de un conjunto básico de reglas es un objetivo deseable. Comprender en detalle cómo funcionan los mecanismos asociados a numerosidad podría tener inmensas repercusiones no tan sólo para las ciencias cognitivas, sino también en áreas como las ciencias sociales y la educación formal. Una manera de sintetizar y analizar en profundidad estos fenómenos es utilizando el lenguaje matemático, lo que permitiría capturar las regularidades de los comportamientos mencionados. Son varios los modelos de numerosidad que se han propuesto en los últimos años. Aquí sólo se analizará algunos.

En relación a la cuantificación de información ejecutada por los sujetos, dos son los principales modelos que se han planteado. Uno es el modelo de modo-control de Meck y Church, el cual trabaja en serie, es decir, los componentes de la información numérica se va procesando secuencialmente; por ejemplo, los puntos de una nube van siendo procesados uno a uno. El modelo propone que cada elemento está codificado por un impulso generado por un contador, que se termina añadiendo a un acumulador. La magnitud en este acumulador al final del proceso de conteo se lee en la memoria, formando de este modo una representación de la cantidad de elementos de un conjunto. Así, el proceso de conteo sería a una suerte de suma, lo cual intentaría asociarse con el comportamiento de las neuronas monótonas de numerosidad.

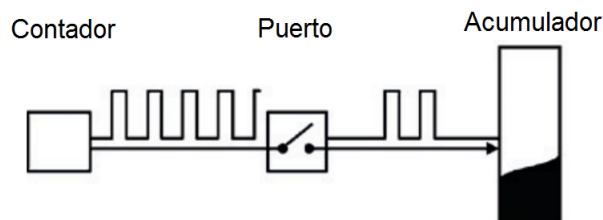


Figura 1.7: Esquema del Modelo Modo-Control de Meck y Church. Fuente: [6]

El otro modelo es el de filtrado neuronal de Dehaene y Changeux, el cual procesa la información numérica en paralelo. En una primera instancia, cada estímulo visual se codifica como una distribución gaussiana local de activación, simulando de este modo el comportamiento de la retina al momento de observar un conjunto de objetos. Luego, los elementos de diferentes tamaños son normalizados, teniendo así un código independiente del tamaño del estímulo. A su vez, las posiciones de las agrupaciones de activaciones locales detectadas también son codificadas. Con estos dos códigos se realiza una especie de suma de agrupaciones; esta operación tienen implementados sesgos que aumentan con números cada vez mayores. La suma de estos conglomerados finalmente se traduce en la numerosidad del conjunto.

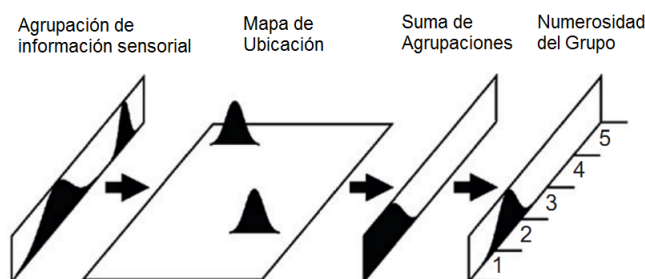


Figura 1.8: Esquema del Modelo de Filtro Neuronal de Dehaene y Changeux. Fuente: [6]

A pesar que las representaciones de numerosidad derivadas con ambos modelos son ruidosas, es decir, incorporan efectos aleatorios, y se ajustan de manera bastante razonable a los datos observados, los dos modelos se diferencian en aspectos importantes. El modelo de modo de control de Meck y Church opera en serie y asume la representación de la cardinalidad en una escala lineal, mientras que en el modelo de filtrado neuronal de numerosidad de Dehaene y Changeux se codifica de forma paralela y la representación termina ajustándose a una escala logarítmica [6].

El debate sobre cuál escala es la más adecuada para modelar la representación mental de los números está aún vigente. Autores como Dehaene proponen que si un sujeto observa un estímulo de numerosidad n , entonces una manera de modelar su representación mental es mediante una variable aleatoria que se distribuye como una normal de media $\log(n)$ y varianza ω^2 . A esto se le conoce como Modelo de Escala Comprimida. Por otra parte, autores como Gallistel señalan que un mejor modelo para la representación mental estaría dado por una variable aleatoria normal de media n y varianza $\sigma \cdot n^2$. A esto se le conoce como el Modelo Lineal con Variabilidad Escalar.

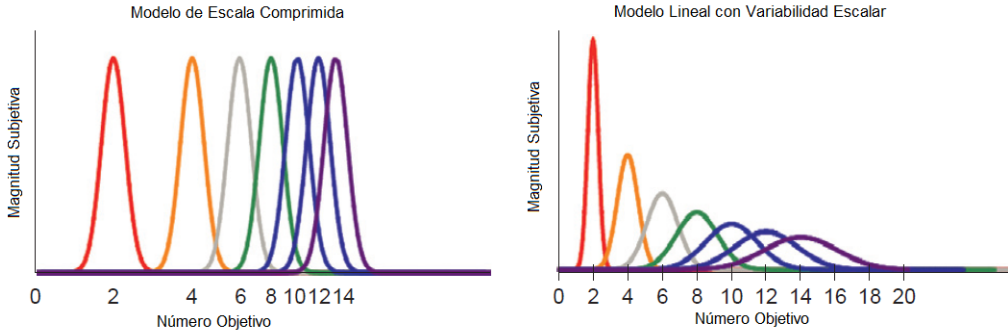


Figura 1.9: Modelos de Escala Comprimida y Lineal con Variabilidad Escalar. A la izquierda se puede observar una descripción gráfica del modelo de Escala Comprimida, en donde se aprecia el escalamiento logarítmico de las numerosidades con varianza constante. A la derecha se ilustra el modelo Lineal con Variabilidad Escalar, donde se observa el escalamiento lineal y el aumento lineal de la desviación estándar de las representaciones de numerosidades. Fuente: [14]

Estos modelos se ajustan bastante bien a los resultados observados en la tarea de comparación de magnitudes. En efecto, si se toma el modelo de Escala Comprimida, y se supone que al sujeto se le presentan dos estímulos, uno de tamaño n y el otro de tamaño m , entonces la declaración de que n es mayor a m está dada por la comparación de las variables aleatorias independientes K_n y K_m donde $K_n \sim N(\log(n), \omega^2)$ y $K_m \sim N(\log(m), \omega^2)$. Así, la probabilidad que el sujeto declare que n es más grande que m está dada por

$$\begin{aligned}
 \mathbb{P}(K_n > K_m) &= \mathbb{P}(0 > K_m - K_n) \\
 &= \mathbb{P}\left(\frac{-(\log(m) - \log(n))}{\omega\sqrt{2}} > \frac{K_m - K_n - (\log(m) - \log(n))}{\omega\sqrt{2}}\right) \\
 &= \Phi\left(\frac{-(\log(m) - \log(n))}{\omega\sqrt{2}}\right) \\
 &= \Phi\left(\frac{1}{\omega\sqrt{2}} \cdot \log\left(\frac{n}{m}\right)\right)
 \end{aligned}$$

donde Φ es la función de distribución de una normal de media 0 y varianza 1.

Por otra parte, si se toma el modelo lineal con variabilidad escalar, de manera análoga se tendrá que la declaración de que n es mayor a m está dada por la comparación de las

variables aleatorias independientes V_n y V_m donde $V_n \sim N(n, \sigma^2 n^2)$ y $V_m \sim N(m, \sigma^2 m^2)$. Con esto, la probabilidad que el sujeto declare que n es mayor que m , según este modelo, está dada por

$$\begin{aligned}
\mathbb{P}(V_n > V_m) &= \mathbb{P}(0 > V_m - V_n) \\
&= \mathbb{P}\left(-\frac{m-n}{\sigma\sqrt{m^2+n^2}} > \frac{V_m - V_n - (m-n)}{\sigma\sqrt{m^2+n^2}}\right) \\
&= \Phi\left(-\frac{m-n}{\sigma\sqrt{m^2+n^2}}\right) \\
&= \Phi\left(\frac{\frac{n}{m}-1}{\sigma\sqrt{1+\left(\frac{n}{m}\right)^2}}\right)
\end{aligned}$$

Como se observa, ambos modelos llegan a dependencias del cuociente entre los números que se comparan, lo cual sucede experimentalmente. Además, si bien las expresiones encontradas no son idénticas, en la práctica es imposible tener un buen ajuste de un modelo con datos experimentales, sin dejar de tenerlo con el otro [15].

Los defensores de la escala comprimida aseveran que este modelo es más plausible que el modelo lineal pues el primero consumiría mucho menos recursos neuronales que el segundo; tener una escala lineal asociada a los números sería poco eficiente, ya que esto se traduciría en carga cognitiva extra y a su vez la chance de encontrarse con números grande es más pequeña comparada con números pequeños. Junto con esto, aseguran que los últimos hallazgos relacionados con neuronas específicas y monótonas daría pie para estar a favor de la escala logarítmica.

Sin embargo, los defensores del modelo lineal argumentan que un modelo de escala comprimida no estaría en concordancia con las capacidades aritméticas de los sujetos. Una escala logarítmica generaría grandes dificultades para realizar operaciones como la suma o la resta. Sin embargo, se observa que incluso aves serían capaces de realizar estas operaciones con un margen de error menor.

La discusión anterior motiva a Gómez para proponer un modelo basado en el origen evolutivo de la escala numérica mental [25]. La escala mental de los números será una secuencia de puntos en el intervalo $[0, 1]$. Dado un valor $N \in \mathbb{N}$ mayor que 2, una escala queda definida por los puntos $x_1 \leq x_2 \leq \dots \leq x_N$ donde x_i sería la posición asociada a la cantidad i . Con esto, la representación de un conjunto de tamaño i está dado por una variable aleatoria ξ_i que se distribuye como una normal de media x_i y varianza fija σ^2 . Así, el error de declarar que i es mayor que j cuando en verdad $i < j$ está dado por la comparación de realizaciones independientes de las variable ξ_i con $i = 1, \dots, N$

$$\mathcal{E}(x, i, j) = \mathbb{P}(\xi_i > \xi_j) = \Phi\left(\frac{x_i - x_j}{\sigma\sqrt{2}}\right)$$

El supuesto del origen evolutivo de la escala queda expresado en la minimización de una

combinación lineal de los errores $\mathcal{E}(x, i, j)$ con $1 \leq i < j \leq N$. El modelo propuesto indica que la función objetivo sería

$$ET(w, x) = \sum_{i=1}^{N-1} w_i \mathcal{E}(x, i, i+1)$$

donde $w \in \mathbb{R}^{N-1}$ es un vector de pesos.

Gómez muestra que si los pesos son decrecientes, entonces existe un valor L entre 1 y N tal que $x_L = x_{L+1} = \dots = x_N$, mostrando de esta manera que para números grandes las representaciones podrían conglomerarse.

1.6. Primeras Conclusiones

A partir de lo expuesto en este primer capítulo, una de las primeras conclusiones es el hecho que el sentido numérico no está presente sólo en seres humanos, sino que es transversal a muchas especies, lo que hace inferir que es un mecanismo que funciona de manera independiente del lenguaje, de modo que sus orígenes podrían deberse a aspectos evolutivos y adaptativos. Además, la evidencia muestra que el sentido numérico tiene componentes innatos, los cuales ya pueden ser observados en los primeros meses de vida.

Otro punto importante es la relación entre el sentido numérico con el aprendizaje de matemáticas superiores. Muchas son las investigaciones que ligan desempeños en tareas que apelan al uso de mecanismos de percepción numérica no simbólica con desempeños en los primeros años de escolaridad. También hay estudios que señalan que el desempeño en tareas de percepción numérica no simbólica se puede entrenar y mejorar, y que a su vez, esto influye en los posteriores resultados académicos en primaria.

Con lo señalado, la repercusión de los hallazgos expuestos en este capítulo en la Educación Matemática resulta ser natural. Conocer que existe un mecanismo incipiente de numerosidad, con bondades y deficiencias, podría facilitar la tarea educativa. Al introducir conceptos matemáticos nuevos en los primeros años de escolaridad, se debería apelar fuertemente al uso del sentido numérico.

El siguiente capítulo aborda desde un punto de vista matemático el problema de la comparación de magnitudes. Para ello se hará un recorrido rápido sobre algunos hechos conocidos de Psicofísica, revizando la literatura reciente y enfocando la teoría hacia aspectos de la cognición matemática. Además se hará una comparación entre los modelos de Escala Comprimida y el Lineal con Variabilidad Escalar.

Capítulo 2

Representaciones y Medidas de Similaridad

“El cerebro puede tener muchos pasos intermedios en los circuitos que median entre el estímulo y la respuesta, y seguir careciendo de mente, si no cumple una condición esencial: la capacidad de representar internamente imágenes y de ordenar dichas imágenes en un proceso denominado pensamiento.”

Antonio Damasio

El Error de Descartes

Como se mencionó en el capítulo anterior, la representación de los números en alguna escala mental es un tema aún debatido en el mundo científico. La elección entre el modelo de escala comprimida o el modelo lineal no pasará por más estudios experimentales, sino por tratar de establecerse diferencias específicas en alguna predicción que se haga con los modelos.

Para tratar de llevar esto a cabo, se estudiará en primer lugar el problema de comparación de magnitudes desde un punto de vista teórico, definiendo qué se entiende por escalamiento y determinando en qué condiciones se puede tener. Además, se analizará qué tipo de representaciones podrían ser consideradas para modelar los fenómenos cognitivos descritos. Posteriormente, se estudiarán los modelos en base a la tarea de declaración de diferencia; esto consiste en que frente a dos estímulos al sujeto se le pide declarar si éstos son iguales o distintos. También será materia de análisis cómo construir una métrica adecuada entre estímulos, y a partir de ello, construir su representación.

Para el estudio de la representación de estímulos, este trabajo se basará fuertemente en el texto de Falmagne, “Elementos de la Teoría Psicofísica” [20], mientras que para la construcción de métricas entre los estímulos se usará como referencia los trabajos de Dzhafarov y Colonius [16] [17] [18]. Finalmente, la representación de estímulos a partir de una métrica será tratado con el método de Escalamiento Multidimensional Estándar [36] [38].

2.1. Representación de Estímulos

En esta sección se analizará la situación experimental en donde a los sujetos se les pide identificar cuál es el mayor de dos estímulos observados. Esta identificación depende del contexto en que son presentados los estímulos. Por ejemplo, frente a dos señales sonoras se podría pedir seleccionar la más aguda, o bien, al observar dos estímulos visuales se podría pedir señalar el de mayor intensidad luminosa. De este modo, el tratamiento matemático aquí expuesto no es propio del problema de discriminación numérica, siendo su carácter general a cualquier problema de discriminación de estímulos.

En esta primera fase se trabajará con estímulos que se identifican con una variable cuantitativa en un dominio continuo. Por ejemplo, un sonido con su frecuencia, una señal luminosa con su intensidad o un segmento con su longitud. Con esto en mente, se denotará con letras minúscula las magnitudes de los estímulos que se comparan.

El objeto principal de este estudio son los valores $P_{a,b}$, los que denotan la probabilidad de que un sujeto declara que a es mayor que b . Lo que en esta sección se analizará es el problema de representación, o de Fechner, el cual consiste en determinar bajo qué condiciones se puede obtener

$$P_{a,b} = F(u(a) - u(b))$$

donde F y u son funciones crecientes. A la función u se le conoce como la función de representación de estímulos, o también como escala psicométrica. La igualdad anterior indica que la probabilidad de declarar un estímulo mayor que otro depende de la diferencia de magnitudes en dicha escala.

Una de las primeras definiciones que surge en este contexto es la de sistema de discriminación.

Definición 2.1 Sea $I \subset \mathbb{R}$ un intervalo, y $C \subset I \times I$. Sea P una función de C en $(0, 1)$. Se dirá que (I, C, P) es un sistema de discriminación si $\forall a, b \in I$:

- 1) $(a, a) \in C$
- 2) $(a, b) \in C$ implica $(b, a) \in C$
- 3) P es creciente en la primera componente, decreciente en la segunda y continua.
- 4) Los conjuntos $\{c \in I / (c, b) \in C\}$ y $\{c \in I / (b, c) \in C\}$ son intervalos abiertos.

En caso que $(a, b) \in C$, se dirá que a es comparable con b . En general, ya sea por razones experimentales o limitaciones físicas, no siempre se tiene que todos los punto $a, b \in I$ son comparables. Sin embargo, se tiene el siguiente teorema

Teorema 2.2 Sea (I, C, P) un sistema de discriminación. Se tiene que para todo $a, b \in I$ existe un natural n y una secuencia a_1, a_2, \dots, a_n tal que $a_1 = a$, $a_n = b$ y $(a_i, a_{i+1}) \in C$ para todo i , $1 \leq i \leq n - 1$.

Las siguientes definiciones serán de gran ayuda en lo que sigue.

Definición 2.3 *Un sistema de discriminación (I, C, P) se dice balanceado si*

$$P_{a,b} + P_{b,a} = 1$$

Definición 2.4 *Un sistema de discriminación (I, C, P) es llamado Fechneriano si existen u y F funciones estrictamente crecientes continuas, tales que*

$$P_{a,b} = F[u(a) - u(b)]$$

Al par (u, F) se le llamará representación Fechneriana de (I, C, P) .

A nivel de observación, se puede decir que la condición de balance no es una condición suficiente para tener representación Fechneriana. En efecto, si se toma

$$P_{a,b} = \Phi \left[\frac{a-b}{\sqrt{a+b}} \right]$$

se verifica fácilmente que el sistema es balanceado. Por otra parte, si se asume que el sistema es Fechneriano con F y u son diferenciables, entonces por la condición de crecimiento tienen derivada positiva. Luego, al derivar con respecto a a queda

$$\Phi' \left(\frac{a-b}{\sqrt{a+b}} \right) \cdot \frac{\sqrt{a+b} - (a-b) \cdot \frac{1}{2} \frac{1}{\sqrt{a+b}}}{a+b} = F'(u(a) - u(b)) \cdot u'(a)$$

y al derivar con respecto a b queda

$$\Phi' \left(\frac{a-b}{\sqrt{a+b}} \right) \cdot \frac{-\sqrt{a+b} - (a-b) \cdot \frac{1}{2} \frac{1}{\sqrt{a+b}}}{a+b} = -F'(u(a) - u(b)) \cdot u'(b)$$

tomando el cociente entre estas dos expresiones y simplificando queda

$$\frac{a+3b}{3a+b} = \frac{u'(a)}{u'(b)}$$

Asumiendo que $b \neq 0$ y que $a = kb$ con $k \geq 0$, se tiene que

$$u'(kb) = \frac{u'(b) \cdot (k+3)}{3k+1}$$

tomando $k = 0$, se tiene que u' es constante, y de la última ecuación se tiene que

$$u'(b) = \frac{u'(1) \cdot (b+3)}{3b+1}$$

luego, $u'(x) = 0$, pero esto contradice el hecho que u era estrictamente creciente. De este modo, el sistema descrito no puede tener representación Fechneriana diferenciable.

En vista de lo anterior, se debe caracterizar aún más a los sistemas de discriminación para tener representación Fechneriana. En ese sentido, las siguientes definiciones son muy importantes.

Definición 2.5 Se dirá que (I, C, P) satisface la condición de bicancelación ssi

$$P_{a,b} \leq P_{a',b'}, P_{b,c} \leq P_{b',c'}, (a, c) \in C, (a', c') \in C \Rightarrow P_{a,c} \leq P_{a',c'}$$

Se dirá que el sistema cumple con la bicancelación débil si la condición anterior se obtiene reemplazando las desigualdades por igualdades.

Definición 2.6 Un sistema (I, C, P) satisface la condición cuádruple ssi

$$P_{a,b} \leq P_{a',b'} \Leftrightarrow P_{a,a'} \leq P_{b,b'}$$

Se llama condición cuádruple débil a la condición anterior con igualdad.

Es claro que las condiciones “fuertes” implican las débiles, pues basta aplicar las propiedades de las desigualdades en ambos sentidos. La recíproca también se tiene para sistemas balanceados y queda establecida por el siguiente teorema.

Teorema 2.7 En un sistema de discriminación balanceado:

- i) Bicancelación es equivalente a la bicancelación débil.
- ii) La condición cuádruple es equivalente a la condición cuádruple débil.

Se tiene que es necesario la bicancelación débil y la condición cuádruple débil para un sistema balanceado sea Fechneriano. Lo sorprendente es que se puede demostrar que estas condiciones son suficientes.

Teorema 2.8 Sea Ψ un sistema de discriminación balanceado. Entonces, son equivalentes:

- i) Ψ es Fechneriano
- ii) Ψ satisface la bicancelación débil.
- iii) Ψ satisface la condición cuádruple débil.

Además, si $(u, F), (u_0, F_0)$ son dos representaciones Fechnerianas de Ψ , entonces $u_0(a) = \alpha u(a) + \beta$ y $F_0(s) = F(s/\alpha)$ para algunas constantes $\alpha > 0$ y β .

Se debe notar que tanto el modelo de escala comprimida como el modelo lineal con variabilidad escalar tienen representaciones Fechnerianas asociadas. En efecto, para el primero se tiene que

$$\begin{aligned} P_{n,m} &= \Phi \left(\frac{1}{\omega\sqrt{2}} \cdot \log \left(\frac{n}{m} \right) \right) \\ &= F_1(u_1(n) - u_1(m)) \end{aligned}$$

donde $F_1(x) = \Phi \left(\frac{1}{\omega\sqrt{2}} \cdot x \right)$ y $u_1(x) = \log(x)$.

Para el modelo lineal con variabilidad escalar se tiene

$$\begin{aligned} P_{n,m} &= \Phi \left(\frac{\frac{n}{m} - 1}{\sigma \sqrt{1 + \left(\frac{n}{m}\right)^2}} \right) \\ &= F_2(u_2(n) - u_2(m)) \end{aligned}$$

donde $F_2(x) = \Phi \left(\frac{e^x - 1}{\sigma \sqrt{1 + (e^x)^2}} \right)$ y $u_2(x) = \log(x)$.

El hecho que ambos modelos se puedan explicar en términos de comparación de variables aleatorias, motiva la siguiente definición

Definición 2.9 Sea $\Psi = (I, C, P)$ un sistema de discriminación. Sea $\mathcal{U} = \{U_a \mid a \in I\}$ una colección de variables aleatorias conjuntamente distribuidas. Entonces \mathcal{U} es una representación aleatoria de Ψ si y sólo si cuando $(a, b) \in C$ y $a \neq b$, entonces

$$P_{a,b} = \mathbb{P}(U_a \geq U_b)$$

Si las variables aleatorias en \mathcal{U} son independientes, entonces decimos que \mathcal{U} es una representación aleatoria independiente de Ψ .

Una de las ventajas de los sistemas con representación aleatoria es que la propiedad de balanceo queda caracterizada de forma muy precisa. Esto lo establece el siguiente teorema

Teorema 2.10 Sea $\{U_a \mid a \in I\}$ una representación aleatoria de un sistema de discriminación $\Psi = (I, C, P)$. Entonces Ψ es balanceado si y sólo si

$$\mathbb{P}(U_a = U_b) = 0$$

cuando $(a, b) \in C$

Debe notarse que en un sistema de discriminación aleatoria balanceado se tiene

$$\begin{aligned} P_{a,c} &= \mathbb{P}(U_a > U_c) \\ &= \mathbb{P}(U_a > U_b > U_c) + \mathbb{P}(U_b > U_a > U_c) + \mathbb{P}(U_a > U_c > U_b) \\ &\leq \mathbb{P}(U_a > U_b > U_c) + \mathbb{P}(U_c > U_a > U_b) + \mathbb{P}(U_a > U_c > U_b) \\ &\quad + \mathbb{P}(U_b > U_a > U_c) + \mathbb{P}(U_b > U_c > U_a) + \mathbb{P}(U_a > U_b > U_c) \\ &= \mathbb{P}(U_a > U_b) + \mathbb{P}(U_b > U_c) \\ &= P_{a,b} + P_{b,c} \end{aligned}$$

Es decir, en un sistema de discriminación balanceado con representación aleatoria se tiene la desigualdad triangular. Esta propiedad es equivalente a varias otras.

Teorema 2.11 En un sistema balanceado de discriminación $\Psi = (I, C, P)$, las cuatro condiciones siguientes son equivalentes; siempre que las probabilidades estén definidas

$$i) P_{a,c} \leq P_{a,b} + P_{b,c} \quad \forall a, b, c \in I.$$

$$ii) 1 \leq P_{a,b} + P_{b,c} + P_{c,a} \quad \forall a, b, c \in I.$$

$$iii) P_{a,b} + P_{b,c} + P_{c,a} \leq 2 \quad \forall a, b, c \in I.$$

$$iv) P_{a_1, a_2} + P_{a_2, a_3} + \dots + P_{a_{n-1}, a_n} \leq n - 1 \quad \forall n \in \mathbb{N} \text{ y } \forall a_i \in I$$

En general, si un sistema tiene representación aleatoria, no necesariamente tiene representación Fechneriana. En efecto, considerese el sistema donde $U_a \sim N(a, a)$. Si se asume que las representaciones son independientes entonces

$$P_{a,b} = \Phi \left[\frac{a-b}{\sqrt{a+b}} \right]$$

y como se vió anteriormente, este sistema no tiene representación.

Un tipo de sistemas de discriminación que relaciona las representaciones Fechnerianas con las representaciones aleatorias son los sistemas de Bradley-Terry-Luce. Para comprenderlos en detalle, se hacen necesarias las siguientes definiciones

Definición 2.12 Sea $\Psi = (I, C, P)$ un sistema de discriminación. Una función v de I en \mathbb{R}_+ se dice representación B.T.L. de Ψ si cuando $(a, b) \in I$

$$P_{a,b} = \frac{v(a)}{v(a) + v(b)}$$

Definición 2.13 Se dice que Ψ satisface la regla del producto si para $(a, b), (b, c), (a, c) \in C$ se tiene

$$\frac{P_{a,b}}{P_{b,a}} \cdot \frac{P_{b,c}}{P_{c,b}} \cdot \frac{P_{c,a}}{P_{a,c}} = 1$$

Definición 2.14 Sea una colección de variables aleatorias $\mathcal{U} = \{U_a \mid a \in I\}$ que satisface

$$\mathbb{P}(U_a \leq t) = \begin{cases} e^{v(a) \cdot t} & \text{si } t \leq 0 \\ 1 & \text{si } t > 0 \end{cases}$$

Se dice que \mathcal{U} es una representación exponencial negativa (independiente) de Ψ si \mathcal{U} es una representación aleatoria (independiente) de Ψ . A la función v se le llama parametrización de \mathcal{U} .

La primera observación que se debe hacer es que si un sistema Ψ tiene representación B.T.L., entonces es Fechneriano. En efecto, si el sistema Ψ tiene representación B.T.L., entonces existe v tal que

$$P_{a,b} = \frac{v(a)}{v(a) + v(b)} = \frac{1}{1 + \frac{v(b)}{v(a)}} = \frac{1}{1 + \left(\frac{v(a)}{v(b)}\right)^{-1}}$$

Si se define $F(x) = \frac{1}{1+e^{-x}}$ y $u(x) = \log(v(x))$, entonces se tiene

$$P_{a,b} = F(u(a) - u(b))$$

Se tiene claramente que F es creciente. Para que u sea creciente es necesario que v lo sea. En tal caso, se tendrá representación Fechneriana.

Las tres definiciones anteriores se enlazan en el siguiente teorema

Teorema 2.15 *En un sistema de discriminación balanceado Ψ , las siguientes tres condiciones son equivalentes:*

- i) Ψ satisface la regla del producto.*
- ii) Ψ tiene una representación B.T.L.*
- iii) Ψ tiene una representación exponencial negativa \mathcal{U} .*

Además, la representación B.T.L de Ψ puede ser ocupada como parámetro de \mathcal{U} , y esta función debe ser estrictamente creciente y continua. Por último, si v y v_0 son dos representaciones B.T.L. de Ψ , entonces $v_0 = \gamma v$ con γ alguna constante positiva.

En el capítulo 4 se volverá a hacer mención de este tipo de modelos.

Los teoremas hasta aquí enunciados son resultados clásicos de Psicofísica y sus demostraciones se pueden encontrar en [20]. El marco teórico anterior permitirá contextualizar la discusión entre los modelos de Escala Comprimida y Lineal con Variabilidad Escalar mencionados en el capítulo 1.

2.1.1. Modelos de Representación Aleatoria para Numerosidades

Una de las razones por las que la discusión entre los modelos de escala comprimida y lineal con variabilidad escalar no ha llegado a buen puerto está fundada principalmente en el hecho que ambos hacen la misma predicción: La probabilidad de declarar a mayor que b depende del cociente entre estas magnitudes y esta probabilidad aumenta en la medida que el cociente lo hace. Ambos modelos son muy populares en el mundo de la psicología cognitiva y la discusión se ha centrado en discernir cuál de los dos es más apropiado para explicar los fenómenos relativos al sentido numérico. Desde el punto de vista matemático, ambos suponen que las representaciones aleatorias de numerosidad tienen distribución normal, lo que facilita los cálculos al momento de calcular probabilidades.

Como parte de este trabajo, se probó que estos modelos pertenecen a familias de representaciones aleatorias más generales, que cumplen que la probabilidad de declarar un estímulo mayor que otro depende del cociente de las cantidades que se comparan. De esta manera, asociar a las representaciones aleatorias distribuciones normales ya no es un supuesto necesario.

Además, se propone otro modelo de discriminación basado en la comparación de los tiempos en que se producen los peaks de activación de neuronas específicas de numerosidad. Este modelo también predice que la probabilidad de declarar un estímulo mayor que otro depende del cociente entre ellos.

Modelos Aditivos

La representación aleatoria del modelo de escala comprimida se puede escribir como

$$K_a = \log(a) + N$$

donde $N \sim N(0, \omega^2)$. Se probó que este modelo es generalizable para casos en donde N no necesariamente tiene que tener una distribución normal. Para ello, se introduce la siguiente definición.

Definición 2.16 Sea $\mathcal{U} = \{U_a \mid a > 0\}$ una familia de variables aleatorias. Se dirá que \mathcal{U} sigue un modelo aditivo si U_a es independiente de U_b , con $a \neq b$, y para todo $a > 0$, U_a tiene la misma distribución que la variable aleatoria $\log(a) + \xi$, donde ξ es una variable aleatoria cuya distribución no depende de a .

Hecha esta definición, en el desarrollo de este trabajo se probó la siguiente proposición

Proposición 2.17 Sea $\mathcal{U} = \{U_a \mid a > 0\}$ una familia de variables aleatorias que sigue un modelo aditivo. Si \mathcal{U} es representación aleatoria de (I, C, P) , entonces $P_{a,b}$ es función de $\frac{a}{b}$, es decir, la probabilidad de declarar un estímulo mayor que otro depende del cociente entre ellos. Además, se tiene que si \mathcal{U} sigue un modelo aditivo, entonces se tiene el efecto distancia, esto es, $P_{a,b} \leq P_{c,b}$ si $b < a < c$.

DEMOSTRACIÓN. Si \mathcal{U} sigue un modelo aditivo, entonces U_a y U_b tienen la misma distribución que las variables aleatorias $\log(a) + \xi$ y $\log(b) + \xi'$, respectivamente. Además se tiene que ξ y ξ' son variables aleatoria iid, luego.

$$\begin{aligned} P_{a,b} &= \mathbb{P}(U_a \geq U_b) \\ &= \mathbb{P}(\log(a) + \xi - [\log(b) + \xi'] \geq 0) \\ &= \mathbb{P}(\xi' - \xi \leq \log(a) - \log(b)) \\ &= G \left[\log \left(\frac{a}{b} \right) \right] \end{aligned}$$

donde G es la función de distribución de $\xi' - \xi$.

Además, dado que G es creciente, si $0 < b < a < c$, entonces $\log \left(\frac{a}{b} \right) < \log \left(\frac{c}{b} \right)$, por lo que $P_{a,b} \leq P_{c,b}$.

□

Modelos Multiplicativos

La representación aleatoria del modelo lineal con variabilidad escalar se puede expresar como

$$V_a = a \cdot W$$

donde $W \sim N(1, \sigma^2)$. De manera similar al caso anterior, el modelo es generalizable para casos donde W no necesariamente tiene distribución normal. Siguiendo el esquema anterior, se introduce la siguiente definición.

Definición 2.18 Sea $\mathcal{U} = \{U_a \mid a > 0\}$ una familia de variables aleatorias. Se dirá que \mathcal{U} sigue un modelo multiplicativo si U_a es independiente de U_b , con $a \neq b$, y para todo $a > 0$, U_a tiene la misma distribución que la variable aleatoria $a \cdot \xi$, donde ξ es una variable aleatoria continua cuya distribución no depende de a .

Usando esta definición, como parte de este trabajo se probó la siguiente proposición

Proposición 2.19 Sea $\mathcal{U} = \{U_a \mid a > 0\}$ una familia de variables aleatorias que sigue un modelo multiplicativo. Si \mathcal{U} es representación aleatoria de (I, C, P) , entonces $P_{a,b}$ es función de $\frac{a}{b}$.

DEMOSTRACIÓN. Si \mathcal{U} sigue un modelo multiplicativo, entonces U_a y U_b tienen la misma distribución que las variables aleatorias $a \cdot \xi$ y $b \cdot \xi'$, respectivamente. Además se tiene que ξ y ξ' son variables aleatoria iid, luego.

$$\begin{aligned} P_{a,b} &= \mathbb{P}(U_a > U_b) \\ &= \mathbb{P}(a \cdot \xi > b \cdot \xi') \\ &= \int_{-\infty}^{\infty} \mathbb{P}(a \cdot x > b \cdot \xi') f_{\xi}(x) dx \\ &= \int_{-\infty}^{\infty} F_{\xi}\left(\frac{a \cdot x}{b}\right) f_{\xi}(x) dx \end{aligned}$$

donde F_{ξ} y f_{ξ} son las funciones de distribución y densidad de ξ respectivamente. Del cálculo anterior se observa la dependencia del cociente entre a y b . \square

Para que se cumpla el efecto distancia se debe imponer una condición extra, esta queda enunciada en la siguiente proposición que se demostró

Proposición 2.20 Si $\mathcal{U} = \{U_a \mid a > 0\}$ es una familia de variables aleatorias que sigue un modelo multiplicativo, entonces una condición suficiente para que se tenga el efecto distancia es

$$f_{\xi}(-x) < f_{\xi}(x)$$

cuando $x > 0$.

Nótese que para el caso normal con media 1 y varianza σ^2 esta condición se cumple. La proposición anterior se probó de la siguiente manera.

DEMOSTRACIÓN. Si \mathcal{U} sigue un modelo multiplicativo, en vista de los cálculos hechos en la demostración anterior, se tiene las siguientes equivalencias.

$$\begin{aligned}
& P_{a,b} > P_{c,d} \\
\Leftrightarrow & \int_{-\infty}^{\infty} \left[F_{\xi} \left(\frac{a \cdot x}{b} \right) - F_{\xi} \left(\frac{c \cdot x}{d} \right) \right] f_{\xi}(x) dx > 0 \\
\Leftrightarrow & \int_0^{\infty} \left[F_{\xi} \left(\frac{a \cdot x}{b} \right) - F_{\xi} \left(\frac{c \cdot x}{d} \right) \right] f_{\xi}(x) dx + \int_{-\infty}^0 \left[F_{\xi} \left(\frac{a \cdot x}{b} \right) - F_{\xi} \left(\frac{c \cdot x}{d} \right) \right] f_{\xi}(x) dx > 0 \\
\Leftrightarrow & \int_0^{\infty} \left[F_{\xi} \left(\frac{a \cdot x}{b} \right) - F_{\xi} \left(\frac{c \cdot x}{d} \right) \right] f_{\xi}(x) dx + \int_0^{\infty} \left[F_{\xi} \left(\frac{-a \cdot x}{b} \right) - F_{\xi} \left(\frac{-c \cdot x}{d} \right) \right] f_{\xi}(-x) dx > 0 \\
\Leftrightarrow & \int_0^{\infty} \left[h \left(\frac{a}{b}, x \right) - h \left(\frac{c}{d}, x \right) \right] dx > 0
\end{aligned}$$

donde

$$h(y, x) = F_{\xi}(y \cdot x) \cdot f_{\xi}(x) + F_{\xi}(-y \cdot x) \cdot f_{\xi}(-x)$$

Así, con la condición impuesta sobre f_{ξ} y considerando $x > 0$, se tiene que $h(\cdot, x)$ es creciente. En efecto

$$\frac{\partial h(y, x)}{\partial y} = f_{\xi}(yx) \cdot x \cdot f_{\xi}(x) - f_{\xi}(-y \cdot x) \cdot x f_{\xi}(-x)$$

así, se tiene que

$$\frac{\partial h(y, x)}{\partial y} > 0$$

Luego, si $\frac{a}{b} > \frac{c}{d}$, entonces

$$h \left(\frac{a}{b}, x \right) > h \left(\frac{c}{d}, x \right)$$

Lo que implica que

$$P_{a,b} \geq P_{c,d}$$

□

Modelos Temporales

Otro tipo de modelo podría hacer supuestos sobre la actividad neuronal. Este modelo supone que neuronalmente hay un proceso de Poisson asociado al estímulo c , dígase $L_t(c)$, de media $\lambda(c) = k \cdot c$, el cuál modela los momentos en que ocurren los peaks de activación neuronal. Suponiendo que hay una variedad de procesos asociados a un mismo estímulo funcionando de forma independiente, los tiempos en que se producen los primeros peaks de activación se distribuyen como exponenciales independientes de media $\lambda(c)^{-1}$. Definiendo $S_{p,c}$

el promedio de los tiempos asociados, donde p es el tamaño de la muestra, entonces se puede plantear

$$P_{a,b} = \mathbb{P}(S_{p,a} \leq S_{p,b})$$

Es decir, se declara que a es mayor que b si los peaks de activación neuronal asociados al estímulo a se producen en promedio antes que los de b .

Considerando p grande, entonces se tiene

$$S_{p,c} \sim N\left(\frac{1}{\lambda(c)}, \frac{1}{\lambda(c)^2 p}\right)$$

luego, dado que $\lambda(c) = k \cdot c$, queda

$$P_{a,b} = \Phi\left\{\frac{p^{\frac{1}{2}} \left[\frac{a}{b} - 1\right]}{\left[\left(\frac{a}{b}\right)^2 + 1\right]^{\frac{1}{2}}}\right\}$$

Obteniendo nuevamente la dependencia del cociente. Nótese que la forma que tiene la ecuación anterior es idéntica a la que genera el modelo de variabilidad escalar.

Comparación entre Modelos Aditivos y Multiplicativos

Una diferencia específica que tienen los modelos aditivos con los multiplicativos tiene que ver con el comportamiento asintótico de las predicciones. Para el caso de los modelos aditivos, se tiene que cuando $a \rightarrow \infty$ entonces, cuando b está fijo se tiene

$$P_{a,b} \rightarrow 1$$

A su vez, cuando a está fijo y $b \rightarrow \infty$ entonces

$$P_{a,b} \rightarrow 0$$

Sin embargo, en general, esto no ocurre necesariamente en los modelos multiplicativos.

Proposición 2.21 *Sea \mathcal{U} es una familia de variables aleatorias que sigue un modelo multiplicativo, es decir, U_a tiene la misma distribución que $a \cdot \xi$ y U_a es independiente de U_b con $a \neq b$, y ξ variable aleatoria continua. Se tiene que cuando $a \rightarrow \infty$ y b está fijo*

$$P_{a,b} \rightarrow \mathbb{P}(\xi > 0)$$

Además, cuando $b \rightarrow \infty$ y a está fijo

$$P_{a,b} \rightarrow \mathbb{P}(\xi < 0)$$

DEMOSTRACIÓN. Para los modelos multiplicativos se tiene

$$\begin{aligned} P_{a,b} &= \mathbb{P}(a \cdot \xi > b \cdot \xi') \\ &= \mathbb{P}(a \cdot \xi > b \cdot \xi' \mid \xi > 0) \cdot \mathbb{P}(\xi > 0) + \mathbb{P}(a \cdot \xi > b \cdot \xi' \mid \xi < 0) \cdot \mathbb{P}(\xi < 0) \\ &= \tilde{F}_1\left(\frac{a}{b}\right) \cdot \mathbb{P}(\xi > 0) + \left(1 - \tilde{F}_2\left(\frac{a}{b}\right)\right) \cdot \mathbb{P}(\xi < 0) \end{aligned}$$

donde \tilde{F}_1 y \tilde{F}_2 son funciones de distribución. Luego, estando b fijo, y tomando $a \rightarrow \infty$ se tiene que

$$P_{a,b} \rightarrow \mathbb{P}(\xi > 0)$$

Además, se tiene

$$\begin{aligned} P_{a,b} &= \mathbb{P}(a \cdot \xi > b \cdot \xi' \mid \xi' > 0) \cdot \mathbb{P}(\xi' > 0) + \mathbb{P}(a \cdot \xi > b \cdot \xi' \mid \xi' < 0) \cdot \mathbb{P}(\xi' < 0) \\ &= \left(1 - \tilde{F}_3\left(\frac{b}{a}\right)\right) \cdot \mathbb{P}(\xi' > 0) + \tilde{F}_4\left(\frac{b}{a}\right) \cdot \mathbb{P}(\xi' < 0) \end{aligned}$$

donde \tilde{F}_3 y \tilde{F}_4 son funciones de distribución. Con esto se concluye que estando a fijo y $b \rightarrow \infty$ entonces

$$P_{a,b} \rightarrow \mathbb{P}(\xi < 0)$$

□

De esta manera, se tiene que el comportamiento asintótico de los modelos multiplicativos está acotado, lo cual podría considerarse como una deficiencia para modelos de este estilo.

Si bien, en el caso que $\mathbb{P}(\xi > 0) = 1$ el comportamiento asintótico en el modelo multiplicativo es el mismo que en el modelo aditivo, también se debe señalar que en tal caso ambos modelos son equivalentes, ya que si $U_a = a \cdot \xi$, y el soporte ξ está contenido en los reales positivos, entonces $V_a = \log(a) + \log(\xi)$ está bien definido y se tiene que $\forall a, b > 0$

$$\mathbb{P}(U_a > U_b) = \mathbb{P}(V_a > V_b)$$

Es decir, si $\mathbb{P}(\xi > 0) = 1$ entonces el modelo multiplicativo se puede reescribir como un modelo aditivo.

Nótese que todo modelo aditivo se puede reescribir como un modelo multiplicativo, ya que si $U_a = \log(a) + \xi$, definiendo $V_a = \exp(U_a)$, se tiene que $V_a = a \cdot \xi_2$, con $\xi_2 = \exp(\xi)$. Obviamente el soporte de ξ_2 está contenido en los reales positivos y se tiene $\mathbb{P}(U_a > U_b) = \mathbb{P}(V_a > V_b)$.

2.2. Construcción de Métricas Fechnerianas

Trabajar con la comparación de magnitudes pidiendo identificar el estímulo mayor requiere que exista una correspondencia natural entre los estímulos y alguna escala. Esto puede ser obvio para estímulos numéricos, luminosos o incluso sonoros. Sin embargo, existen tipos de estímulos donde esto no es claro, como por ejemplo en el problema de identificación de rostros o voces; a los sujetos se les podría preguntar si dos frases que escuchan fueron generadas por la misma persona. En este tipo de casos pedir identificar el mayor de los estímulos, en principio, carece de todo sentido. Así, como plantea Dzharfarov, el objeto de estudio en esta sección será la función ψ , donde $\psi(x, y)$ representa la probabilidad de declarar que x es distinto de y .

A diferencia de lo estudiado anteriormente, el dominio de la función ψ estará en correspondencia a un conjunto numerable arbitrario de estímulos, denotado por \mathcal{M} . Formalmente, ψ es un función de $\mathcal{M} \times \mathcal{M}$ en $[0, 1]$.

Una propiedad importante que debe cumplir la función ψ es la llamada Minimalidad Regular, esto es

$$\psi(x, x) < \begin{cases} \psi(x, y) \\ \psi(y, x) \end{cases}$$

para todo $x, y \in \mathcal{M}$ con $x \neq y$. En general, no se exigirá que $\psi(x, y) = \psi(y, x)$ ya que las componentes de los argumentos de ψ podrían tener consideraciones experimentales precisas. Por ejemplo, si a los sujetos se les pide comparar dos imágenes que observan en momentos distintos, entonces la primera componente podría estar asociada a la imagen que se observa por primera vez, mientras que la segunda componente estaría asociada con la segunda imagen.

Otra propiedad que no se exigirá es el hecho que $\psi(x, x) = k$, para todo $x \in \mathcal{M}$. Experimentalmente se podría encontrar estímulos más difíciles de diferenciar de si mismos que otros. Esta situación es extrapolable al caso de números; dos nubes de dos puntos cada una son más fáciles de declarar como iguales que dos nubes con 50 puntos cada una.

El objetivo de esta parte es encontrar una métrica razonable para el conjunto de estímulos a partir de la función ψ , que de alguna manera permita inferir que tan parecido son cognitivamente los estímulos estudiados. Para ello se hará uso de las cantidades positivas

$$\begin{aligned} \Psi^{(1)}(x, y) &= \psi(x, y) - \psi(x, x) \\ \Psi^{(2)}(x, y) &= \psi(y, x) - \psi(x, x) \end{aligned}$$

las que se llaman incrementos psicométricos de primer y segundo tipo, respectivamente.

Si $a, b \in \mathcal{M}$, se entenderá por cadena que va desde a hasta b a cualquier secuencia $x_1, x_2, \dots, x_k \in \mathcal{M}$ tal que $x_1 = a$ y $x_k = b$. El largo psicométrico del tipo i de una cadena, con $i \in \{1, 2\}$, está dado por

$$L^{(i)}[(x_1, x_2, \dots, x_k)] = \sum_{j=1}^{k-1} \Psi^{(i)}(x_j, x_{j+1})$$

Si $C_{a,b}$ denota el conjunto de cadenas que van desde a hasta b , se define

$$G_i(a, b) = \inf_{c \in C_{a,b}} L^{(i)}[c]$$

Nótese que la función G_i cumple con todas las propiedades de una métrica, salvo la simetría. Ahora bien, si $x = (x_1, x_2, \dots, x_k)$ es una cadena que va desde a hasta b , entonces se define $\overleftarrow{x} = (y_1, y_2, \dots, y_k)$ como la cadena que va desde b hasta a tal que $y_i = x_{k-i+1}$, es decir, es la cadena inversa.

Dada las definiciones anteriores, es fácil probar el siguiente teorema.

Teorema 2.22 *Dados $a, b \in \mathcal{M}$ y una cadena x que va de a en b , entonces se tiene*

$$L^{(1)}[x] - L^{(2)}[\overleftarrow{x}] = L^{(2)}[x] - L^{(1)}[\overleftarrow{x}] = \psi(b, b) - \psi(a, a)$$

Como corolario del teorema anterior se tiene que

Corolario 2.23 Sean $a, b \in \mathcal{M}$ y sean las cadenas $x, y \in C_{a,b}$, entonces

$$L^{(1)}[x] + L^{(1)}[\overleftarrow{y}] = L^{(2)}[\overleftarrow{x}] + L^{(2)}[y]$$

Tomando ínfimo sobre las cadenas en la igualdad anterior, queda el siguiente teorema

Teorema 2.24 Sean $a, b \in \mathcal{M}$, entonces

$$G_1(a, b) + G_1(b, a) = G_2(a, b) + G_2(b, a)$$

De este modo, si se define $G(a, b) = G_1(a, b) + G_2(a, b)$, se tiene que G es una métrica en \mathcal{M} . A esta métrica se le denominará Métrica Fechneriana. Las demostraciones de estas y otras proposiciones relacionadas se pueden ver en [16] [17] [18].

Trabajar con una métrica facilita la tarea de representar los estímulos como puntos en algún espacio; esto consiste en asociar a a y b vectores x_a y x_b respectivamente, de modo que $G(a, b) = \|x_a - x_b\|$. Existen distintos métodos para realizar esto. En este trabajo se analizará el más conocido, el método de escalamiento multidimensional. Antes de ello, se estudiará la factibilidad de construir una métrica de este estilo suponiendo que las funciones ψ son calculadas a partir de los modelos de escala comprimida y variabilidad escalar.

2.2.1. Aplicaciones a Modelos de Numerosidad

En adelante, se considerará que $\mathcal{M} = \mathbb{N}$. En primer lugar, se estudiará la construcción de la métrica fechneriana con el modelo de escala comprimida, encontrando varias propiedades interesantes. Posteriormente se probará la infactibilidad de construir la métrica fechneriana con el modelo de variabilidad escalar, dado que no se cumple la propiedad de minimalidad regular.

Modelo de Escala Comprimida

Como se ha dicho, el modelo de Escala Comprimida supone que al visualizar un estímulo de tamaño n se realiza una variable aleatoria

$$K_n \sim N(\log(n), \omega^2)$$

Así, declarar que un estímulo es mayor que otro, o bien que dos estímulos son distintos, se corresponde a comparaciones de realizaciones de estas variables aleatorias, las cuales se supondrán independientes. En base a esto, la probabilidad de declarar x distinto de y , según este modelo, estaría dada por

$$\psi_{EC}(x, y) = \mathbb{P}(|K_x - K_y| > \delta) = \Phi\left(\frac{-\delta - \log\left(\frac{x}{y}\right)}{\omega\sqrt{2}}\right) + \Phi\left(\frac{\log\left(\frac{x}{y}\right) - \delta}{\omega\sqrt{2}}\right)$$

Donde Φ es la función de distribución de una normal estándar y δ es un parámetro que determina la capacidad de discriminación en la tarea de determinar si dos estímulos son distintos o no. Así, dos estímulos numéricos se declararán distintos si las realizaciones de las variables aleatorias asociadas a los estímulos se encuentran a una distancia mayor que δ .

Una función que será útil para efectos del análisis es

$$\varphi(x) = \Phi\left(\frac{-\delta - x}{\omega\sqrt{2}}\right) + \Phi\left(\frac{x - \delta}{\omega\sqrt{2}}\right)$$

con lo que se tiene

$$\psi_{EC}(x, y) = \varphi\left(\log\left(\frac{x}{y}\right)\right)$$

Nótese que, si bien no son propiedades necesarias para construir la métrica fechneriana, se cumple

1. $\psi_{EC}(x, x) = cte$
2. $\psi_{EC}(x, y) = \psi_{EC}(y, x)$

Lo que sí es necesario que se cumpla es la propiedad de Minimalidad Regular. Referente a esto se tiene la siguiente proposición

Proposición 2.25 *El modelo de Escala Comprimida cumple con la Minimalidad Regular, es decir*

$$\psi_{EC}(x, x) < \begin{cases} \psi_{EC}(x, y) \\ \psi_{EC}(y, x) \end{cases}$$

para $x \neq y$

DEMOSTRACIÓN. Se mostrará que la función φ tiene su mínimo en 0. En efecto, derivando φ queda

$$\varphi'(x) = -\frac{1}{\omega\sqrt{2}} \cdot \Phi'\left(\frac{-\delta - x}{\omega\sqrt{2}}\right) + \frac{1}{\omega\sqrt{2}} \cdot \Phi'\left(\frac{x - \delta}{\omega\sqrt{2}}\right)$$

Se tiene que Φ' es la función de densidad de una normal de media 0 y varianza 1. Esta función, además de ser simétrica, es creciente en la parte negativa y decreciente en la parte positiva. Supóngase que $x > 0$, entonces, como $\delta > 0$ se tiene que $-\delta - x < x - \delta$. Dado que hay distintos comportamiento de acuerdo al signo del argumento de la función Φ' , hay que plantear dos casos.

Si $0 < x$ y $x < \delta$, entonces $-\delta - x < x - \delta < 0$, lo que implica que

$$\frac{1}{\omega\sqrt{2}} \cdot \Phi'\left(\frac{-\delta - x}{\omega\sqrt{2}}\right) < \frac{1}{\omega\sqrt{2}} \cdot \Phi'\left(\frac{x - \delta}{\omega\sqrt{2}}\right)$$

es decir $\varphi'(x) > 0$

Si $\delta < x$, entonces se tiene que $0 < x - \delta < x + \delta$, y dado el decrecimiento de Φ' en la parte positiva y su simetría se tiene

$$\frac{1}{\omega\sqrt{2}} \cdot \Phi' \left(\frac{-\delta - x}{\omega\sqrt{2}} \right) < \frac{1}{\omega\sqrt{2}} \cdot \Phi' \left(\frac{x - \delta}{\omega\sqrt{2}} \right)$$

luego se concluye que para valores positivos φ es creciente. En vista de la paridad de la función φ , se tiene que para valores negativos la función es decreciente. Así se tiene que el mínimo se alcanza en $x = 0$.

Ahora bien, dado que

$$\psi_{EC}(x, y) = \varphi \left(\log \left(\frac{x}{y} \right) \right)$$

se obtiene que a x fijo, el valor de y que minimiza ψ_{EC} sería $y = x$.

□

Con la proposición anterior, se tiene que los incrementos psicométricos están bien definidos. Es más, se tiene la siguiente teorema

Teorema 2.26 *En el modelo de Escala Comprimida, se tiene*

$$G(a, b) = 2 \cdot G_1(a, b)$$

DEMOSTRACIÓN. Sean $a, b \in \mathcal{M}$. En vista que $\psi_{EC}(a, b) = \psi_{EC}(b, a)$, se tiene

$$\Psi^{(1)}(a, b) = \psi_{EC}(a, b) - \psi_{EC}(a, a) = \psi_{EC}(b, a) - \psi_{EC}(a, a) = \Psi^{(2)}(a, b)$$

Además, dado que $\psi_{EC}(a, a) = \psi_{EC}(b, b)$, se tiene

$$\Psi^{(1)}(a, b) = \Psi^{(2)}(a, b) = \Psi^{(1)}(b, a) = \Psi^{(2)}(b, a)$$

Con esto se tiene que si $x \in C_{a,b}$, cadena de a en b , entonces

$$L^{(1)}[x] = L^{(1)}[\overleftarrow{x}] = L^{(2)}[x] = L^{(2)}[\overleftarrow{x}]$$

Finalmente

$$\begin{aligned} G_2(a, b) &= \inf_{y \in C_{a,b}} L^{(1)}[\overleftarrow{y}] \\ &= \inf_{y \in C_{a,b}} L^{(1)}[y] \\ &= G_1(a, b) \end{aligned}$$

luego

$$G(a, b) = 2 \cdot G_1(a, b)$$

□

Un elemento que es importante estudiar es el camino $x^* \in C_{a,b}$ tal que

$$G(a, b) = 2 \cdot L^{(1)}(x^*)$$

En principio, no es obvio que x^* exista, sin embargo, la siguiente proposición que se demostró ayuda a aclarar esto.

Proposición 2.27 Sean $m, n \in \mathbb{N}$, tales que $m < n$. Para toda cadena $c \in C_{m,n}$ existe $\tilde{c} \in C_{m,n}$ tal que

$$L^{(1)}(\tilde{c}) \leq L^{(1)}(c)$$

Tal que todos los elementos de \tilde{c} se encuentran entre m y n .

DEMOSTRACIÓN. Sea $c \in C_{m,n}$. Si c tiene todos sus elementos entre m y n , entonces no hay nada que probar. Supóngase entonces que c tiene al menos un elemento que no cumple con esta condición. En primer lugar, se supondrá que existe al menos un elemento en la cadena que es mayor a n . Así, si $c = (m, x_2, x_3, \dots, x_{k-1}, n)$, se define

$$i = \inf\{k : x_k > n\}$$

Esto permite construir la cadena $\hat{c} = (m, x_2, \dots, x_{i-1}, n)$. Naturalmente, todos los elementos de la cadena \hat{c} son menores a n . Nótese que el largo psicométrico de c es mayor que el de \hat{c} . En efecto,

$$L^{(1)}(c) = L^{(1)}(c_1) + L^{(1)}(c_2) + L^{(1)}(c_3)$$

donde

$$c_1 = (m, x_2, \dots, x_{i-1})$$

$$c_2 = (x_{i-1}, x_i)$$

$$c_3 = (x_i, x_{i+1}, \dots, n)$$

Y

$$L^{(1)}(\hat{c}) = L^{(1)}(c_1) + L^{(1)}(\hat{c}_2)$$

con

$$\hat{c}_2 = (x_{i-1}, n)$$

en vista que $x_{i-1} < n < x_i$, entonces $0 < \log\left(\frac{n}{x_{i-1}}\right) < \log\left(\frac{x_i}{x_{i-1}}\right)$, y dado que φ es creciente en la parte positiva se tiene que

$$\psi_{EC}(x_{i-1}, n) < \psi_{EC}(x_{i-1}, x_i)$$

lo que implica que $L^{(1)}(\hat{c}_2) < L^{(1)}(c_2)$, y se tiene $L^{(1)}(\hat{c}) < L^{(1)}(c)$.

Ahora bien, si \hat{c} no contiene ningún elemento menor a m , entonces esta es la cadena buscada. Nótese que el procedimiento es válido si se supone que existe algún elemento mayor a n . Sin embargo, se puede dar que todos los elementos sean menores a n pero que existan elementos menores a m . Así, en lo que sigue, se supondrá que la cadena c que une a m con n tiene elementos menores a n . En caso que exista un elemento mayor a n , ya se sabe cómo construir una cadena de largo psicométrico menor con elementos menores a n .

En base a lo señalado, si $c = (m, x_2, \dots, x_{k-1}, n)$, se define

$$j = \sup\{k : x_k < m\}$$

Con esto se construye la cadena $\tilde{c} = (m, x_{j+1}, \dots, n)$. Nótese que el largo psicométrico de c se puede escribir como

$$L^{(1)}(c) = L^{(1)}(c_4) + L^{(1)}(c_5) + L^{(1)}(c_6)$$

donde

$$c_4 = (m, x_2, \dots, x_j)$$

$$c_5 = (x_j, x_{j+1})$$

$$c_6 = (x_{j+1}, x_{j+1}, \dots, n)$$

Y

$$L^{(1)}(\tilde{c}) = L^{(1)}(\tilde{c}_1) + L^{(1)}(c_6)$$

con

$$\tilde{c}_1 = (m, x_{j+1})$$

en vista que $x_j < m < x_{j+1}$, entonces $\log\left(\frac{x_j}{x_{j+1}}\right) < \log\left(\frac{m}{x_{j+1}}\right) < 0$, y dado que φ es decreciente en la parte negativa se tiene que

$$\psi_{EC}(m, x_{j+1}) < \psi_{EC}(x_j, x_{j+1})$$

lo que implica que $L^{(1)}(\tilde{c}_1) < L^{(1)}(c_5)$, y se tiene $L^{(1)}(\tilde{c}) < L^{(1)}(c)$. De esta manera, a partir de una cadena c que une m con n , se construyó una cadena que une dichos puntos, de largo psicométrico menor y posee todos sus elementos entre m y n . \square

La proposición anterior implica que el ínfimo de los largo psicométricos de las cadenas que unen a los puntos m y n , es igual al ínfimo de los largos psicométricos de las cadenas que unen a los dichos puntos con elementos entre m y n . Pero dado que este último conjunto es finito, ya que la cantidad de elementos entre m y n es finita, se tiene que este ínfimo se alcanza. Luego, siempre existe $x^* \in C_{m,n}$ tal que $G(n, m) = 2 \cdot L^{(1)}(x^*)$. Esto motiva la siguiente definición

Definición 2.28 *Si x es una cadena que une a m con n es tal que $G(n, m) = 2 \cdot L^{(1)}(x^*)$, entonces se dirá que x genera la métrica entre m y n .*

Para los siguientes análisis, el siguiente lema será de gran utilidad.

Lema 2.29 *Sea $A \subset \mathbb{R}_+$, conjunto convexo tal que $0 \in A$. Sea $f : A \rightarrow \mathbb{R}$ estrictamente convexa, tal que $f(0) = 0$. Si $x, y, x + y \in A$, entonces*

$$f(x) + f(y) < f(x + y)$$

DEMOSTRACIÓN. Primero nótese que para $\lambda \in (0, 1)$ y $x > 0$, se tiene

$$f(\lambda x) = f(\lambda x + (1 - \lambda)0) < \lambda f(x) + (1 - \lambda)f(0)$$

pero como $f(0) = 0$, entonces

$$f(\lambda x) < \lambda f(x)$$

con esto se tiene que

$$\begin{aligned} f(x) + f(y) &= f\left(\frac{x+y}{x+y} \cdot x\right) + f\left(\frac{x+y}{x+y} \cdot y\right) \\ &< \frac{x}{x+y} f(x+y) + \frac{y}{x+y} f(x+y) \\ &= f(x+y) \end{aligned}$$

□

Ahora bien, si eventualmente la función $\tilde{\varphi}$ fuera estrictamente convexa, donde

$$\tilde{\varphi}(x) = \varphi(x) - 2\Phi\left(\frac{-\delta}{\sqrt{2\omega}}\right)$$

entonces se tendría que

$$\tilde{\varphi}\left(\log\left(\frac{h}{m}\right)\right) + \tilde{\varphi}\left(\log\left(\frac{n}{h}\right)\right) < \tilde{\varphi}\left(\log\left(\frac{n}{m}\right)\right)$$

donde $m < h < n$, con $m, h, n \in \mathbb{N}$. Dado que

$$\Psi^{(1)}(a, b) = \tilde{\varphi}\left(\log\left(\frac{b}{a}\right)\right) = \varphi\left(\log\left(\frac{b}{a}\right)\right) - 2\Phi\left(\frac{-\delta}{\sqrt{2\omega}}\right)$$

entonces se tiene

$$\Psi^{(1)}(m, h) + \Psi^{(1)}(h, n) < \Psi^{(1)}(m, n)$$

Es decir, la cadena (m, h, n) tiene menor largo psicométrico que la cadena (m, n) . Así, repitiendo el argumento con valores $h', h^{(2)} \in \mathbb{N}$, tal que $m < h' < h$ y $h < h^{(2)} < n$, e iterando hasta cubrir todos los enteros entre m y n , se tendría que la cadena que genera la métrica Fechneriana entre m y n sería $(m, m+1, \dots, n-1, n)$. Es importante notar que para aplicar este argumento, se usa fuertemente el hecho que los valores con los que se evalúa la función $\tilde{\varphi}$ sean positivos, por lo que no se podría ocupar un valor de h que no esté entre m y n . Pero dado que la cadena que genera la métrica entre m y n sólo posee elementos entre estos números, entonces no es necesario analizar el caso de $h > n$ o de $h < m$.

Sin embargo, como es de esperar, la función $\tilde{\varphi}$ no es convexa en todo el dominio. Esto se puede observar claramente en la parte positiva, ya que la función es creciente y acotada. Además, se puede encontrar situaciones en donde la cadena que genera la métrica entre dos estímulos no contiene a todos los números entre dichos estímulos. Por ejemplo, si se toma $\delta = 0,2$ y $\omega = 0,1$, entonces se tiene $\Psi^{(1)}(1,2) = 0,8425$, $\Psi^{(1)}(2,3) = 0,7696$ y $\Psi^{(1)}(1,3) = 0,8427$. Como se observa, la cadena que genera la métrica entre 1 y 3 es simplemente $(1,3)$.

Aún así, se pudo encontrar una condición suficiente para asegurar que la cadena que une m y n , y genera la métrica, contiene a todos los puntos j , tal que $m < j < n$.

Teorema 2.30 Sean $m, n \in \mathbb{N}$, tal que $m < n$. Si se cumple que

$$\log\left(\frac{n}{m}\right) + \delta > \left(\log\left(\frac{n}{m}\right) - \delta\right) \left(\frac{n}{m}\right)^{\frac{\delta}{\omega^2}}$$

entonces $\forall i \in \mathbb{N}$ tal que $m \leq i \leq n$ está en la cadena que genera la métrica entre m y n .

DEMOSTRACIÓN. Primero se caracterizará la región en donde la función $\tilde{\varphi}$ es convexa. Para ello se debe ver en qué parte la segunda derivada es positiva.

$$\begin{aligned} \tilde{\varphi}''(x) > 0 &\Leftrightarrow \Phi''\left(\frac{-\delta - x}{\omega\sqrt{2x}}\right) > -\Phi''\left(\frac{x - \delta}{\omega\sqrt{2x}}\right) \\ &\Leftrightarrow \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{-\delta - x}{\omega\sqrt{2}}\right)^2\right) \frac{\delta + x}{\sqrt{2\omega}} > \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - \delta}{\omega\sqrt{2}}\right)^2\right) \frac{x - \delta}{\sqrt{2\omega}} \\ &\Leftrightarrow \exp\left(-\frac{1}{2} \left(\frac{x + \delta}{\omega\sqrt{2}}\right)^2\right) (\delta + x) > \exp\left(-\frac{1}{2} \left(\frac{x - \delta}{\omega\sqrt{2}}\right)^2\right) (x - \delta) \\ &\Leftrightarrow (\delta + x) > \exp\left(-\frac{1}{2} \left(\frac{x - \delta}{\omega\sqrt{2}}\right)^2 + \frac{1}{2} \left(\frac{x + \delta}{\omega\sqrt{2}}\right)^2\right) (x - \delta) \\ &\Leftrightarrow (\delta + x) > \exp\left(\frac{x\delta}{\omega^2}\right) (x - \delta) \end{aligned}$$

Con esto, se define el conjunto

$$\Gamma_{\omega, \delta} = \left\{ x \in \mathbb{R} \mid x + \delta > (x - \delta) \exp\left(\frac{x\delta}{\omega^2}\right) \right\}$$

Dado que la función $\tilde{\varphi}$ es par, se tiene que $\tilde{\varphi}(x) > 0$ si y solo si $\tilde{\varphi}(-x) > 0$, luego, el conjunto $\Gamma_{\omega, \delta}$ es simétrico, es decir, si $x \in \Gamma_{\omega, \delta}$, entonces $-x \in \Gamma_{\omega, \delta}$.

Además, se verifica fácilmente que si $0 \leq x \leq \delta$, entonces $x \in \Gamma_{\omega, \delta}$, luego, por simetría se tiene

$$[-\delta, \delta] \subset \Gamma_{\omega, \delta}$$

Por otra parte, nótese que la función

$$f(x) = (x - \delta) \exp\left(\frac{x\delta}{\omega^2}\right)$$

es convexa para $x \geq \delta$. En efecto, se puede calcular que

$$f''(x) = \exp\left(\frac{x\delta}{\omega^2}\right) \left[\frac{2\delta}{\omega^2} + (x - \delta) \frac{\delta^2}{\omega^4} \right]$$

obteniendo así, dado que $\delta > 0$, que $f''(x) > 0$ para valores de x mayores a δ . Con esto, si $x_1, x_2 \in \Gamma_{\omega, \delta}$ tal que $x_1, x_2 \geq \delta$, entonces

$$x_1 + \delta > f(x_1)$$

$$x_2 + \delta > f(x_2)$$

luego, tomando $\lambda \in (0, 1)$, se tiene

$$(\lambda x_1 + (1 - \lambda)x_2) + \delta > \lambda f(x_1) + (1 - \lambda)f(x_2)$$

Pero dada la convexidad de f en valores mayores que δ se tiene

$$(\lambda x_1 + (1 - \lambda)x_2) + \delta > f(\lambda x_1 + (1 - \lambda)x_2)$$

Así se tiene que si un valor $x > \delta$ está en $\Gamma_{\omega, \delta}$, entonces, por convexidad se tiene $[\delta, x] \subset \Gamma_{\omega, \delta}$, y por lo dicho anteriormente, se obtiene

$$x \in \Gamma_{\omega, \delta} \Rightarrow [-x, x] \subset \Gamma_{\omega, \delta}$$

Ahora bien, el hecho que m y n cumplan las hipótesis del teorema es equivalente a

$$\log\left(\frac{n}{m}\right) \in \Gamma_{\omega, \delta}$$

Así, si $m \leq h \leq n$, entonces, dado que $1 < \frac{h}{m}, \frac{n}{h} < \frac{n}{m}$, entonces

$$\log\left(\frac{n}{h}\right), \log\left(\frac{h}{m}\right) \in \Gamma_{\omega, \delta}$$

Pero como $\Gamma_{\omega, \delta}$ es la región convexa en donde la función $\tilde{\varphi}$ es estrictamente convexa, se tiene que

$$\tilde{\varphi}\left(\log\left(\frac{h}{m}\right)\right) + \tilde{\varphi}\left(\log\left(\frac{n}{h}\right)\right) < \tilde{\varphi}\left(\log\left(\frac{h}{m}\right) + \log\left(\frac{n}{h}\right)\right)$$

lo que equivale a

$$\Psi^{(1)}(m, h) + \Psi^{(1)}(h, n) < \Psi^{(1)}(m, n)$$

Así, iterando este argumento con las parejas m y h , y h y n , las cuales, ya se sabe que cumplen con las hipótesis del teorema, se concluye que todos los puntos entre m y n están en la cadena que genera la métrica entre estos números \square

La hipótesis del teorema anterior podría parecer poco manejable. Sin embargo es fácil encontrar valores de m y n tales que la cumplan. En particular, basta considerar que $\log\left(\frac{n}{m}\right) < \delta$. Lo que intenta señalar el teorema recién demostrado es que cuando el cociente entre dos números se mantiene acotado, entonces la cadena que genera la métrica Fechneriana entre dichos números contiene todos los números entre m y n .

A continuación se estudiará el problema de declarar dos estímulos numéricos distintos bajo el modelo lineal con variabilidad escalar.

Modelo Lineal con Variabilidad Escalar

De manera análoga al modelo anterior, se considera que al visualizar un estímulo de tamaño n se realiza una variable aleatoria

$$V_n \sim N(n, \omega^2 n^2)$$

así se tiene la probabilidad de declarar x distinto de y está dada por

$$\psi_{EL}(x, y) = \mathbb{P}(|V_x - V_y| > \delta) = \Phi\left(\frac{-\delta - (x - y)}{\omega\sqrt{x^2 + y^2}}\right) - \Phi\left(\frac{\delta - (x - y)}{\omega\sqrt{x^2 + y^2}}\right) + 1$$

Donde $\delta > 0$ es un parámetro asociado al problema de discriminación. Este parámetro tiene la misma interpretación que en el caso anterior. Además, en el cálculo anterior, se supone la independencia entre V_x y V_y .

Si se quiere repetir en análisis hecho con el modelo de escala comprimida, una de las primeras cosas que se debería probar es el cumplimiento de la Minimalidad Regular. Sin embargo, la función ψ_{EL} no cumple con esta propiedad. Gráficamente, esto se puede observar en la figura 2.1, en donde se grafica la función $\psi_{EL}(7, \cdot)$ con parámetros $\delta = 0,6$ y $\omega = 0,9$. En la figura se observa que el mínimo no se alcanza en 7, si no más bien cerca de 4.

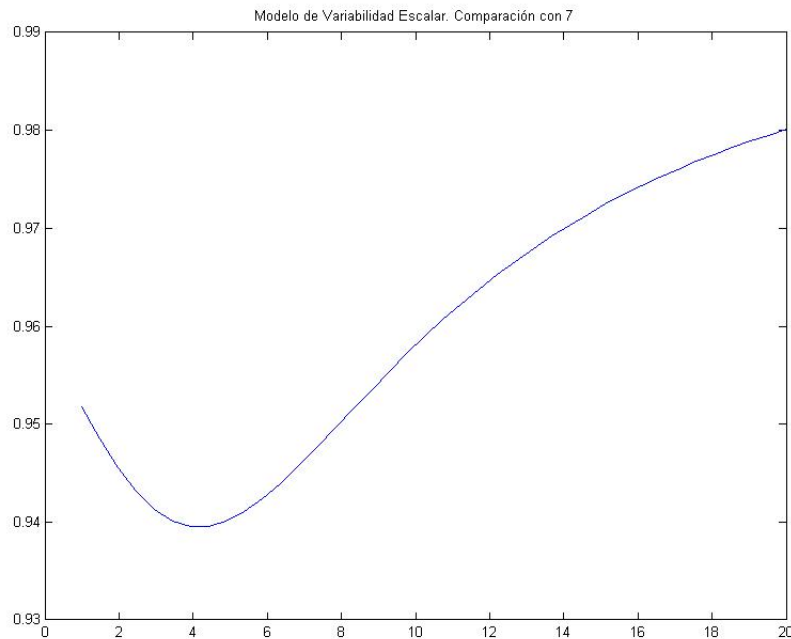


Figura 2.1: No cumplimiento de la Minimalidad Regular en el Modelo Lineal con Variabilidad Escalar. Se grafica la probabilidad de declarar un estímulo distinto de 7 según este modelo, considerando $\delta = 0,6$ y $\omega = 0,9$. Se observa que el mínimo de la función no se alcanza en 7.

Formalmente, se probó la siguiente proposición

Proposición 2.31 En el modelo de variabilidad escalar, se tiene que a x fijo, el mínimo de $\psi(x, \cdot)$ no se alcanza en x .

DEMOSTRACIÓN. Primero calculemos $\frac{\partial \psi(x, y)}{\partial y}$

Se tiene que

$$\begin{aligned} \frac{\partial \psi(x, y)}{\partial y} &= \phi(-h(x, y) + g(x, y)) \left[-\frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y} \right] \\ &\quad - \phi(h(x, y) + g(x, y)) \left[\frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y} \right] \end{aligned}$$

donde

$$\phi(x) = \frac{\exp\left[-\frac{x^2}{2}\right]}{\sqrt{2\pi}} \quad h(x, y) = \frac{\delta}{\omega\sqrt{x^2 + y^2}} \quad g(x, y) = \frac{-(x - y)}{\omega\sqrt{x^2 + y^2}}$$

La condición de mínimo implica que

$$\phi(-h(x, y) + g(x, y)) \left[-\frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y} \right] = \phi(h(x, y) + g(x, y)) \left[\frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y} \right]$$

como $\phi(x) \neq 0 \forall x$, se tiene que

$$\frac{\phi(-h(x, y) + g(x, y))}{\phi(h(x, y) + g(x, y))} \left[-\frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y} \right] = \frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y}$$

pero

$$\begin{aligned} \frac{\phi(-h(x, y) + g(x, y))}{\phi(h(x, y) + g(x, y))} &= \frac{\frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(-h(x, y) + g(x, y))^2}{2}\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(h(x, y) + g(x, y))^2}{2}\right)} \\ &= \exp\left(\frac{-(-h(x, y) + g(x, y))^2}{2} + \frac{(h(x, y) + g(x, y))^2}{2}\right) \\ &= \exp\left(\frac{-h^2 + 2hg - g^2 + h^2 + 2hg + g^2}{2}\right) \\ &= \exp(2h(x, y) \cdot g(x, y)) \end{aligned}$$

Así se tiene que el valor de y mínimo satisface

$$\exp\left(\frac{-2\delta(x - y)}{\omega^2(x^2 + y^2)}\right) \left[-\frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y} \right] = \frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y}$$

De este modo, si suponemos que $x = y$, $x \neq 0$, se tiene que cumplir que

$$-\frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y} = \frac{\partial h(x, y)}{\partial y} + \frac{\partial g(x, y)}{\partial y}$$

o bien

$$\frac{\partial h}{\partial y}(x, x) = 0$$

pero

$$\frac{\partial h}{\partial y}(x, x) = -\frac{\delta}{2\omega}(2x^2)^{-\frac{3}{2}}2x$$

por lo que en general, no se tiene que x sea mínimo de $\psi(x, \cdot)$

□

2.3. Método de Escalamiento Multidimensional

Una vez que se tiene una matriz de distancias para un conjunto de estímulos, resulta natural tratar de representar dichos estímulos en algún espacio de modo que las distancias entre las representaciones sea igual, o similar, a las distancias originales. Un método que se ajusta a estos requerimientos es el Método de Escalamiento Multidimensional.

Para partir, se va a suponer que se tiene n estímulos, los que serán enumerados de forma arbitraria desde 1 hasta n , y una matriz D de dimensión n , la cual contiene la distancia entre ellos. Así, el valor $D_{i,j}$ es igual a la distancia entre los estímulos i y j . Además, cada estímulo se le asociará un punto en \mathbb{R}^q , que en principio se desconocen, de modo que

$$D_{i,j} = \|x_i - x_j\|$$

con x_i y $x_j \in \mathbb{R}^q$. El objetivo de este método es encontrar n vectores de manera de recuperar las distancias por medio de la norma de las diferencias. Obviamente si el conjunto $\{x_i \mid i \in \{1, \dots, n\}\}$ conserva las distancia de la matriz D , entonces también lo hará el conjunto $\{x_i + c \mid i \in \{1, \dots, n\}\}$ con c un vector fijo. Es más, cualquier transformación isométrica sobre los puntos también será una solución del problema en cuestión. Una primera condición que se impondrá sobre el conjunto de vectores es

$$\sum_{i=1}^n x_i = 0$$

Por otra parte, si se asume que a partir de la matriz distancias D se puede construir la matriz de productos internos $W \in \mathcal{M}_{n \times n}(\mathbb{R})$, esto es

$$W_{i,j} = \langle x_i, x_j \rangle$$

entonces cualquier descomposición de la matriz W de la forma

$$W = X^t \cdot X$$

con $X \in \mathcal{M}_{q \times n}(\mathbb{R})$, permite resolver el problema, siendo las columnas de X los vectores que se buscan. Ahora bien, si W es efectivamente una matriz de productos internos, entonces se tiene que es simétrica semidefinida positiva, de modo que se puede descomponer como

$$W = P \cdot \Lambda \cdot P^t$$

donde Λ es una matriz diagonal, en que el término $\Lambda_{i,i}$ corresponde al i -ésimo valor propio más grande de W , y P es la matriz que contiene a la base ortonormal de vectores propios de W , donde la i -ésima columna de P es el vector propio asociado al valor propio $\Lambda_{i,i}$.

Así, una matriz X posible se escribe como

$$X = \Lambda^{\frac{1}{2}} \cdot P^t$$

Una de las principales ventajas de elegir esta descomposición de la matriz W para encontrar X , es el criterio de optimalidad que subyace en este método. En efecto, si eventualmente se quisiera representar los estímulos en un espacio de dimensión p , con $p < q$, entonces una manera de buscar las representaciones $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n \in \mathbb{R}^p$ sería imponiendo que dichos vectores minimicen

$$\sum_{i,j=1}^n (W_{i,j} - \langle \hat{x}_i, \hat{x}_j \rangle)^2$$

Sin embargo, es sabido que la solución a este problema de optimización está dada por

$$\hat{X} = I_{p \times n} \cdot \Lambda^{\frac{1}{2}} \cdot P^t$$

donde $I_{p \times n}$ tiene la forma

$$(I_{p \times n})_{i,j} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Con esto se tiene que la representación en un espacio de dimensión p queda definida por los p valores propios más grande de la matriz W y sus respectivos vectores propios. Aumentar o disminuir la dimensión de la representación corresponde a agregar o quitar vectores propios, respectivamente.

Con este resultado, sólo queda establecer cuál es la manera de reconstruir la matriz de productos internos W a partir de la matriz de distancias D . En primer lugar, nótese que

$$D_{i,j}^2 = \|x_i\|^2 + \|x_j\|^2 - 2 \cdot \langle x_i, x_j \rangle$$

luego

$$\langle x_i, x_j \rangle = \frac{1}{2} [\|x_i\|^2 + \|x_j\|^2 - D_{i,j}^2]$$

Obviamente, aún queda determinar cuánto valen las normas de las representaciones. Pero, en vista de la restricción impuesta sobre la suma de las representaciones, se tiene que

$$\|x_i\|^2 = \|x_i - \bar{x}\|^2$$

donde $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0$. Por otra parte, nótese que

$$\begin{aligned} D_{i,j}^2 &= \|x_i - x_j\|^2 \\ &= \|x_i - z - (x_j - z)\|^2 \\ &= \|x_i - z\|^2 + \|x_j - z\|^2 - 2 \langle x_i - z, x_j - z \rangle \end{aligned}$$

Así, sumando sobre i queda

$$\sum_{i=1}^n D_{i,j}^2 = \sum_{i=1}^n \|x_i - z\|^2 + n \cdot \|x_j - z\|^2 - 2 \left\langle \sum_{i=1}^n x_i - n \cdot z, x_j - z \right\rangle$$

luego, tomando $z = \bar{x}$, queda que

$$\sum_{i=1}^n D_{i,j}^2 = \sum_{i=1}^n \|x_i - \bar{x}\|^2 + n \cdot \|x_j - \bar{x}\|^2$$

notando $D_{:j}^2 = \frac{1}{n} \sum_{i=1}^n D_{i,j}^2$, se sigue que

$$\|x_j - \bar{x}\|^2 = D_{:j}^2 - \frac{1}{n} \sum_{i=1}^n \|x_i - \bar{x}\|^2$$

Lo que queda por averiguar es cómo expresar la cantidad $\sum_{i=1}^n \|x_i - \bar{x}\|^2$ en términos de la matriz de distancias. Ocupando la ecuación anterior se tiene que

$$\begin{aligned} \sum_{i=1}^n \|x_i - \bar{x}\|^2 &= \sum_{i=1}^n \left(D_{:i}^2 - \frac{1}{n} \sum_{j=1}^n \|x_j - \bar{x}\|^2 \right) \\ &= \sum_{i=1}^n D_{:i}^2 - \sum_{i=1}^n \|x_i - \bar{x}\|^2 \end{aligned}$$

por consiguiente

$$\frac{1}{n} \sum_{i=1}^n \|x_i - \bar{x}\|^2 = \frac{\sum_{i=1}^n D_{:i}^2}{2n}$$

Definiendo $D_{..}^2 = \frac{\sum_{i,j} D_{i,j}^2}{n^2}$ queda, finalmente que

$$\|x_j - \bar{x}\|^2 = D_{:j}^2 - \frac{1}{2} D_{..}^2$$

Con esto se tiene

$$\begin{aligned} W_{i,j} &= \langle x_i, x_j \rangle \\ &= \frac{1}{2} [\|x_i\|^2 + \|x_j\|^2 - D_{i,j}^2] \\ &= \frac{1}{2} \left[D_{:i}^2 - \frac{1}{2} D_{..}^2 + D_{:j}^2 - \frac{1}{2} D_{..}^2 - D_{i,j}^2 \right] \\ &= \frac{1}{2} [D_{:i}^2 + D_{:j}^2 - D_{..}^2 - D_{i,j}^2] \end{aligned}$$

Así, se ha conseguido expresar la matriz de productor internos en función de la matriz de distancia. Esto matricialmente se puede escribir como

$$W = \frac{1}{2} (I - A) \cdot R \cdot (I - A)$$

siendo $A = \frac{1}{n} u \cdot u^t$ donde $u \in \mathbb{R}^n$ y todas sus componentes son iguales a 1, y $R_{i,j} = D_{i,j}^2$

Los cálculos expuestos en esta sección son estándar, y el método se encuentra implementado en distintos softwares. Para analizar otros métodos similares se recomienda revisar [38].

2.3.1. Un Ejemplo Aplicado

Para apreciar una aplicación conjunta de la construcción de la métrica Fechneriana y el método de escalamiento multidimensional, se analizaron los datos de un experimento de aprendizaje. Los sujetos observaban frente a un computador una variedad de kanjis, símbolos propios de la escritura japonesa, y debían declarar si era iguales o distintos.

Específicamente, cada trial del experimento comenzaba cuando los sujetos presionaban la barra espaciadora del teclado, luego pasaban entre 1,5 a 2 segundos, apareciendo el primer estímulo por 100 milisegundos. Después de esto, por 500 milisegundos se presentaba una máscara, esto es, una pantalla ruidosa, y luego por 1 segundo una pantalla negra. Seguido de esto, se presentaba un segundo estímulo kanji por 100 milisegundos, para continuar con otra máscara por 500 milisegundos y una pantalla negra por 1 segundo. Luego, se presentaba por 100 milisegundos un tercer estímulo kanji, el cual estaba rotado en 90° a la izquierda o a la derecha. Finalmente se presentaba una última máscara por 500 milisegundos y se esperaba la respuesta del sujeto. Éste debía contestar si el primer con el tercer estímulo era iguales o no.

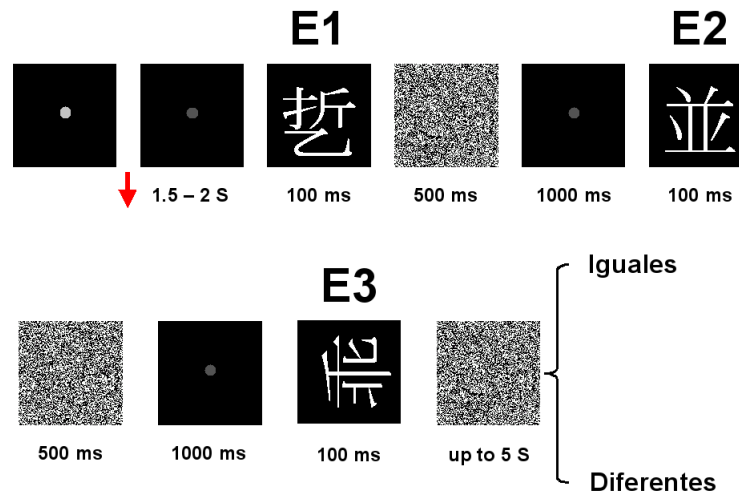


Figura 2.2: Esquema de cada trial del Experimento

Se trabajó con los datos de 7 sujetos, los cuales respondieron 8 bloques de 50 trials cada uno. Dentro del experimento se presentaron 30 símbolos kanjis, y las preguntas fueron generadas de forma aleatoria.

El análisis aquí expuesto corresponde a un estudio global de la muestra, por lo que no se presentarán conclusiones a nivel de individuos. Para analizar la evolución de las respuestas a medida que el tiempo avanza se decidió dividir el conjunto de datos en dos grupos. El primero contienen las respuestas de los primeros cuatro bloques, mientras que el segundo contiene a los datos restantes. Con esto, y una vez etiquetados los estímulos del 1 al 30, se construyeron dos tablas de probabilidad, ambas de 30 filas y 30 columnas, en donde la componente (i, j) contiene una estimación de la probabilidad de declarar que i es distinto de j . De esta manera, las tablas de probabilidad corresponden a estimaciones de la función ψ necesaria para la construcción de la métrica Fechneriana.

Para evitar sobrestimar las probabilidad, se decidió construir las estimaciones considerando un suavizamiento de Laplace, es cual se define como

$$\hat{\psi}(i, j) = \frac{x_{i,j} + 1}{n_{i,j} + 2}$$

donde $x_{i,j}$ corresponde a la cantidad de veces que se declara i distinto de j , mientras que $n_{i,j}$ es la cantidad de veces que se contrastaron los estímulos i y j . Este tratamiento es estándar para estimar probabilidades [52].

Construidas estas tablas, se procedió a verificar el cumplimiento de la propiedad de Minimalidad Regular. Esto no se cumplió, por lo que se decidió retirar un conjunto de estímulos del análisis. De este modo, con 22 de los 30 estímulos, se cumplió la propiedad necesaria para la construcción de la métrica Fechneriana. Así, para cada una de las tablas se construyó la métrica Fechneriana y con la matriz de distancias obtenida se aplicó el método de escalamiento multidimensional.

Para el primer conglomerado de bloques, la matriz de distancias Fechnerianas se reconstruye con vectores de dimensión 13, mientras que para el segundo grupo esto se logra con vectores de dimensión 14. Además, se calculó la varianza promedio de las representaciones, obteniendo para el primer conglomerado un valor de 0.5915 y para el segundo un valor de 0.6457. Esto hace inferir, que al menos en promedio, las representaciones del segundo conglomerado se encuentran más alejadas entre sí, es decir, es más fácil discriminar.

A continuación, en las figuras 2.3 y 2.4 se presentan los estímulos representados en las dos primeras dimensiones para cada conglomerado. El propósito de estos gráficos es tratar de resumir la información entrega por los experimentos para poder visualizar de manera sencilla qué estímulos se consideran con alta probabilidad distintos y cuáles no. Estímulos que se encuentran muy cercanos entre sí se declararán en general como similares, mientras que aquellos alejados serán declarados como distintos. La representación de los dos grupos fue orientada principalmente para observar si a lo largo del tiempo podían o no cambiar las medidas de similaridad. Sin embargo, a primera vista no se observa un patrón evidente.

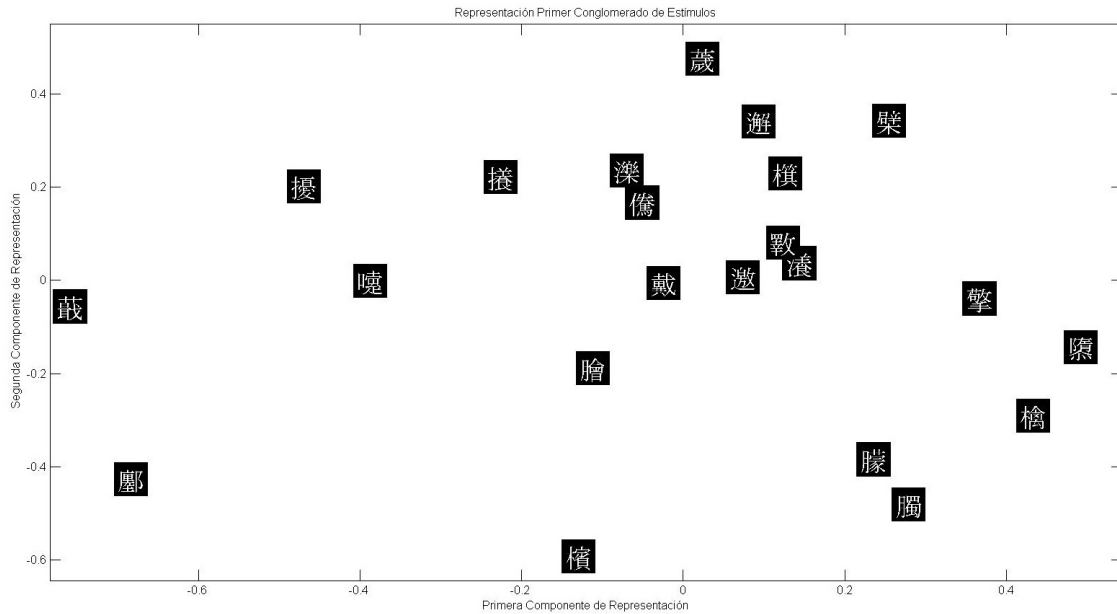


Figura 2.3: Representación en dos dimensiones del Primer Conglomerado de Estímulos

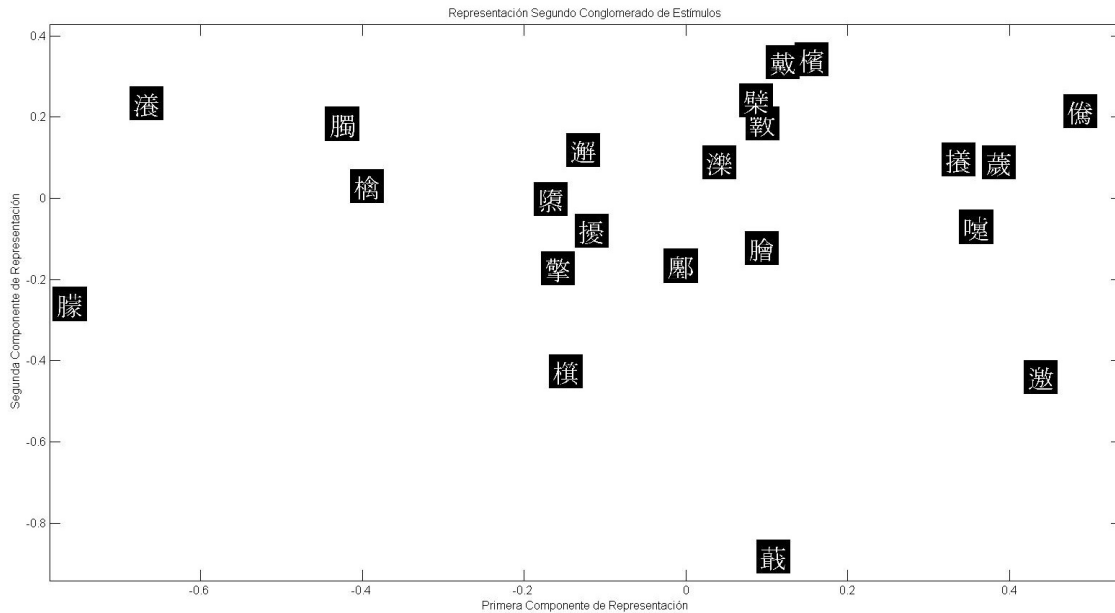


Figura 2.4: Representación en dos dimensiones del Segundo Conglomerado de Estímulos

2.4. Conclusiones

El principal propósito de este capítulo fue comparar los modelos de escala comprimida y el lineal con varibilidad escalar. Para ello se analizaron los modelos bajo el problema de discriminación en dos enfoques. El primer enfoque, consistió en analizar los modelos bajo la situación en que frente a dos estímulos, se pide identificar el de mayor numerosidad. Ambos modelos indican que la probabilidad de identificar el mayor de los estímulos depende del cuociente de las cantidades que se comparan. Además, se mostró que ambos modelos son

casos particulares de modelos más generales que hacen la misma predicción. Conjuntamente, se mostró que el modelo lineal tiene un comportamiento asintótico distinto al modelo de escala comprimida, esto es, en la medida que se hace crecer la numerosidad de uno de los estímulos, según el modelo lineal, la probabilidad de identificar el mayor siempre será acotada por algo menor a 1, mientras que el modelo de escala comprimida indica que dicha probabilidad converge a uno en la medida que uno de los estímulos crezca a infinito. Esta es una de las primeras diferencias teóricas que se encontraron entre los modelos.

El segundo enfoque, consistió en analizar el problema de la declaración de diferencia, es decir, frente a dos estímulos numéricos, declarar si los estímulos son de igual numerosidad o no. Una propiedad que parece ser razonable, y que debería cumplirse, es la propiedad de minimalidad regular, esto es, que siempre la probabilidad de declarar que dos estímulos de tamaño n son distintos es menor que la probabilidad de declarar que un estímulo de tamaño n es distinto a un estímulo de tamaño m , con $n \neq m$. Esta propiedad se cumple para el modelo de escala comprimida, lo que permite calcular lo que se conoce como métrica Fechneriana entre estímulos. Sin embargo, al trabajar con el modelo lineal con variabilidad escalar, resulta que la propiedad de minimalidad regular no se cumple, es decir, se podría dar que declarar “7 es distinto de 4” es menos probable que declarar “7 es distinto de 7”, esto en experimentos con numerosidades no simbólicas.

A partir de lo expuesto en este capítulo, se considera que el modelo de escala comprimida se ajusta mejor a los hechos experimentales que el modelo de lineal con variabilidad escalar. De este modo, trabajar con la escala logarítmica para representar a las numerosidades es más conveniente que trabajar con la escala lineal. Esto será la base para el modelo que se propone en el siguiente capítulo.

Capítulo 3

Estimación y Modelo de Acumuladores

“Hay quienes... creen que el número de granos de arena es infinito... Otros, aun sin considerarlo infinito, piensan que todavía no se ha mencionado un número lo bastante grande... Pero voy a tratar de mostrarte números que superen no sólo el de una masa de arena equivalente a la Tierra... sino el de una masa igual en magnitud al Universo”

Arquímedes

El Arenario

La estimación de cantidades es otro de los experimentos conductuales clásicos en cognición matemática. En esta tarea a los sujetos se les presenta una estímulo numérico no simbólico, generalmente una nube de puntos, y se les pide dar una estimación de la cantidad asociada al estímulo. Como se dijo en el capítulo 1, hay varios resultados sistemáticos que se observan experimentalmente, entre ellos la tendencia a subestimar y la denominada propiedad de variabilidad escalar; esto indica que el cociente entre el promedio y la desviación estándar de una cantidad estimada es independiente de dicha cantidad.

Además, dentro de esta tarea aparece el fenómeno de la Calibración; si el experimentador entrega información al sujeto de cuál es la numerosidad de una nube de puntos específica, las estimaciones de todas las numerosidades mejoran. Esta pequeña cantidad de información dada al inicio del experimento hace que la curva estímulo respuesta pueda ajustarse sin problemas a una recta.

Un modelo matemático para describir este fenómeno es el desarrollado por Dehaene [31] [13]. Éste se basa en el modelo de Escala Comprimida y en la Teoría de Detección de Señales, haciendo el supuesto que cuando se observa un estímulo de tamaño n , se realiza una variable aleatoria $N(\log(n), \sigma^2)$ en la escala numérica mental, la que a su vez se encuentra particionada en zonas asociadas a los números con los cuales se hará la estimación pedida. Así, si la realización de la variable aleatoria cae, por ejemplo, en la zona del 30, se declarará que el estímulo es de cardinal 30 (ver figura 3.1).

La partición de la recta numérica mental, la cual se hace corresponder a la recta real, queda

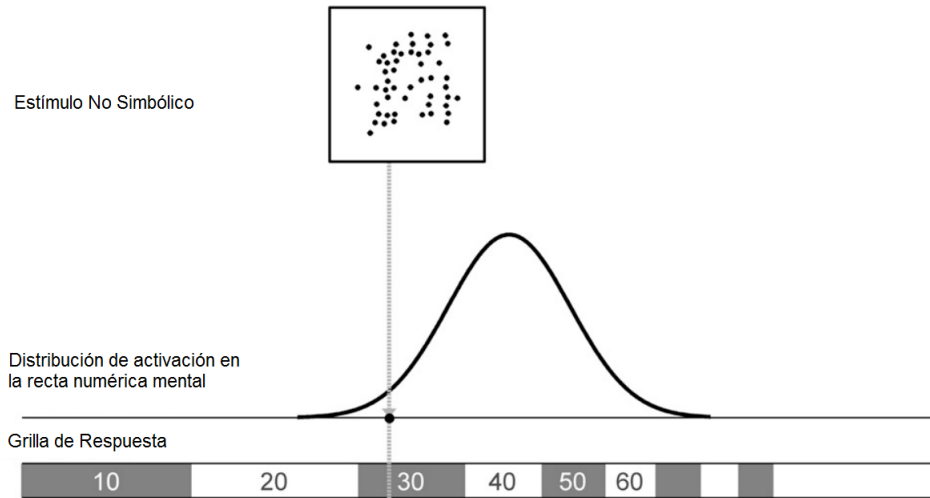


Figura 3.1: Esquema del Modelo de Estimación Numérica de Dehaene. Se ejemplifica un caso en donde se muestra una nube de 40 puntos. La realización de la variable aleatoria cae en la recta numérica mental en la zona del número 30, por lo que se declarará que la nube contiene 30 puntos. Fuente: [31]

en principio sujeta al tipo de respuestas que pueda dar el sujeto. En algunos experimentos la respuesta es libre, mientras que en otros, por razones de diseño experimental, las respuestas quedan acotadas a una serie de alternativas. Lo que prevalece es que en ambos casos las estimaciones deben ser números enteros positivos. El modo en que se realiza la partición de la recta numérica mental se basa en un criterio de optimalidad. La región de la recta asociada a n será aquella en donde la función de densidad de $K_n \sim N(\log(n), \omega^2)$ sea mayor a todas las funciones de densidad de K_m con $m \neq n$. Esto se traduce, en el caso de declaración libre que la región de una numerosidad de tamaño n sería el intervalo

$$\left[\frac{\log(n) + \log(n-1)}{2}, \frac{\log(n) + \log(n+1)}{2} \right]$$

A su vez, para el caso de respuestas restringidas, si se tiene que las respuestas son $x_1 < x_2 < \dots < x_n$, entonces la región asociada a la respuesta x_i sería

$$\left[\frac{\log(x_i) + \log(x_{i-1})}{2}, \frac{\log(x_i) + \log(x_{i+1})}{2} \right]$$

En la figura 3.2 se muestra un caso de respuestas restringidas a múltiplos de 10 hasta 100.

Dada la descripción del modelo, se tiene que éste predice que los sujetos tienden a sobreestimar la numerosidad de los estímulos, cosa que en general no se observa. Para que el modelo se ajuste a los resultados experimentales, Dehaene propone una corrección de las particiones de la recta numérica mental, de modo que si las respuestas admisibles son $x_1 < x_2 < \dots < x_n$,

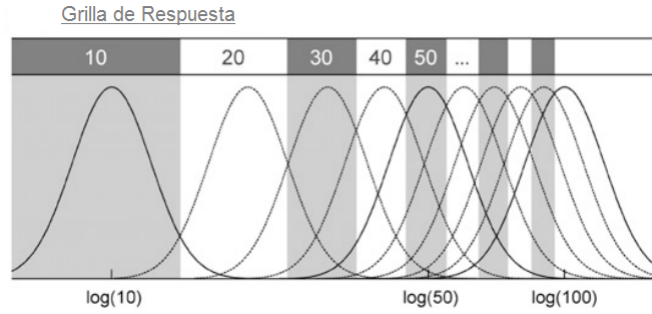


Figura 3.2: Ejemplo de Partición de la Recta Numérica Mental. Fuente: [31]

entonces la región asociada a la respuesta x_i está dada por

$$\left[a \cdot \frac{\log(x_i) + \log(x_{i-1})}{2} + b, a \cdot \frac{\log(x_i) + \log(x_{i+1})}{2} + b \right]$$

donde a y b son parámetros a determinar. Por medio de simulaciones, el autor identifica que para valores de a mayores a 1 se tiene un buen ajuste de la subestimación de la numerosidad. Por último, el fenómeno de la Calibración, en términos del modelo, se explica como un cambio de los parámetros a y b tal que se logre una mejora en la calidad de las estimaciones.

Con todo lo mencionado, si bien se considera que el modelo tiene bases razonables, como el trabajo con la recta numérica mental y la partición de ésta para generar las estimaciones, el hecho de que el ajuste a la evidencia empírica quede determinado por parámetros que en principio no tienen significado teórico, genera un grado de disconformidad y se considera razón suficiente para buscar un modelo alternativo.

En este capítulo se introducirá y estudiará dos modelos para la estimación de numerosidad. Ambos hacen uso de la recta numérica mental, pero están constituidos por distintos supuestos. El primer modelo, denominado Modelo de Acumuladores Difusos, se fundamenta en que cada punto en la recta numérica mental tiene un peso al momento de observar un estímulo de numerosidad ε . Estos pesos tienen una dinámica asociada, la cual se basa en un proceso de difusión. La manera de estimar dependerá de la realización de una variable aleatoria cuya función de densidad es proporcional a los pesos descritos. El segundo modelo, el cual se denominó Modelo de Competencia, se fundamenta en el hecho que los pesos sobre cada punto en la recta dependen tanto del estímulo expuesto como de la cantidad de neuronas asociadas a cada punto de la recta numérica.

Qué tan bien se ajustan estos modelos a los resultados experimentales y qué tipo de predicciones pueden hacer, será discutido en las siguientes secciones.

3.1. Modelo de Acumuladores Difusos

Este modelo tiene como base algunos supuestos considerados por Dehaene en su modelo. El elemento central será la recta numérica mental. Se asumirá que existe una correspondencia entre todo punto de la recta numérica mental con los números positivos. Esto se puede

apoyar en la evidencia que indica que incluso las fracciones se pueden representar una la recta numérica y no los números negativos [22]. Con lo indicado, a cada posición x en la recta numérica mental se le asociará los valores $a(x)$, el cual es el valor positivo representado en la posición x de la recta.

El modelo aquí desarrollado trata de sostenerse en el descubrimiento de las neuronas específicas de numerosidad. Así, un grupo de neuronas afín a una numerosidad dada quedará representada por una posición x en la recta, de modo que cuando se observe un estímulo de numerosidad ε , la actividad neuronal de dicho grupo debería ser proporcional a la cantidad $A_\varepsilon(x, t)$, que se denominará como el valor del acumulador de x frente al estímulo ε en el tiempo t .

Además, se asumirá que

$$A_\varepsilon(x, t) \geq 0 \quad \forall x \in \mathbb{R}, \forall t > 0$$

y que

$$0 < \bar{A}_\varepsilon(t) = \int_{-\infty}^{\infty} A_\varepsilon(x, t) dx < \infty \quad \forall t > 0$$

Luego, definiendo

$$f_{\varepsilon, t}(x) = \frac{A_\varepsilon(x, t)}{\bar{A}_\varepsilon(t)}$$

se tiene una función de densidad en la recta numérica mental.

Se considera entonces, que al observar un estímulo de numerosidad ε , después de un tiempo t , se realiza la variable aleatoria $X_{\varepsilon, t}$, que tiene por densidad de probabilidad a la función $f_{\varepsilon, t}$. Así, la estimación de la numerosidad del estímulo queda dada por $a(X_{\varepsilon, t})$.

En lo que sigue se dará una dinámica a los valores de los acumuladores e intentará estudiar los valores de

$$\mathbb{E}[a(X_{\varepsilon, t})]$$

3.1.1. Dinámica de Acumuladores

Basado en la idea que la actividad de neuronas afines a numerosidades similares son influyentes unas sobre las otras, se asume la siguiente dinámica sobre los acumuladores cuando se observa un estímulo de numerosidad ε

$$\begin{aligned} A_\varepsilon(x, t + \Delta t) &= \left(\frac{\lambda}{2} [A_\varepsilon(x + \Delta x, t) + A_\varepsilon(x - \Delta x, t)] + (1 - \lambda) A_\varepsilon(x, t) \right) (1 - \omega \cdot \Delta t) \\ &\quad + g_\varepsilon(x) \cdot \Delta t \end{aligned}$$

Es decir, el valor del acumulador en la posición x y en el tiempo $t + \Delta t$ es igual a una combinación convexa entre el valor del acumulador en el tiempo t y en la posición x con

los valores de los acumuladores en las posiciones vecinas $x - \Delta x$ y $x + \Delta x$ en el tiempo t . Además se considera que los valores de los acumuladores decaen en la medida que el tiempo transcurre, lo que explica el factor $(1 - \omega \cdot \Delta t)$. Por último, se asume que la observación del estímulo de numerosidad ε sigue teniendo efecto despues de un tiempo Δt . Este efecto sobre los acumuladores en la posición x queda establecido por la suma de la cantidad $g_\varepsilon(x) \cdot \Delta t$. A la función g_ε , que depende de ε , se le denominará función de evidencia, y se exigirá que sea positiva e integrable.

Nótese que la ecuación anterior se puede reescribir como

$$\frac{A_\varepsilon(x, t + \Delta t) - A_\varepsilon(x, t)}{\Delta t} = \left[\frac{A_\varepsilon(x + \Delta x, t) + A_\varepsilon(x - \Delta x, t) - 2A_\varepsilon(x, t)}{2 \cdot \Delta x^2} \right] \lambda \frac{\Delta x^2}{\Delta t} (1 - \omega \cdot \Delta t) - \omega A_\varepsilon(x, t) + g_\varepsilon(x)$$

asumiendo que $\frac{\Delta x^2}{\Delta t} \rightarrow D$, cuando $\Delta t \rightarrow 0$ y $\Delta x \rightarrow 0$, se llega a la ecuación

$$\frac{\partial A_\varepsilon(x, t)}{\partial t} = \frac{\lambda D}{2} \frac{\partial^2 A_\varepsilon(x, t)}{\partial x^2} - \omega A_\varepsilon(x, t) + g_\varepsilon(x) \quad (3.1)$$

Como condición inicial de los valores de los acumuladores, se prefirió dejar

$$A_\varepsilon(x, 0) = h(x) \quad (3.2)$$

donde h es una función positiva e integrable. Así, esta función modelará el tipo de sesgo que tienen los sujetos hacia los números pequeños, o la incorporación de información al inicio del experimento.

Finalmente, llamando $\eta = \frac{\lambda D}{2}$, se tiene la ecuación

$$\begin{cases} \frac{\partial A_\varepsilon(x, t)}{\partial t} &= \eta \frac{\partial^2 A_\varepsilon(x, t)}{\partial x^2} - \omega A_\varepsilon(x, t) + g_\varepsilon(x) \\ A_\varepsilon(x, 0) &= h(x) \end{cases}$$

3.1.2. Solución

Al ser una ecuación diferencial parcial 3.1 lineal, exigiendo que tanto la función g_ε y h sean continuas y tengan decaimiento exponencial cuando x tienda a ∞ o $-\infty$, se tiene que dicha ecuación tiene solución única y positiva [19].

Para encontrar la solución de manera explícita, se hará uso de la Transformada de Fourier. Aplicándola sobre la ecuación (3.1) en la variable x se tiene

$$\frac{\partial \hat{A}_\varepsilon(s, t)}{\partial t} = -s^2 \eta \hat{A}_\varepsilon(s, t) - \omega \hat{A}_\varepsilon(s, t) + \hat{g}_\varepsilon(s)$$

Es fácil ver que esta ecuación diferencial ordinaria, considerando como condición inicial la aplicación de la transformada de Fourier sobre la ecuación 3.2, tiene por solución a

$$\hat{A}_\varepsilon(s, t) = \frac{\hat{g}_\varepsilon(s)}{\eta s^2 + \omega} + e^{-[\eta s^2 + \omega]t} \left(\hat{h}(s) - \frac{\hat{g}_\varepsilon(s)}{\eta s^2 + \omega} \right)$$

Pero, debe notarse que

$$\frac{\hat{g}_\varepsilon(s)}{\eta s^2 + \omega} = \frac{1}{2\sqrt{\omega\eta}} \cdot e^{-\widehat{\sqrt{\frac{\omega}{\eta}}|x|}}(s) \cdot \hat{g}_\varepsilon(s)$$

y que

$$\begin{aligned} e^{-[\eta s^2 + \omega]t} &= e^{-\omega t} \cdot e^{-\eta s^2 t} \\ &= e^{-\omega t} \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot 2\sqrt{\pi\eta t} \cdot e^{-\frac{4\eta t}{4}s^2} \\ &= e^{-\omega t} \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\widehat{\frac{x^2}{4\eta t}}}(s) \end{aligned}$$

Así se tiene, aplicando antitransformada, que

$$A_\varepsilon(x, t) = \frac{1}{2\sqrt{\omega\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * g_\varepsilon(x) + e^{-\omega t} \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\frac{x^2}{4\eta t}} * \left[h(x) - \frac{1}{2\sqrt{\omega\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * g_\varepsilon(x) \right]$$

Lo que implica

$$\bar{A}_\varepsilon(t) = \int_{-\infty}^{\infty} A_\varepsilon(x, t) dx = \frac{1}{\omega} G + e^{-\omega t} \left(H - \frac{1}{\omega} G \right)$$

donde $G = \int_{-\infty}^{\infty} g_\varepsilon(x) dx$ y $H = \int_{-\infty}^{\infty} h(x) dx$. Se considerará que $\int_{-\infty}^{\infty} g_\varepsilon(x) dx$ no depende de ε

3.1.3. Evaluación para caso particular

Si se asume que la recta numérica mental está en escala logarítmica, entonces, una posibilidad es considerar que la posición x en la recta se asocia a la numerosidad $a(x)$ como lo hace el modelo de Dehaene, esto es

$$a(x) = n \text{ si } x \in \left[\frac{\log(n) + \log(n-1)}{2}, \frac{\log(n) + \log(n+1)}{2} \right]$$

Sin embargo, para ganar analiticidad, se considerará

$$a(x) = e^x$$

La similitud entre estas dos funciones se puede apreciar en la figura 3.3. Dado que $a(X_{\varepsilon,t})$ es la variable aleatoria que modela la estimación que entrega el sujeto cuando observa un estímulo de numerosidad ε después de un tiempo t , podría ser criticable que la función a no tenga como recorrido el conjunto de los números naturales. Sin embargo, en este estudio se está interesado en el comportamiento de la $\mathbb{E}(a(X_{\varepsilon,t}))$, por lo que trabajar con la función definida por Dehaene se hace tremendamente engorroso. Preferir la función exponencial permite llegar a resultados analíticos bastante precisos.

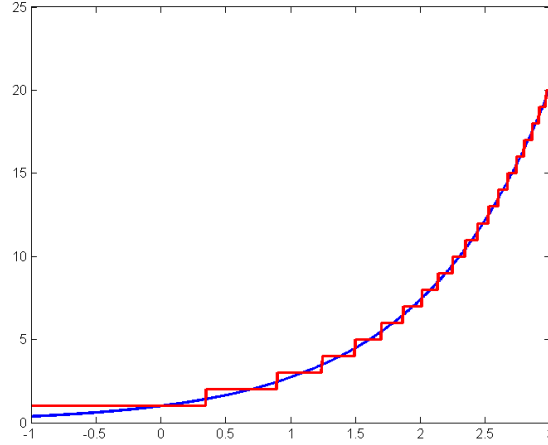


Figura 3.3: Comparación de funciones de asociación a números positivos. La línea azul es el gráfico de la función $a(x) = e^x$, mientras que la línea roja es el gráfico de la función $a(x) = n$ si $x \in \left[\frac{\log(n-1)+\log(n)}{2}, \frac{\log(n)+\log(n+1)}{2} \right]$

Considerando que $a(x) = e^x$, se tiene

$$\mathbb{E}[a(X_{\varepsilon,t})] = \int_{-\infty}^{\infty} e^x \cdot \frac{A_{\varepsilon}(x,t)}{\bar{A}_{\varepsilon}(t)} dx$$

Dada la dinámica descrita para $A_{\varepsilon}(x,t)$, se tiene la siguiente proposición

Proposición 3.1 *Sea $X_{\varepsilon,t}$ la variable aleatoria con densidad $f_{\varepsilon,t}$. Si $a(x) = e^x$, entonces*

$$\mathbb{E}[a(X_{\varepsilon,t})] = \xi_1(t) \cdot \mathbb{E}[e^{Z_{\varepsilon}}] + \xi_2(t) \cdot \mathbb{E}[e^Q]$$

donde Z_{ε} es una variable aleatoria que tiene como función de densidad $\frac{g_{\varepsilon}}{G}$, y Q la variable aleatoria que tiene por densidad $\frac{h}{H}$

DEMOSTRACIÓN. Dado los cálculos anteriores, se tiene

$$\begin{aligned}
\int_{-\infty}^{\infty} e^x \cdot A_\varepsilon(x, t) dx &= \int_{-\infty}^{\infty} e^x \frac{1}{2\sqrt{\omega\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * g_\varepsilon(x) dx \\
&+ \int_{-\infty}^{\infty} e^x \cdot e^{-\omega t} \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\frac{x^2}{4\eta t}} * h(x) dx \\
&- \int_{-\infty}^{\infty} e^x \cdot e^{-\omega t} \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\frac{x^2}{4\eta t}} * \frac{1}{2\sqrt{\omega\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * g_\varepsilon(x) dx
\end{aligned}$$

Ahora bien para el primer término se tiene

$$\begin{aligned}
\int_{-\infty}^{\infty} e^x \frac{1}{2\sqrt{\omega\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * g_\varepsilon(x) dx &= \frac{G}{\omega} \int_{-\infty}^{\infty} e^x \cdot \frac{1}{2} \sqrt{\frac{\omega}{\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * \frac{g_\varepsilon(x)}{G} dx \\
&= \frac{G}{\omega} \mathbb{E} [e^{Y+Z_\varepsilon}]
\end{aligned}$$

con Y variable aleatoria que se distribuye como una Laplace $(0, \sqrt{\frac{\eta}{\omega}})$ y $Z_\varepsilon \sim G_\varepsilon$, donde G_ε es la distribución de probabilidad asociada a la densidad $\frac{g_\varepsilon}{G}$ teniendo además que Y y Z_ε son independientes.

Luego, se tiene, asumiendo $\eta < \omega$,

$$\begin{aligned}
\mathbb{E} [e^{Y+Z_\varepsilon}] &= \mathbb{E} [e^Y] \cdot \mathbb{E} [e^{Z_\varepsilon}] \\
&= \frac{1}{1 - \frac{\eta}{\omega}} \cdot \mathbb{E} [e^{Z_\varepsilon}] \\
&= \frac{\omega}{\omega - \eta} \cdot \mathbb{E} [e^{Z_\varepsilon}]
\end{aligned}$$

con lo que se tiene

$$\int_{-\infty}^{\infty} e^x \frac{1}{2\sqrt{\omega\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * g_\varepsilon(x) dx = \frac{G}{\omega - \eta} \cdot \mathbb{E} [e^{Z_\varepsilon}]$$

Por otra parte. el segundo término se puede calcular como

$$\begin{aligned}
\int_{-\infty}^{\infty} e^x \cdot e^{-\omega t} \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\frac{x^2}{4\eta t}} * h(x) dx &= e^{-\omega t} \cdot H \int_{-\infty}^{\infty} e^x \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\frac{x^2}{4\eta t}} * \frac{h(x)}{H} dx \\
&= e^{-\omega t} \cdot H \cdot \mathbb{E} [e^{W_t+Q}]
\end{aligned}$$

donde Q y W_t son variables aleatoria independientes tales que $W_t \sim N(0, 2\eta t)$ y $Q \sim H(x)$, donde $H(x)$ es la función de distribución asociada a la densidad $\frac{h(x)}{H}$

Así se tiene

$$\begin{aligned}\mathbb{E} [e^{W_t+Q}] &= \mathbb{E} [e^{W_t}] \mathbb{E} [e^Q] \\ &= e^{\eta t} \cdot \mathbb{E} [e^Q]\end{aligned}$$

quedando que

$$\int_{-\infty}^{\infty} e^x \cdot e^{-\omega t} \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\frac{x^2}{4\eta t}} * h(x) dx = H e^{-(\omega-\eta)t} \cdot \mathbb{E} [e^Q]$$

Finalmente, se tiene que el tercer término se puede escribir como

$$\begin{aligned}\int_{-\infty}^{\infty} e^x \cdot e^{-\omega t} \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\frac{x^2}{4\eta t}} * \frac{1}{2\sqrt{\omega\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * g_\varepsilon(x) dx &= e^{-\omega t} \frac{G}{\omega} \int_{-\infty}^{\infty} e^x \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\frac{x^2}{4\eta t}} * \frac{1}{2} \sqrt{\frac{\omega}{\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * \frac{g_\varepsilon(x)}{G} dx \\ &= e^{-\omega t} \frac{G}{\omega} \cdot \mathbb{E} [e^{W_t+Y+Z_\varepsilon}]\end{aligned}$$

donde W_t , Y , Z_ε se distribuyen como se mencionó anteriormente y son independientes. Luego

$$\begin{aligned}\mathbb{E} [e^{W_t+Y+Z_\varepsilon}] &= \mathbb{E} [e^{W_t}] \cdot \mathbb{E} [e^Y] \cdot \mathbb{E} [e^{Z_\varepsilon}] \\ &= e^{\eta t} \cdot \frac{\omega}{\omega - \eta} \cdot \mathbb{E} [e^{Z_\varepsilon}]\end{aligned}$$

por lo que se tiene

$$\int_{-\infty}^{\infty} e^x \cdot e^{-\omega t} \cdot \frac{1}{2\sqrt{\pi\eta t}} \cdot e^{-\frac{x^2}{4\eta t}} * \frac{1}{2\sqrt{\omega\eta}} \cdot e^{-\sqrt{\frac{\omega}{\eta}}|x|} * g_\varepsilon(x) dx = \frac{G}{\omega - \eta} \cdot e^{-(\omega-\eta)t} \cdot \mathbb{E} [e^{Z_\varepsilon}]$$

Con todo esto, se puede decir que

$$\begin{aligned}\int_{-\infty}^{\infty} e^x \cdot A_\varepsilon(x, t) dx &= \varepsilon \cdot \frac{G}{\omega - \eta} \cdot e^{\frac{\sigma^2}{2}} + H e^{\frac{\sigma^2}{2}} e^{-(\omega-\eta)t} - \varepsilon \frac{G}{\omega - \eta} \cdot e^{-(\omega-\eta)t} \cdot e^{\frac{\sigma^2}{2}} \\ &= \frac{G}{\omega - \eta} \cdot \mathbb{E} [e^{Z_\varepsilon}] (1 - e^{-(\omega-\eta)t}) + H \cdot e^{-(\omega-\eta)t} \cdot \mathbb{E} [e^Q]\end{aligned}$$

y se tiene que

$$\mathbb{E} [a(X_{\varepsilon, t})] = \frac{\frac{G}{\omega - \eta} \cdot \mathbb{E} [e^{Z_\varepsilon}] (1 - e^{-(\omega-\eta)t}) + H \cdot e^{-(\omega-\eta)t} \cdot \mathbb{E} [e^Q]}{\frac{1}{\omega} G + e^{-\omega t} (H - \frac{1}{\omega} G)}$$

luego, definiendo

$$\xi_1(t) = \frac{\frac{G}{\omega - \eta} (1 - e^{-(\omega-\eta)t})}{\frac{1}{\omega} G + e^{-\omega t} (H - \frac{1}{\omega} G)}$$

y

$$\xi_2(t) = \frac{H \cdot e^{-(\omega-\eta)t}}{\frac{1}{\omega}G + e^{-\omega t} \left(H - \frac{1}{\omega}G \right)}$$

queda

$$\mathbb{E} [a(X_{\varepsilon,t})] = \xi_1(t) \cdot \mathbb{E} [e^{Z_\varepsilon}] + \xi_2(t) \cdot \mathbb{E} [e^Q]$$

□

Recuérdese que Z_ε tiene directa relación con la función g , la que a su vez representa la incorporación de peso a los acumuladores $A_\varepsilon(x, t)$ cuando se observa un estímulo de numerosidad ε . Por otra parte, la variable aleatoria Q se relaciona con la función h , la cuál determinaba los valores de los acumuladores al inicio del proceso de observación del estímulo. Este se puede relacionar, como se dijo, con un sesgo hacia los números pequeños o la incorporación de información al inicio del experimento. Dar distribuciones a las variables Z_ε y Q que se ajusten a las ideas anteriores, siendo consistente con el escalamiento logarítmico permite tener una idea más concreta de cómo son las estimaciones que indica el modelo cuando se observa un estímulo. La siguiente proposición intenta hacerse cargo de esto.

Proposición 3.2 *Si*

$$Z_\varepsilon \stackrel{\text{ley}}{=} \log(\varepsilon) + Z$$

y

$$Q \stackrel{\text{ley}}{=} \log(x_0) + Z$$

donde Z es una variable aleatoria dada, tal que $\mathbb{E}(Z) = 0$ y $\mathbb{E}(e^Z) = K < \infty$, entonces

$$\mathbb{E} [a(X_{\varepsilon,t})] = \tilde{\xi}_1(t) \cdot \varepsilon + \tilde{\xi}_2(t) \cdot x_0$$

DEMOSTRACIÓN. Dado que

$$\mathbb{E}(e^{Z_\varepsilon}) = \mathbb{E}(e^{\log(\varepsilon)+Z}) = \varepsilon \cdot K$$

Además, por desigualdad de Jensen, $K > 1$. De igual manera se tiene

$$\mathbb{E}(e^Q) = \mathbb{E}(e^{\log(x_0)+Z}) = x_0 \cdot K$$

lo que implica

$$\mathbb{E} [a(X_{\varepsilon,t})] = \tilde{\xi}_1(t) \cdot \varepsilon + \tilde{\xi}_2(t) \cdot x_0$$

donde $\tilde{\xi}_i(t) = \xi_i(t) \cdot K$ con $i \in 1, 2$.

□

La proposición anterior indica que los valores de $\mathbb{E}(a(X_{\varepsilon,t}))$ a t fijo, se representan mediante funciones lineales afines de la numerosidad del estímulo. En relación a la hipótesis de esta proposición, no es muy difícil encontrar funciones g_ε y h que generen las variables aleatorias Z_ε y Q que cumplan lo pedido. Un ejemplo sería

$$g_\varepsilon(x) = e^{-\frac{(\log(\varepsilon)-x)^2}{\sigma^2}}$$

$$h(x) = e^{-\frac{(\log(x_0)-x)^2}{\sigma^2}}$$

Por otra parte, hay que observar que $\tilde{\xi}_1(0) = 0$ y $\tilde{\xi}_2(0) = K$, por lo que las estimaciones cuando $t = 0$ no dependen del valor de ε , si no más bien de x_0 . Una explicación para esto sería que para el inicio del experimento el sujeto no tiene información del estímulo, por lo que la estimación sería la declaración de un número al azar. Está estudiado que cuando se les pide a los sujetos declarar números al azar, estos tienen a decir más números pequeños [39]. El valor de x_0 también puede asociarse a la información que recibe el sujeto, de modo que si recibe información sobre numerosidades altas, la predicción de este modelo es que todas las declaraciones aleatorias en el instante $t = 0$, es decir, antes de comenzar a observar el estímulo, deberían aumentar.

A su vez, cuando $t \rightarrow \infty$ se tiene que

$$\tilde{\xi}_1(t) \rightarrow \frac{\omega}{\omega - \eta}$$

$$\tilde{\xi}_2(t) \rightarrow 0$$

De modo que la esperanza cuando t tiende a infinito converge a una recta que pasa por el origen y tiene pendiente mayor a 1. Naturalmente, la influencia de la condición inicial desaparece y se tiene una tendencia a la sobreestimación. Además, por continuidad de las funciones ξ_1 y ξ_2 se tiene que existen tiempos en donde la estimación tiene a subestimar las numerosidades de números grandes (mayores a $K \cdot x_0$).

Una potencial crítica al modelo es la dependencia del tiempo; el análisis anterior se hizo a tiempo t fijo. Sin embargo, no es claro el comportamiento de los tiempos que tardan en declarar las estimaciones los sujetos. Una alternativa es considerar que el tiempo en que se tarda en hacer la estimación desde que se observa el estímulo es una variable aleatoria T , por lo que los cálculos anteriores corresponden a una esperanza condicional, por lo que se tiene, manteniendo los supuestos del modelo que

$$\mathbb{E}[a(X_{\varepsilon,t}) | T = t] = \tilde{\xi}_1(t) \cdot \varepsilon + \tilde{\xi}_2(t) \cdot x_0$$

Si se asume que T es independiente de ε y de x_0 , entonces se tiene que

$$\mathbb{E}[a(X_{\varepsilon,t})] = \mathbb{E}(\mathbb{E}[a(X_{\varepsilon,t}) | T = t]) = K_1 \cdot \varepsilon + K_2 \cdot x_0$$

Por lo que la dependencia lineal de las numerosidades sobre las estimaciones se mantiene. Es importante aclarar que no hay información precisa de los tiempos de respuestas en la

tarea de estimación de magnitudes, de modo que, si bien el supuesto de independencia puede ser considerado poco plausible, tampoco existe una relación clara entre los tiempos y las numerosidades a estimar.

Si bien, el modelo logra incorporar condiciones para predecir subestimación, se tiene dos problemas. Primero, no se recupera el hecho experimental que las estimaciones tienden a ajustarse a funciones potencia sobre la numerosidad, es decir, funciones de la forma $K \cdot \varepsilon^\alpha$. El segundo problema es cómo se modela el fenómeno de la Calibración. Como se ha dicho, la incorporación de información se asociaría al modelo mediante una modificación del valor x_0 . Por ejemplo, si al sujeto no se le entrega información, entonces se puede asumir que $x_0 = 1$ o algún otro valor pequeño. Sin embargo, si al sujeto se le presenta una nube de 30 puntos y se le indica su numerosidad al inicio del experimento, entonces se puede considerar $x_0 = 30$.

Lamentablemente, esta modificación en las condiciones iniciales no logra ajustarse a lo observado experimentalmente, ya que en vez de una rectificación de la curva de estimaciones, sólo se logra una traslación de la curva ya rectificada.

Esta disparidad con la evidencia experimental hace necesario perfeccionar el modelo o buscar una alternativa. El siguiente modelo tiene una base común, el uso de los acumuladores $A_\varepsilon(x, t)$. La diferencia radica en no incorporar el efecto de difusión pero agregar un efecto sobre la cantidad de neuronas asociadas a cada numerosidad.

3.2. Modelo de Competencia

A diferencia del modelo anterior, se considerará la siguiente dinámica sobre los valores de los acumuladores

$$\begin{aligned} \frac{\partial A_\varepsilon(x, t)}{\partial t} &= -\omega A_\varepsilon(x, t) + g_\varepsilon(x) \\ A_\varepsilon(x, 0) &= h(x) \end{aligned}$$

Se puede observar que no se consideró el efecto de difusión sobre valores cercanos. De la ecuación diferencial anterior, es fácil notar que su solución está dada por

$$\begin{aligned} A_\varepsilon(x, t) &= \frac{g_\varepsilon(x)}{\omega} + e^{-\omega t} \left[h(x) - \frac{g_\varepsilon(x)}{\omega} \right] \\ &= \frac{g_\varepsilon(x)}{\omega} [1 - e^{-\omega t}] + e^{-\omega t} \cdot h(x) \end{aligned}$$

Se considerará esta vez la función $c(x)$, cuyo valor es proporcional a la cantidad de neuronas asociadas al valor x en la recta numérica mental. De este modo, se generará la variable aleatoria $\tilde{X}_{\varepsilon, t}$ tal que su función de densidad esté dada por

$$\tilde{f}_{\varepsilon, t}(x) = \frac{c(x) \cdot A_\varepsilon(x, t)}{\int_{-\infty}^{\infty} c(x) \cdot A_\varepsilon(x, t) dx}$$

Nuevamente se estudiará

$$\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]$$

3.2.1. Esperanza de la Estimación

En vista de la definición de $\tilde{X}_{\varepsilon}(x, t)$, se tiene

$$\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right] = \frac{\int_{-\infty}^{\infty} a(x) \cdot c(x) \cdot A_{\varepsilon}(x, t) dx}{\int_{-\infty}^{\infty} c(x) \cdot A_{\varepsilon}(x, t) dx} \quad (3.3)$$

pero

$$\begin{aligned} \int_{-\infty}^{\infty} c(x) \cdot A_{\varepsilon}(x, t) dx &= \frac{(1 - e^{-\omega t}) G}{\omega} \int_{-\infty}^{\infty} c(x) \cdot \frac{g_{\varepsilon}(x)}{G} dx + e^{-\omega t} \cdot H \cdot \int_{-\infty}^{\infty} c(x) \cdot \frac{h(x)}{H} dx \\ &= \frac{(1 - e^{-\omega t}) G}{\omega} \cdot \mathbb{E} [c(Z_{\varepsilon})] + e^{-\omega t} \cdot H \cdot \mathbb{E} [c(Q)] \end{aligned}$$

haciendo un cálculo análogo para el numerador de la expresión 3.3, se llega a

$$\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right] = \frac{\frac{(1 - e^{-\omega t}) G}{\omega} \cdot \mathbb{E} [a(Z_{\varepsilon}) \cdot c(Z_{\varepsilon})] + e^{-\omega t} \cdot H \cdot \mathbb{E} [a(Q) \cdot c(Q)]}{\frac{(1 - e^{-\omega t}) G}{\omega} \cdot \mathbb{E} [c(Z_{\varepsilon})] + e^{-\omega t} \cdot H \cdot \mathbb{E} [c(Q)]}$$

donde Q es una variable aleatoria que tiene por función de densidad $\frac{h(x)}{H}$ y Z_{ε} una variable aleatoria con densidad $\frac{g_{\varepsilon}}{G}$, con $G = \int_{-\infty}^{\infty} g_{\varepsilon}(x) dx$ y $H = \int_{-\infty}^{\infty} h(x) dx$.

3.2.2. Evaluación de Caso Particular

Al igual que en el estudio del modelo anterior, se considerará

$$a(x) = e^x$$

La cantidad de neuronas asociadas a la posición de la recta numérica mental x , se considerará decreciente en x ; en particular se tomará

$$c(x) = e^{-\lambda x}$$

Además, de igual manera que en el modelo anterior, se impondrá

$$Z_{\varepsilon} \stackrel{\text{ley}}{=} \log(\varepsilon) + Z$$

$$Q \stackrel{\text{ley}}{=} \log(x_0) + Z$$

donde Z es una variable aleatoria dada.

Con todo esto se tiene que

$$\mathbb{E} [a(Z_\varepsilon) \cdot c(Z_\varepsilon)] = \varepsilon^{(1-\lambda)} \cdot \mathbb{E} [e^{(1-\lambda)Z}]$$

$$\mathbb{E} [a(Q) \cdot c(Q)] = x_0^{(1-\lambda)} \cdot \mathbb{E} [e^{(1-\lambda)Z}]$$

y

$$\mathbb{E} [c(Z_\varepsilon)] = \varepsilon^{-\lambda} \cdot \mathbb{E} [e^{-\lambda Z}]$$

$$\mathbb{E} [c(Q)] = x_0^{-\lambda} \cdot \mathbb{E} [e^{-\lambda Z}]$$

Por lo que

$$\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right] = \frac{\mathbb{E} [e^{(1-\lambda)Z}]}{\mathbb{E} [e^{-\lambda Z}]} \cdot \frac{\frac{(1-e^{-\omega t})G}{\omega H} \cdot \varepsilon^{(1-\lambda)} + e^{-\omega t} \cdot x_0^{(1-\lambda)}}{\frac{(1-e^{-\omega t})G}{\omega H} \cdot \varepsilon^{-\lambda} + e^{-\omega t} \cdot x_0^{-\lambda}}$$

En particular, si tomamos

$$Z \sim N(0, \sigma^2)$$

entonces

$$\frac{\mathbb{E} [e^{(1-\lambda)Z}]}{\mathbb{E} [e^{-\lambda Z}]} = e^{\frac{1}{2}\sigma^2[(1-\lambda)^2 - \lambda^2]} = e^{\frac{1}{2}\sigma^2(1-2\lambda)}$$

Por lo que finalmente se llega a

$$\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right] = e^{\frac{1}{2}\sigma^2(1-2\lambda)} \cdot \frac{\frac{(1-e^{-\omega t})G}{\omega H} \cdot \varepsilon^{(1-\lambda)} + e^{-\omega t} \cdot x_0^{(1-\lambda)}}{\frac{(1-e^{-\omega t})G}{\omega H} \cdot \varepsilon^{-\lambda} + e^{-\omega t} \cdot x_0^{-\lambda}} \quad (3.4)$$

El comportamiento de las estimaciones generadas por el modelo cumplen varias propiedades

Proposición 3.3 *Si $\lambda < 1$, entonces a t fijo los valores de $\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]$ dados por la ecuación 3.4 son crecientes en ε , cuando $\varepsilon > 0$*

DEMOSTRACIÓN. A t fijo, se tiene que la forma de las estimaciones es de la forma

$$\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right] = K \cdot \frac{\alpha \cdot \varepsilon^{1-\lambda} + \beta}{\alpha \cdot \varepsilon^\lambda + \gamma}$$

derivando con respecto a ε queda

$$\frac{d\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]}{d\varepsilon} = K \cdot \frac{\alpha \cdot (1 - \lambda) \varepsilon^{-\lambda} \cdot (\alpha \cdot \varepsilon^\lambda + \gamma) + (\alpha \cdot \varepsilon^{1-\lambda} + \beta) (\alpha \lambda \varepsilon^{-1-\lambda})}{(\alpha \cdot \varepsilon^\lambda + \gamma)^2}$$

y se tiene que esta expresión es positiva si $\lambda < 1$ □

De manera totalmente análoga a la proposición anterior, se puede probar que

Proposición 3.4 *Los valores de $\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]$ son crecientes en x_0 cuando x_0 es positivo.*

Además se tiene que en la medida que el tiempo avanza, las estimaciones van aumentando o disminuyendo de acuerdo a la numerosidad del estímulo y del valor de x_0 .

Proposición 3.5 *A ε fijo, los valores de $\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]$ son crecientes en t si $\varepsilon > x_0$, y son decrecientes si $\varepsilon < x_0$.*

DEMOSTRACIÓN. A ε fijo, la forma de las estimaciones son

$$\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right] = K \cdot \frac{(1 - e^{-\omega t}) \alpha_1 + e^{-\omega t} \cdot \beta_1}{(1 - e^{-\omega t}) \alpha_2 + e^{-\omega t} \cdot \beta_2}$$

Derivando con respecto a t queda

$$\frac{d\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]}{dt} = K \cdot \frac{\omega \cdot e^{-\omega t} [\alpha_1 \cdot \beta_2 - \alpha_2 \cdot \beta_1]}{((1 - e^{-\omega t}) \alpha_2 + e^{-\omega t} \cdot \beta_2)^2}$$

Que esta expresión sea positiva equivale a

$$\alpha_1 \cdot \beta_2 - \alpha_2 \cdot \beta_1 > 0$$

pero esto se puede escribir como

$$\frac{G}{\omega H} \varepsilon^{1-\lambda} \cdot x_0^{-\lambda} - \frac{G}{\omega H} \varepsilon^{-\lambda} \cdot x_0^{1-\lambda} > 0$$

lo que equivale a decir

$$\varepsilon - x_0 > 0$$

Así, si ε es mayor a x_0 , entonces la esperanza de las estimaciones aumenta en la medida que el tiempo aumenta. A su vez, si ε es menor a x_0 , entonces la esperanza de las numerosidades disminuye en la medida que el tiempo disminuye. □

Una interpretación que se le puede dar a la proposición anterior es que, a tiempos pequeños desde que se comienza a observar el estímulo una vez dada la información para calibrar, las numerosidades menores al estímulo de calibración tienden a sobreestimarse, mientras que numerosidades mayores al estímulo de calibración tienen a subestimarse.

Por último, es fácil notar que se tiene las siguientes propiedades

Proposición 3.6 *Bajo las condiciones descritas, se tiene*

$$i) \mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,0} \right) \right] = e^{\frac{1}{2}\sigma^2(1-2\lambda)} \cdot x_0$$

$$ii) \mathbb{E} \left[a \left(\tilde{X}_{x_0,t} \right) \right] = e^{\frac{1}{2}\sigma^2(1-2\lambda)} \cdot x_0$$

$$iii) \text{ Cuando } t \rightarrow \infty, \text{ entonces } \mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right] \rightarrow e^{\frac{1}{2}\sigma^2(1-2\lambda)} \cdot \varepsilon$$

Es bueno notar que si bien el modelo no indica que la esperanza de las estimaciones es una función potencial de las numerosidades de los estímulos, se tiene que la forma que toma $\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]$ es similar a una de estas funciones. Un punto importante es que el fenómeno de la Calibración es bien descrito por el modelo, ya que queda definido por un ajuste de la condición inicial de los acumuladores, lo que se traduce en aumentar el valor x_0 . Esto logra que la curva de estimaciones se rectifique, mejorando la calidad de éstas. En la figura 3.4 se observa cómo afecta sobre las estimaciones cambiar la condición inicial de los acumuladores.

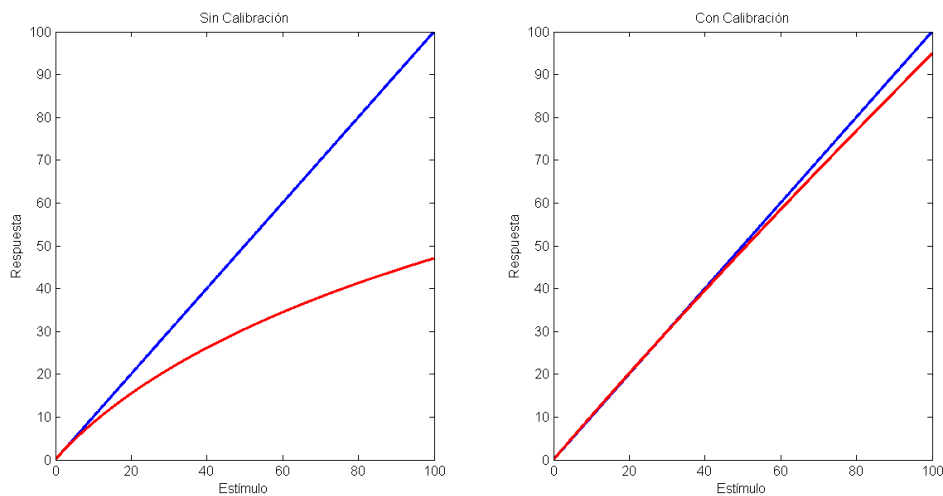


Figura 3.4: Estimaciones según modelo de Competencias Sin y Con Calibración. La línea azul en ambos gráficos muestra la función identidad. En ambos gráficos se ocuparon los parámetros $\omega = 0,1$, $H = 1$, $G = 1$, $\sigma = 0,1$ y $\lambda = 0,8$. En el gráfico de la izquierda se muestra las estimaciones que genera el modelo con $x_0 = 1$, lo que se considera como modelación de la estimación sin dar información. En el gráfico de la derecha se tomó $x_0 = 30$, lo que en términos del modelo se traduce como incorporar información, replicando la Calibración de la Estimación.

3.2.3. Variabilidad Escalar

En vista que el modelo tiene similitudes con el comportamiento de la esperanza de las estimaciones en función de la numerosidad del estímulo, una propiedad que debería estudiarse es la variabilidad escalar. En términos del modelo, se debería estudiar la cantidad

$$\frac{\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]}{\sqrt{\text{Var} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]}}$$

o bien estudiar la cantidad

$$\frac{\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]^2}{\mathbb{E} \left[a^2 \left(\tilde{X}_{\varepsilon,t} \right) \right] - \mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]^2} \quad (3.5)$$

Si bien, hay reportes en donde se indica que esta propiedad no siempre se registra [31], también es cierto que es ampliamente aceptada por los científicos dedicados al tema [6] [14].

Mostrar que la cantidad dada por 3.5 no depende de ε equivale a mostrar que la siguiente expresión tampoco lo hace

$$\frac{\mathbb{E} \left[a^2 \left(\tilde{X}_{\varepsilon,t} \right) \right]}{\mathbb{E} \left[a \left(\tilde{X}_{\varepsilon,t} \right) \right]^2} \quad (3.6)$$

Una igualdad necesaria para estudiar la cantidad 3.6 , a la cual se llega de manera similar que en 3.4, es

$$\mathbb{E} \left[a^2 \left(\tilde{X}_{\varepsilon,t} \right) \right] = e^{\sigma^2(2-2\lambda)} \cdot \frac{\frac{(1-e^{-\omega t})G}{\omega H} \cdot \varepsilon^{(2-\lambda)} + e^{-\omega t} \cdot x_0^{(2-\lambda)}}{\frac{(1-e^{-\omega t})G}{\omega H} \cdot \varepsilon^{-\lambda} + e^{-\omega t} \cdot x_0^{-\lambda}}$$

Así se tiene que 3.6 es igual a

$$e^{\sigma^2} \cdot \frac{\left[\frac{(1-e^{-\omega t})G}{\omega H} \cdot \varepsilon^{-\lambda} + e^{-\omega t} \cdot x_0^{-\lambda} \right] \left[\frac{(1-e^{-\omega t})G}{\omega H} \cdot \varepsilon^{(2-\lambda)} + e^{-\omega t} \cdot x_0^{(2-\lambda)} \right]}{\left[\frac{(1-e^{-\omega t})G}{\omega H} \cdot \varepsilon^{(1-\lambda)} + e^{-\omega t} \cdot x_0^{(1-\lambda)} \right]^2}$$

La expresión anterior sí depende de ε , sin embargo, cuando $t \rightarrow \infty$ se tiene que ésta converge a e^{σ^2} , por lo que cuando t tiene a infinito se tiene la propiedad de variabilidad escalar.

3.3. Conclusiones y Preguntas Abiertas

En este capítulo se presentaron dos modelos para explicar el fenómeno de la Estimación de Cantidades y su Calibración. El primer modelo, basado en un proceso de difusión en la recta numérica mental, terminó por indicar que las estimaciones tienden a ser rectas y que la entrega de información haría que las estimaciones aumentaran de manera uniforme, lo que se traduce en un cambio del coeficiente de posición de las rectas mencionadas. Ninguna de estas dos cosas se observa experimentalmente.

El segundo modelo, consideraba un efecto sobre la cantidad de neuronas asociadas a cada posición de la recta numérica mental. Este modelo logra generar estimaciones acordes con los datos experimentales y el fenómeno de la calibración. Esto último se considera en el modelo como un cambio de la condición inicial de la dinámica, generando una rectificación de las estimaciones. Además se tiene la variabilidad escalar como propiedad asintótica.

Este último modelo cuenta con dos predicciones que deberían ser testeadas experimentalmente

- A mayor tiempo de exposición del estímulo, el valor esperado de las estimaciones debería tender a rectificarse.
- A mayor el valor del estímulo con que se realiza la calibración, mayores deberían ser las estimaciones.

Una predicción adicional estaría relacionada con el efecto que tiene un trail con los que le prosiguen en un experimento. Podría ser que la misma tarea de estimar afecte las estimaciones de un intento al otro. Esto, en términos del modelo, se podría explicar a partir de los acumuladores, ya que entre trial y trial podría darse que las condiciones iniciales no serían exactamente las mismas. Es decir, la condición inicial al momento de estimar la numerosidad del trail n podría ser la numerosidad del trail $n - 1$.

Capítulo 4

Un Estudio sobre la Enseñanza de Fracciones

“Nuestros niños aprenden matemáticas con un cerebro diseñado inicialmente para sobrevivir en la sabana africana... Una gran cantidad de niños encuentra las fracciones difíciles de aprender porque su maquinaria cortical se resiste a un concepto contrario a la intuición.”

Stanislas Dehaene

El Sentido Numérico

En los capítulos anteriores se desarrollaron algunos modelos matemáticos para el sentido numérico. Estos, se basaron en la enorme evidencia empírica que existe en experimentaciones con estímulos no simbólicos, tales como nubes de puntos, secuencias de tonos, u otros.

El objetivo de este capítulo es utilizar algunas de las ideas desarrolladas anteriormente, principalmente la de recta numérica mental, para tratar de abordar un problema crucial en la investigación en educación matemática; el aprendizaje de las fracciones.

En primer lugar, se expondrá algunos hechos conocidos sobre las distintas dificultades que existen en torno al concepto de fracción. Luego, se hará un rápido recorrido de las distintas investigaciones que se han desarrollado estudiando este tema. Posteriormente se intentará dar un marco conceptual para tratar de explicar el origen de estas dificultades y cómo tratar de resolverlas. Finalmente, para tratar de validar las ideas del modelo conceptual, se describe un experimento que se llevó a cabo como parte de este trabajo, el cual, contrastó tres métodos de enseñanza de fracciones en más de 200 niños chilenos. Además, se proponen algunos métodos estadísticos, los cuales permitirían hacer una inferencia sobre el tipo de criterios que utilizan los niños para comparar fracciones.

4.1. Marco Teórico

4.1.1. La dificultad en la enseñanza de fracciones

Para una gran cantidad de investigadores, la tarea de enseñar fracciones es declarada como “el mayor desafío pedagógico” [21] [8]. Diversos son los estudios que indican las serias deficiencias que tienen los estudiantes en este tema. Por ejemplo, el 2004 en Estados Unidos, se detectó que el 50 % de los alumnos de 8vo grado no podían ordenar tres fracciones de menor a mayor [46]. Junto con esto, menos del 30 % de los alumnos de 17 años pudieron interpretar 0,029 como $\frac{29}{1000}$ [33]. También se indica que la mayoría de los alumnos de 5to y 6to grado considera que 0,274 es mayor que 0,83 [60].

Además, existen dificultades al relacionar las fracciones con la recta numérica. Al momento en que se les pide a los niños ubicar fracciones en la recta, algunos ubican un determinado número racional de acuerdo a la fracción del segmento de recta que se utiliza en la tarea, así por ejemplo, se observa que en algunos casos $\frac{1}{2}$ es ubicado al medio del segmento de recta con independencia de los números que aparecen en ella. Por otra parte, hay estudiantes que consideran que la fracción $\frac{a}{b}$ debe ubicarse entre los números a y b [65]. Un ejemplo de esto se muestra en la figura 4.1, en donde incluso se ubica en la recta $\frac{1}{0}$.

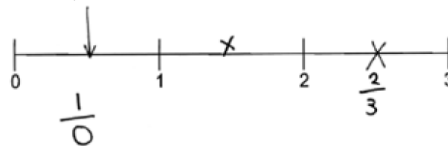


Figura 4.1: Error común al momento de ubicar fracciones en la recta numérica. Cuando a un niño se le pide determinar la posición en que ubicaría $\frac{1}{2}$, él lo posiciona al centro del segmento. Cuando se le pide ubicar $\frac{2}{3}$, lo hace entre 2 y 3. Cuando se le pide señalar la fracción que se ubica donde indica la fecha, él indica $\frac{1}{0}$.

Existen distintas explicaciones para entender el origen de estas dificultades. Algunos apelan a la estructura del curriculum escolar, en donde parece ser que el principal propósito es que en determinados niveles los niños logren hacer determinadas tareas matemáticas. Esto hace sentir a los estudiantes que lo aprendido en un año lo pueden olvidar en el siguiente. En relación a las fracciones, éstas son enseñadas simultáneamente como partes de un entero, una división y una razón sin hacer un vínculo claro entre estas tres formas de representación. Además, los números decimales son enseñados de forma independiente de las fracciones, siendo el eje fundamental el de números enteros. El no hacer vínculos claros entre las distintas ideas que orbitan en torno al concepto de fracción podría causar serios daños en el aprendizaje.

Ejemplo típico de esto es cuando se enseña a sumar fracciones. En general, se enfatiza que para sumar de fracciones se debe encontrar el mínimo común múltiplo entre los denominadores, haciendo de la tarea algo netamente procedimental. El enlace con la suma de números enteros queda difuso y esta nueva tarea termina convirtiéndose en una combinación extraña de elementos. Hay una variedad de reportes que aseveran que gran parte de la fobia hacia

las matemáticas comienza cuando a los niños se les enseña a sumar fracciones.

Otro factor que entra en juego es el nivel de preparación que tienen los profesores de enseñanza primaria. Al parecer las dificultades para comprender las fracciones llegan incluso a los docentes. Existe una diversidad de afirmaciones erróneas en torno al concepto de fracción las cuales los profesores usan permanentemente, por ejemplo, que “simplificar una fracción es convertir la fracción en una más pequeña”, o “un medio no es lo mismo que dos cuartos”, o simplemente que “las fracciones no son números”. Además, parece ser que las habilidades procedimentales también son una dificultad en los docentes; Liping Ma detectó que tan solo un 43 % de los profesores norteamericanos de primaria llegan a la respuesta correcta en una tarea de división de fracciones [40].

Para hacer frente a todas estas problemáticas, en 2010 un comité de expertos, encabezados por Robert Siegler, dieron 5 recomendaciones básicas para dirigir el aprendizaje de fracciones durante la educación escolar [54]. Estas son:

1. Basarse en el entendimiento informal que tienen los alumnos de repartición y proporcionalidad para desarrollar de forma inicial el concepto de fracción.
2. Ayudar a los estudiantes a entender que las fracciones son números y que expanden el sistema de los números enteros. Usar la recta numérica como representación lo más tempranamente posible.
3. Ayudar a los estudiantes a entender por qué los procedimientos para hacer cálculos con fracciones tienen sentido.
4. Desarrollar en los estudiantes la comprensión conceptual sobre estrategias para resolver problemas de proporciones antes de presentar como método de resolución la multiplicación cruzada.
5. Los programas de desarrollo profesional deberían tomar un lugar prioritario para mejorar la comprensión por parte de los profesores de las fracciones y de cómo enseñarlas.

Sin duda, las recomendaciones anteriores son una buena guía para intentar mejorar la enseñanza de las fracciones, sin embargo, estas no implican que basta con seguirlas para solucionar el problema. La investigación en relación a este tema aún está en pleno desarrollo. Todavía hay aristas que no están del todo claras y varios puntos en donde no hay consenso. Lo que sí es claro es la importancia de este tema a nivel educacional.

4.1.2. La importancia del buen aprendizaje de las fracciones

En 2006 fue creado en los Estados Unidos el National Mathematics Advisory Panel, comité que tuvo como objetivo examinar y sintetizar la evidencia científica sobre la enseñanza y el aprendizaje de las matemáticas. El trabajo de este panel de expertos terminó en el año 2008 con la entrega de un informe, en donde se señala que uno de los problemas graves en la educación escolar es la enseñanza del álgebra [47]. Para intentar cambiar este escenario, se propuso que el aprendizaje de este contenido debe tener como base tres ejes fundamentales:

1. La fluidez con números enteros. Se debe desarrollar el sentido numérico, fomentar el conocimiento de propiedades, representaciones y habilidades de cálculo.

2. La fluidez con fracciones. Se debe enseñar a ubicar fracciones en la recta numérica, comparar fracciones y representarlas.
3. El conocimiento de aspectos particulares de Geometría y Medición. Entender la semejanza de triángulos es fundamental para la comprensión del concepto de pendiente. Se pueden usar las figuras para contextualizar expresiones algebraicas, como fórmulas para el área o perímetro.

Junto con esto, la influencia de las fracciones sobre el algebra es estudiada en una reciente investigación. A partir de datos de alumnos de Estados Unidos y Reino Unido, se revela que el nivel de conocimiento de fracciones y división en quinto grado predice el nivel de comprensión de algebra y matemática en general en la secundaria [55]. Esta deducción se hizo controlando estadísticamente por otros tipos de conocimiento matemático, la capacidad intelectual en general, la memoria de trabajo y el ingreso y educación familiar. A su vez es sabido que el desempeño en matemática en la secundaria es un predictor de la matrícula en universidades y del término de una carrera.

Las explicaciones van en dos corrientes. La primera considera que para las tareas asociadas al algebra y la matemática más avanzada, la operatoria con fracciones y la realización de divisiones son cruciales para un buen desempeño. Otra explicación considera que el trabajo con fracciones es por si misma una tarea que requiere mayor esfuerzo cognitivo. Sin embargo, se observa que la predicción es válida para la generalidad de los alumnos, tanto para aquellos con buen desempeño como los con mal desempeño en matemática.

Mejoras sustanciales en la enseñanza de fracciones a nivel básico podrían entonces repercutir de manera significativa en el desempeño logrado por los alumnos en niveles muy superiores.

4.1.3. Representaciones Privilegiadas y Mecanismos Ecológicamente Válidos

Aparentemente hay representaciones que son privilegiadas sobre otras en el dominio del razonamiento matemático. Esto serviría para explicar la causa de los errores comunes que se cometen en fracciones y el trabajo con números decimales. Como se dijo en el Capítulo 1 de este trabajo, en general hay acuerdo que los fundamentos del razonamiento matemático están basados en habilidades innatas. Al aceptar la existencia de representaciones privilegiadas se asume que el sujeto tiene un rol activo en el aprendizaje, pero que además existe una fuerte influencia del medio ambiente [4].

Una segunda implicancia de la aceptación de representaciones privilegiadas es el rechazo del supuesto de equipotencialidad de los conceptos. La equipotencialidad significa que no hay conceptos que a priori sean más fáciles o más difíciles que otros.

En general, la equipotencialidad no se asume como cierta; es claro que enseñar números enteros es mucho más fácil que enseñar trigonometría; sin embargo, cuando los docentes analizan un dominio conceptual en particular, por ejemplo las fracciones, generalmente se asume la equipotencialidad. De este modo, se considera que la dificultad en visualizar las

fracciones como razones, números racionales, divisiones o números decimales es la misma. En vista de lo que se ha expuesto, distintas investigaciones indicarían que el supuesto de equipotencialidad es erróneo, en particular para el razonamiento matemático.

Una idea similar es la desarrollada por Gigerenzer [68], quien plantea que la comprensión de determinados conceptos depende fuertemente del formato en que sean estudiados. En particular, Gigerenzer trabajó la comprensión del concepto de probabilidad, concluyendo que existe una diferencia sustancial en entender la probabilidad como una magnitud entre 0 y 1, por ejemplo que la probabilidad que ocurra un determinado evento es 0,1, y entenderla asociado a una tasa de frecuencias, por ejemplo que un evento ocurra en 10 de cada 100 casos. Este último formato sería muchos más comprensible y permitiría trabajar con probabilidades de mejor forma.

Aparentemente, el formato frecuentista de las probabilidades podría tener orígenes evolutivos, lo que explicaría su ventaja sobre el formato decimal. Tener una medida de ocurrencia de eventos dada la experiencia podría facilitar reconocer situaciones benéficas o riesgosas para el individuo. De este modo, los seres humanos estaríamos ecológicamente adaptados para comprender las probabilidades como frecuencias. Este y otros hallazgos motivan a Gigerenzer a hablar de mecanismos ecológicamente válidos, siendo aquellos mecanismos cognitivos que podrían deber su origen a la evolución.

Desde este punto de vista, valdría la pena observar qué tipo de conductas están presentes en animales que están bajo los seres humanos en la escala evolutiva y que a la vez intenten ser base para la situación aquí estudiada, es decir, el aprendizaje de fracciones.

Según algunos estudios, la capacidad de intercambiar no es propia de los seres humanos. Experimentos conductuales con monos indican que éstos podrían negociar. El tipo de intercambio entre monos se produce entre bienes y servicios, y se ajusta de acuerdo a la escasez de los bienes; los monos subordinados acicalan más a los monos dominantes con el fin de tener más comida en períodos cuando ésta es escasa [62] [61], las hembras acicalan más a una hembra que recién ha sido madre para que éste permita a las otras tomar en brazos al bebé en tiempos en que la tasa de natalidad es baja [27] [28], los machos tienden a acicalar más a las hembras fértiles cuando es período de apareamiento [26]. Los primates tienen mecanismos de ajuste de sus tasas de intercambios muy sensibles, capaces de adaptarse a cambios en las situaciones de su tribu. A este fenómeno suele denominarse Mercado Ecológico.

Tener en mente estos hallazgos podrían motivar la búsqueda de una representación privilegiada asociada a las fracciones, en la línea de tasas de frecuencias o de intercambio de objetos.

4.1.4. Otros estudios sobre fracciones

El problema del aprendizaje de fracciones no es tema de investigación nuevo. Existe una diversidad de estudios que intentan determinar cuál es el origen de las dificultades en su aprendizaje y cuál es la mejor manera de enseñarlas. Algunas investigaciones señalan que la enseñanza de fracciones debe estar orientada principalmente por el concepto de medida. Así,

lo fundamental para trabajar con fracciones es la definición de una unidad de medida, lo que permitirá dar significado a las fracciones [65].

Otros proponen que una buena manera de abordar este concepto es mediante la idea de intensidad. Así, una fracción se debe asociar a las nociones de velocidad, presión o densidad, entre otras [29]. De esta manera, los investigadores apelan al sentido de proporcionalidad que tendrían los estudiantes para enseñar las fracciones. Además, se indica que existirían diferencias en desempeño cuando se construyen intensidades con distintos tipos de cantidades. Cuando la intensidad se construye con cantidades continuas, como distancia y tiempo, los resultados serían mejores que cuando las intensidades se construyen con cantidades discretas.

También hay estudios que indican que existe una estrecha relación entre el aprendizaje de fracciones con las capacidades inhibitorias que tienen los alumnos. Esto se plantea ya que se considera que la dificultad asociada a las fracciones está ligada al sesgo hacia los números enteros [43] [44]. En vista que los niños comienzan, en general, el aprendizaje de fracciones una vez que han trabajado de manera extensa sobre los números enteros, las dificultades en fracciones resultan de no reconocer la diferencia entre los dos sistemas numéricos, por lo que aplican equivocadamente conceptos y algoritmos propios de los números enteros.

Sin embargo, este no es un tema totalmente consensuado. Algunos investigadores consideran que el problema surge de establecer analogías impresas sobre los números enteros. De este modo, los números enteros deben ser vistos como parte de un sistema mayor, asociado a la recta numérica mental [56].

Para entender en profundidad de qué manera se representan las fracciones en la recta numérica mental, la principal tarea que se ha ocupado en distintos estudios es la de comparación. A los estudiantes se les presentan dos fracciones y se les pide determinar cuál es la mayor. Las conclusiones obtenidas apuntan en distintas direcciones; algunos consideran que al comparar fracciones, estas pueden ser vistas de manera holística [22] [32] [53], mientras que otros consideran que para determinado tipo de fracciones que se comparan se utilizan determinadas estrategias, así por ejemplo, cuando se comparan dos fracciones con igual numerador, se suele utilizar la comparación de números enteros. Otras estrategias consisten en transformar la fracción a número decimal y luego comparar.

Lo cierto es que cuando los niños comienzan el aprendizaje de fracciones, en la tarea de comparar suele manifestarse el sesgo hacia los números enteros, es decir, se considera como fracción mayor aquella que contiene en su escritura el número entero mayor. Con el fin de intentar resolver esta dificultad, Wu recomienda usar la recta numérica para comparar fracciones [66]. Si se ocupa la representación de la figura 4.2, esta puede causar serios inconvenientes. De ahí que, en caso de ocupar esta representación, se debe resaltar permanentemente que la unidad es la misma en ambas fracciones que se comparan.

Wu también recomienda trabajar con fracciones unitarias (numerador igual a 1). Luego, comparar fracciones con igual numerador es simplemente replicar una fracción unitaria tantas veces como el numerador, así por ejemplo, $\frac{5}{6}$ se puede entender como 5 veces $\frac{1}{6}$. Además, tener números de referencia, como el 1 o $\frac{1}{2}$ podría facilitar la representación de fracciones en la recta numérica.

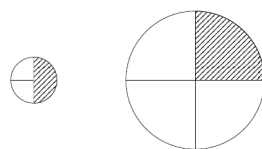


Figura 4.2: Problema en la Representación Geométrica de Fracciones. Si no se insiste en que para comparar fracciones se necesita representarlas con la misma unidad, se podrían llegar a conclusiones erradas.

4.2. Presentación del Experimento

4.2.1. Objetivo

Se pretendió con este estudio comparar el impacto de tres métodos de enseñanza en estudiantes que han iniciado el aprendizaje de fracciones. La enseñanza en los tres casos estuvo orientada a que los alumnos aprendieran a comparar fracciones.

A cada uno de los estudiantes participantes de este estudio se les enseñó sólo uno de los siguientes tres formatos de representación de fracciones:

- Fraccionar. Fracción como parte de un cuadrado. Por ejemplo, $\frac{1}{2}$ se representa como la mitad de un cuadrado.
- Temporal. La fracción asociada a la entrega de objetos cada cierta cantidad de días. Por ejemplo, $\frac{1}{2}$ equivale a recibir un dulce cada dos días.
- Intercambio. La fracción asociada a una situación de intercambio de objetos. Por ejemplo, $\frac{1}{2}$ es igual a recibir un objeto por cada dos monedas que doy.

En base a la literatura y experimentos realizados anteriormente, se ha detectado una tipología en las preguntas de comparación de fracciones. En las preguntas en que se presentan dos fracciones y se le pide al estudiante identificar la mayor, se encuentran dos grupos de preguntas:

- Congruentes: son aquellas preguntas cuya solución coincide con la fracción que tiene al mayor de los cuatro números enteros que aparecen en la tarea; dos numerados y dos denominadores.
- Incongruentes: son las no congruentes.

La evidencia ha mostrado que las preguntas congruentes tienen mayor tasa de acierto que las incongruentes. Esto es consistente con los hallazgos relacionados con el sesgo hacia los números enteros.

Un subconjunto de las preguntas incongruentes son las denominadas “Incongruentes Difíciles”. Estas son aquellas preguntas incongruentes en que el segundo número mayor de los cuatro se encuentra en la fracción menor. Se denominará “Incongruentes Simples”, a las preguntas incongruentes que no son difíciles.

Parte de este estudio tiene como propósito analizar el impacto de los distintos formatos en los distintos tipos de preguntas.

4.2.2. Participantes

En total, participaron 213 estudiantes, 109 mujeres y 104 hombres, todos cursando el cuarto año de enseñanza básica. La edad promedio de los alumnos fue cercana a los 10 años. Los alumnos participantes pertenecían a dos colegios de la región metropolitana, ambos ubicados en sectores vulnerables; un colegio pertenece a la comuna de Pedro Aguirre Cerda y el otro a la comuna de La Pintana. Estos alumnos estaban comenzando su aprendizaje formal en el área de fracciones.

Para comparar los métodos de enseñanza se separaron de forma aleatoria en tres grupos, los cuales se denominaron Fraccionar, Temporal e Intercambio.

4.2.3. Materiales

Para realizar este estudio, los alumnos asistieron al laboratorio de computación de sus colegios. Cada alumno ocupó un computador de manera personal, en el que se trabajó con el programa SAGDE (<http://sagde.metaforas.cl/>), el cuál permite crear preguntas y registrar las respuestas de los alumnos. Además, cada computador contaba con audífonos.

En parte de este estudio los alumnos debieron contestar preguntas, todas ellas de alternativas. Para contestar, los alumnos tuvieron conectar un punto, que se ubica debajo del enunciado, con la respuesta correcta. Es importante notar que los alumnos estaban familiarizados con el programa, ya que a la fecha, hacían uso de él una vez a la semana.

La realización de la tarea duró entre 30 y 45 minutos.

4.2.4. Descripción del Test

Este estudio consideró un diseño experimental de tres etapas. La primera etapa corresponde a un Pre Test común para todos los alumnos. La segunda etapa es la denominada etapa de Entrenamiento, en donde se separaron a los alumnos de manera aleatoria en los tres grupos señalados. Finalmente, en la última etapa se realizó un Post Test común a todos los estudiantes. Un esquema del experimento se presenta en la figura 4.3. Especificaciones de cada etapa se encontrarán en las siguientes secciones.

Pre Test

Todos los alumnos, con independencia al grupo definido, contestaron 18 preguntas al iniciar el test. Estas preguntas se presentaron en un orden definido con antelación. Una vez

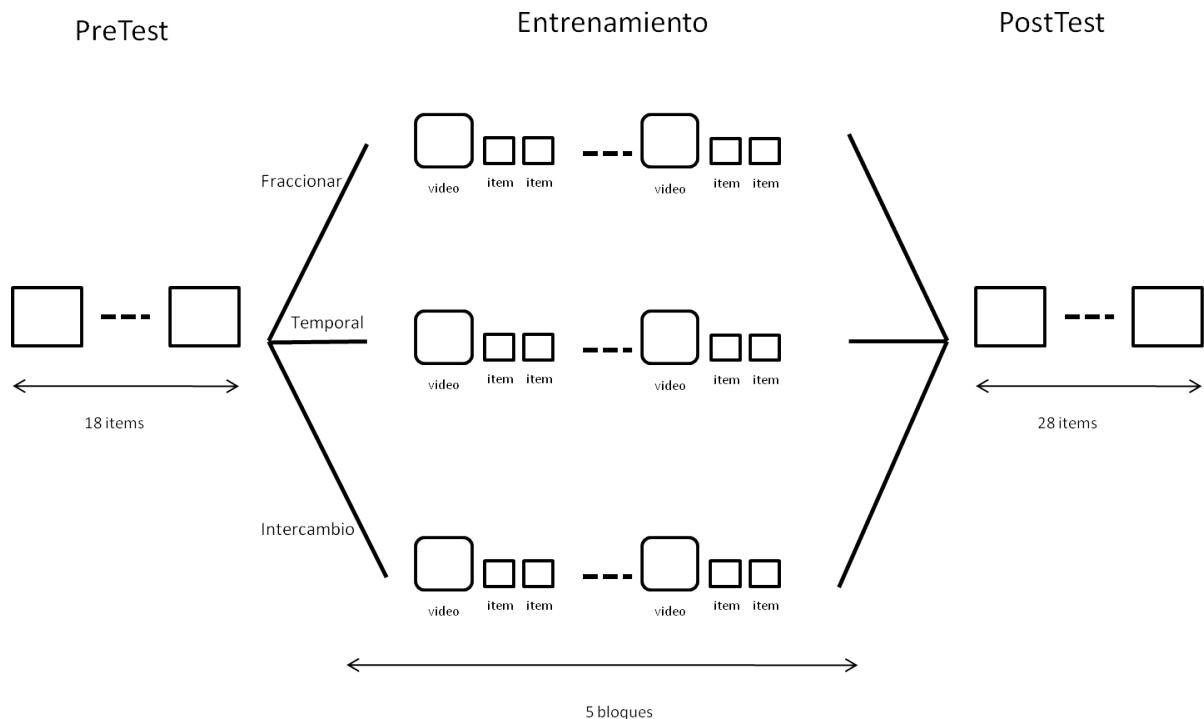


Figura 4.3: Esquema del Diseño Experimental

que el alumno respondió una pregunta, pasó a la siguiente sin recibir feedback y sin tener la posibilidad de cambiar las respuestas ya emitidas.

Las seis primeras preguntas pidieron identificar cuál es la mayor fracción de dos que se presentaron escritas en formato simbólico. Dos de estas preguntas fueron congruentes, dos incongruentes simples y dos incongruentes difíciles. Todas estas preguntas tuvieron como enunciado: “Conecta el mayor de los dos”.

Las seis preguntas siguientes tuvieron como objetivo identificar la habilidad de los alumnos de utilizar las representaciones para comparar magnitudes utilizando el formato temporal, pero sin hacer uso de la notación simbólica de fracción. El enunciado de estas preguntas fue: “La profesora da dulces a sus alumnos. Conecta a quien prefiere”. Las alternativas fueron dos, y eran de la forma “Juan: Recibe x dulces cada y días”. Los números utilizados en las alternativas de estas preguntas fueron los mismos que se utilizaron en las primeras series preguntas del Pre Test, por lo que por cada pregunta en formato simbólico existe una pregunta en formato temporal. Obviamente, esto hace que hayan dos preguntas congruentes, dos incongruentes simples y dos incongruentes difíciles.

Las seis últimas tienen el mismo objetivo anterior pero en esta parte se utilizó la metáfora de intercambio, también sin hacer uso de la notación simbólica de fracción. El enunciado fue: “¿Qué máquina me conviene más?”. Las alternativas fueron del estilo “Máquina A: me entrega x tickets por cada y fichas que pongo”. Al igual que en el caso anterior, los números ocupados en las alternativas fueron los mismos que en los casos anteriores. La correspondencia entre cada una de estas preguntas se observa en la figura 4.4.

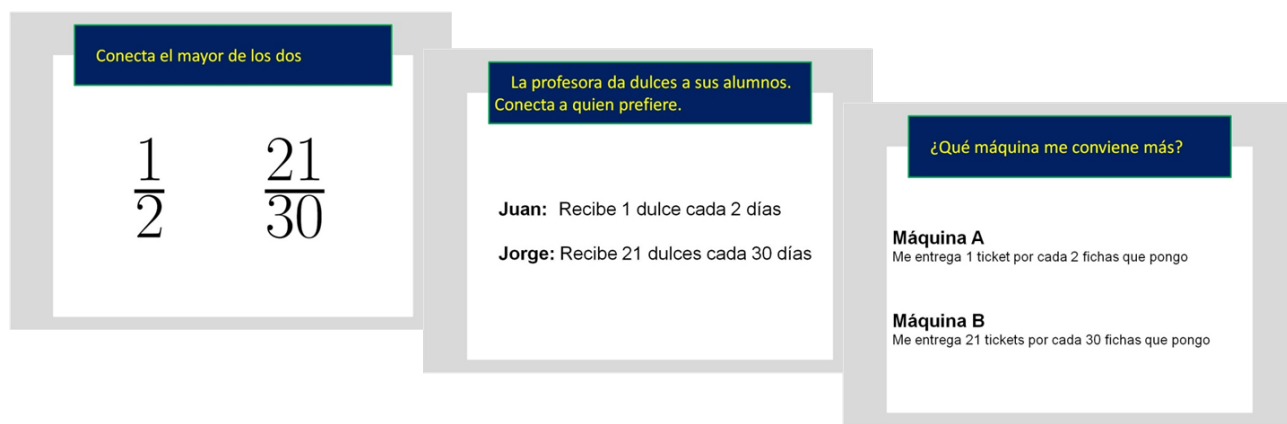


Figura 4.4: Ejemplo de Preguntas del Pre Test. En el Pre Test hubo tres tipos de preguntas; aquellas en formato simbólico, en donde se les pedía identificar la mayor de dos fracciones escritas simbólicamente, aquellas en formato temporal, en donde se les pedía identificar qué alumno terminaba siendo más favorecido, y finalmente aquellas en formato de intercambio, en donde los alumnos debían identificar qué situación era más conveniente. Para cada pregunta en formato simbólico, se planteó su símil en formato temporal y de intercambio utilizando los mismo números enteros, como se indica en la figura.

Entrenamiento

En esta etapa, las actividades hechas por los alumnos dependieron del grupo al que fueron asignados. El entrenamiento consistió en 5 bloques, en donde cada uno contó de un video y dos preguntas. Tanto los videos como las preguntas a responder dependieron del grupo que el alumno perteneció. Todos los alumnos de un mismo curso realizaron esta etapa de manera simultánea, y ninguno sabía que habían tres posibles entrenamientos.

En los primeros cuatro videos se presentó una fracción específica y el modo de representarla. En los videos del grupo Fraccionar, las fracciones fueron enseñadas como parte de un cuadrado. Así, representar $\frac{1}{2}$ equivale a marcar la mitad de un cuadrado. A los alumnos del grupo Temporal se les enseñó fracciones asociándolas a la entrega de dulces cada cierta cantidad de días. Así $\frac{1}{2}$ equivale a recibir 1 dulce cada 2 días. Por último, en los videos del Grupo Intercambio, se les enseñó a los alumnos asociar las fracciones a una situación en donde recibían cierta cantidad de tickets por cada cierta cantidad de fichas que entregan, así representar $\frac{1}{2}$ corresponde a recibir 1 ticket por cada 2 fichas que entrego.

Después de cada uno de estos videos, se les pidió a los estudiantes contestar dos preguntas. En la primera se pidió identificar cuál es la representación que se ajusta a una fracción dada, la cuál siempre coincide con la fracción enseñada en el video. En la segunda pregunta se mostró una representación y se pidió identificar en un grupo de alternativas cuál es la fracción que se ajusta a ésta; la respuesta correcta coincide con la fracción enseñada en el video.

En el último video, se enseñó a los alumnos a comparar fracciones ocupando las representaciones correspondientes. Para el grupo Fraccionar la comparación de fracciones se basó en la comparación de áreas sombreadas de cuadrados de igual tamaño. Para el grupo Temporal e Intercambio la comparación de fracciones quedó sujeta a encontrar la situación más con-

veniente. Así, identificar el mayor entre $\frac{1}{2}$ y $\frac{1}{4}$ equivale a indentificar si es más conveniente tener un dulce cada dos días o un dulce cada cuatro días, en el caso del grupo Temporal, o un ticket por cada dos fichas que pongo o un ticket por cada cuatro fichas que pongo, en el caso del grupo Intercambio. Luego del video, se les hicieron a los alumnos dos preguntas, en cada una se les pedía identificar la fracción mayor, siendo una pregunta congruente y la otra incongruente.

En esta etapa los alumnos recibieron feedback y podían modificar su respuestas en caso de responder de manera incorrecta. Sólo pudieron avanzar en el test si encontraron la pregunta de forma correcta.

La tabla 4.1 muestra las fracciones utilizadas en cada video y sus duraciones según cada grupo.

Bloque	Fracción enseñada	Video Fraccionar	Video Temporal	Video Intercambio
1	$\frac{1}{2}$	1:17	1:23	1:08
2	$\frac{1}{4}$	1:35	1:23	1:04
3	$\frac{3}{8}$	2:20	1:50	1:54
4	$\frac{2}{9}$	1:26	1:17	1:15
5	$\frac{1}{2}$ v/s $\frac{1}{4}$ $\frac{2}{8}$ v/s $\frac{1}{9}$	1:33	1:39	1:42

Tabla 4.1: Tiempo y Contenido de Videos de la Etapa de Entrenamiento

Es importante notar que en la etapa de aprendizaje no se hizo mención a los nombres propios de las fracciones, así por ejemplo, cuando se presentó $\frac{1}{2}$ no se hizo referencia a “un medio”, si no se indicaba que ese símbolo tiene un determinado significado, de acuerdo al tipo de instrucción. Más aún, en ningún video se mencionó la palabra “fracción”.

Post Test

En esta etapa se deseaba estudiar la habilidad de comparar fracciones. Se eligieron 8 fracciones, de modo que cada pregunta se genera comparando dos de estas. Así se tiene un total de 28 preguntas. Las fracciones escogidas fueron:

$$\frac{4}{40} \quad \frac{6}{20} \quad \frac{1}{2} \quad \frac{21}{30} \quad \frac{9}{10} \quad \frac{6}{5} \quad \frac{3}{2} \quad \frac{7}{4}$$

En todas las preguntas se pidió identificar la fracción mayor, ocupando siempre el enunciado “Conecta el mayor de los dos”. Nótese que la distancia entre todas las fracciones es mayor a 0,2.

En esta etapa se dió prioridad a preguntas incongruentes. Así, del total de 28 preguntas generadas, se tiene que hubo 8 congruentes y 20 incongruentes. Las preguntas según su clasificación se muestran en la tabla 4.2, donde C identifica a las preguntas Congruentes, IS a las Incongruentes Simples e ID a las Incongruentes Difíciles.

	$\frac{4}{40}$	$\frac{6}{20}$	$\frac{1}{2}$	$\frac{21}{30}$	$\frac{9}{10}$	$\frac{6}{5}$	$\frac{3}{2}$	$\frac{7}{4}$
$\frac{4}{40}$	-	IS	ID	IS	IS	IS	ID	IS
$\frac{6}{20}$		-	ID	C	IS	IS	ID	IS
$\frac{1}{2}$			-	C	C	C	C	C
$\frac{21}{30}$				-	ID	ID	ID	ID
$\frac{9}{10}$					-	ID	ID	ID
$\frac{6}{5}$						-	ID	C
$\frac{3}{2}$							-	C
$\frac{7}{4}$								-

Tabla 4.2: Clasificación de Preguntas

Las 28 preguntas fueron presentadas a cada estudiante de manera aleatoria. Cabe notar que no se presentaron todos los ordenes posibles de las preguntas, por ejemplo, se mostró la pregunta en donde $\frac{7}{4}$ se ubicaba a la izquierda y $\frac{6}{20}$ a derecha, y no aquella en donde $\frac{7}{4}$ se ubica a la derecha y $\frac{6}{20}$ a la izquierda. Sin embargo, se procuró que en el caso de las preguntas Congruentes la mitad de las respuestas correctas se ubicaran a la izquierda y la otra mitad a la derecha. Lo mismo se hizo para las preguntas Incongruentes Simples y para las Incongruentes Difíciles. Esto se hizo pensando que en caso contrario se tendría un Post Test de 56 preguntas, lo que se consideró como demasiado extenso para niños de cuarto básico.

Otro punto importante, es el hecho que las primeras seis preguntas del Pre Test coinciden con algunas preguntas del Post Test, lo que permitió medir el aprendizaje alcanzado por los estudiantes.

4.3. Principales Resultados

Cada pregunta del Pre Test puede ser clasificada de acuerdo a dos factores, el tipo de pregunta, es decir, si fue formulada usando fracciones en formato Simbólico, formato Temporal o formato Intercambio, y según la congruencia, en Congruente, Incongruente Simple o Incongruente Difícil. Esto hace que las preguntas se puedan clasificar en 9 tipos, combinando los dos factores mencionados.

Si se toma las preguntas en formato Simbólico, es decir aquellas en donde aparecían los símbolos fraccionarios, comparando por tipo de congruencia se tiene las preguntas congruentes

tienen un desempeño superior al azar, no siendo así para las incongruentes simples y difíciles. Junto con esto, se tiene que las preguntas Incongruentes Simples tiene una mejor correctitud que las preguntas Incongruentes Difíciles ($p\text{-valor} < 0,001$).

Si se consideran las preguntas en formato Temporal, se tiene que las preguntas Congruentes e Incongruentes Simples tienen una correctitud sobre el azar, pero no las Incongruentes Difíciles. Además, se tiene que las preguntas Congruentes tienen asociado un mejor desempeño que las preguntas Incongruentes Simples ($p\text{-valor} < 0,001$), y estas últimas un mejor desempeño que las Incongruentes Difíciles ($p\text{-valor} < 0,001$).

Una situación similar se observa para las preguntas en formato Intercambio. De igual manera al caso anterior, las preguntas Congruentes e Incongruentes Simples tienen una correctitud sobre el azar, pero no las Incongruentes Difíciles. Sin embargo, en este caso no se puede decir que las preguntas Congruentes tienen un mejor desempeño que las preguntas Incongruentes Simples ($p\text{-valor} > 0,05$). Si observa además que en las Incongruentes Simples se obtiene un mejor desempeño que en las Incongruentes Difíciles ($p\text{-valor} < 0,001$).

Un resultado interesante resulta al comparar el tipo de preguntas a un mismo nivel de congruencia. En el caso de las preguntas Congruentes, en todos los formatos, como se mencionó anteriormente, se tiene una correctitud superior al azar. Sin embargo, las preguntas en formato Temporal tienen una correctitud mayor que las preguntas en formato Simbólico e Intercambio ($p\text{-valor} < 0,05$). Para el caso de las Incongruentes Simples, sólo las preguntas en formato Temporal e Intercambio superan al azar, sin encontrar diferencias estadísticamente significativas en estos dos formatos ($p\text{-valor} > 0,1$), mientras que las preguntas en formato Simbólico no superan el 30 % de correctitud. Este hallazgo es muy importante, ya que indicaría que parte de la dificultad que tienen los alumnos para comparar fracciones se puede resolver apelando a la representación adecuada. Finalmente, para el caso de las preguntas Incongruentes Difíciles, si bien todos los formatos tienen desempeños inferiores al azar, se tiene que el formato Intercambio tiene un mejor resultado que el Temporal y el Simbólico ($p\text{-valor} < 0,001$), y el formato Temporal mejor al formato Simbólico ($p\text{-valor} < 0,001$). Toda esta información se encuentra sintetizada en el gráfico de la figura 4.5.

Para el caso del Post Test, se calculó por separado el desempeño obtenido por cada alumno en las preguntas Congruentes, Incongruentes Simples e Incongruentes Difíciles. Dado que en este escenario el entrenamiento es un factor clave, se realizó un test ANOVA para estudiar la existencia de diferencias en cada uno de los promedios indicados. Se encontró que en las preguntas Congruentes e Incongruentes Difíciles existe una diferencia estadísticamente significativa entre los distintos tipos de entrenamiento ($p\text{-valor} < 0,02$ en ambos casos), no así para las preguntas Incongruentes Simples ($p\text{-valor} > 0,5$). Haciendo un análisis Post-Hoc, se tiene que en el caso de las preguntas Congruentes el grupo Fraccionar tiene un desempeño superior al Grupo Temporal ($p\text{-valor} < 0,05$), no registrándose ninguna otra diferencia importante. Para el caso de las preguntas Incongruentes Difíciles, se tiene que el desempeño del Grupo Temporal es superior al Grupo Fraccionar. No se puede declarar que el Grupo Intercambio es, términos estadísticos, distinto a los otros dos grupos. Esta situación se resume en el gráfico de la figura 4.6.

Por otra parte, se quiso estudiar si hubo un aumento de desempeño después del entrenamiento. Para ello, se tomaron los resultados de las preguntas en formato Simbólico del Pre

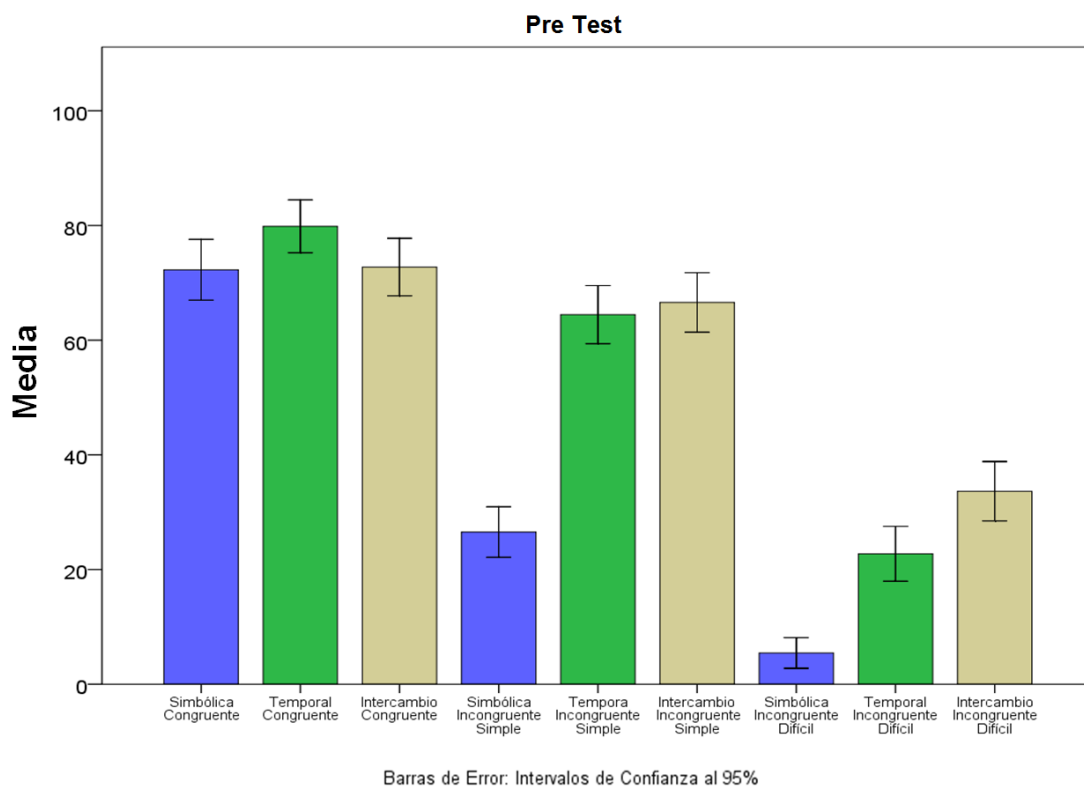


Figura 4.5: Resultados Pre Test

Test y se compararon con su desempeño en el Post Test. Recuérdese que todas las preguntas en formato Simbólico del Pre Test se repitieron en el Post Test. Así, para cada pregunta se calculó la diferencia entre el puntaje obtenido en el Post Test con obtenido en el Pre Test, luego se promediaron las preguntas que tenían el mismo tipo de congruencia.

Inferir si un determinado grupo aprendió se traduce en determinar si el promedio de las diferencias de desempeño es mayor que 0 en términos estadísticos. En el caso de las preguntas Congruentes esto no ocurre para ningún grupo ($p\text{-valor} > 0,05$ en todos los casos). Sin embargo, en las preguntas Incongruentes Simples y Difíciles se encontró que en todos los grupos la diferencia de desempeño es superior a cero ($p\text{-valor} < 0,05$ en todos los casos). Además, se encontró que existe una diferencia en la contribución que hacen los métodos de enseñanza. El Grupo Temporal tiene más de 14 puntos de diferencia sobre el Grupo Fraccionar ($p\text{-valor} < 0,05$). La figura 4.7 hace síntesis de estos hallazgos. La diferencia entre el grupo Intercambio y Fraccionar en este caso no es significativa.

4.3.1. Primeras Conclusiones del Estudio

De los análisis estadísticos hasta aquí expuestos se puede decir varias cosas. En primer lugar, de los resultados del Pre Test se concluye el evidente sesgo hacia los números enteros en la tarea de comparar fracciones en lenguaje simbólico, pues en general, se considera la fracción mayor a aquella que posee números grandes. Sin embargo, cuando se lleva a los estudiantes a una situación en donde deben identificar la situación más beneficiosa, el sesgo

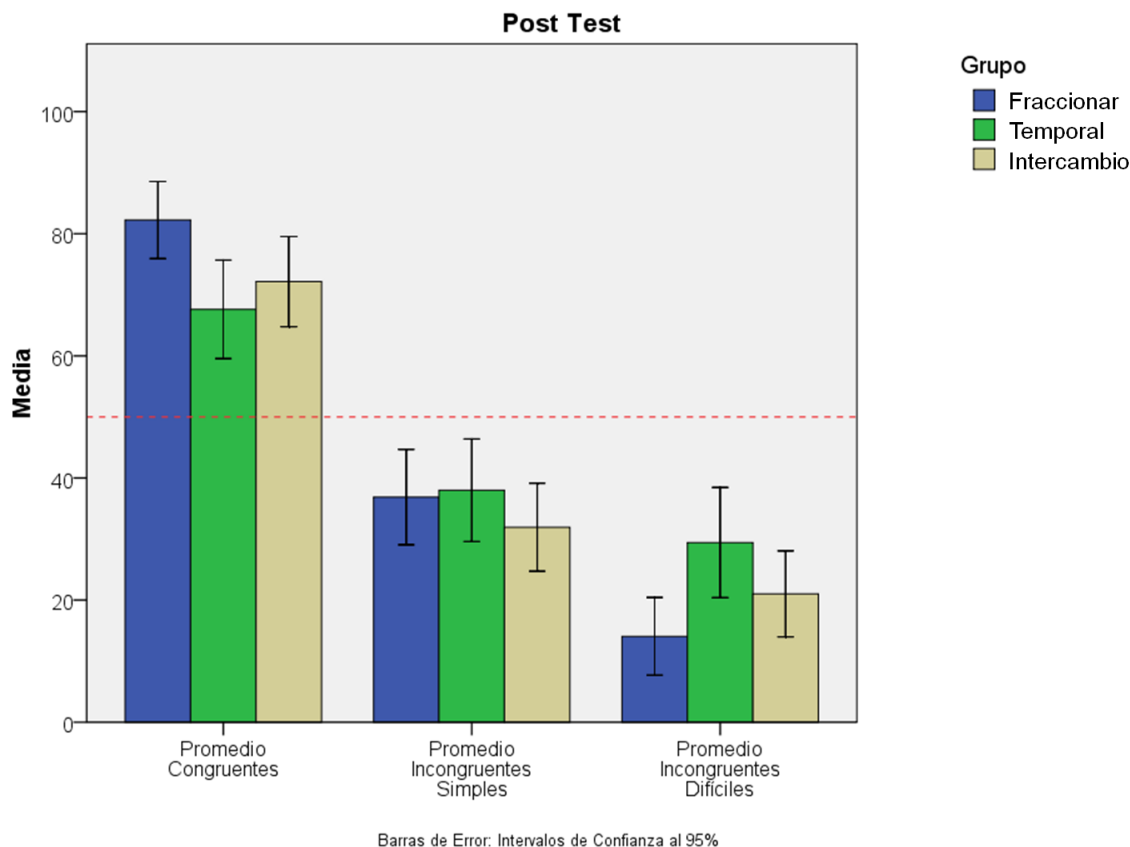


Figura 4.6: Resultados PostTest

tiende a disminuir. Esto se observa en los resultados del Pre Test en las preguntas en formato Temporal e Intercambio. Esto hace pensar que para tener éxito en la tarea de comparación de fracciones podría ser útil representarlas en uno de estos dos formatos.

Por otra parte, los resultados del Post Test indican que si bien las diferencias de desempeño entre las preguntas Congruentes, Incongruentes Simples e Incongruentes Difíciles, también es cierto que hay un aumento en el desempeño con respecto al Pre Test. Al momento de comparar los tres métodos de enseñanza se observa que el que tiene mejores desempeños en las preguntas Incongruentes es el método Temporal, siguiéndole el método de Intercambio.

Hasta acá, se podría decir que el método Temporal aporta más que el método tradicional de enseñanza (en términos estadísticos). Sin embargo, estos análisis no logran capturar el tipo de estrategia que podrían estar utilizando los alumnos al momento de contestar las preguntas. La siguiente sección intenta hacerse cargo de este problema, proponiéndose una serie de métodos para lograr identificar qué tipo de estrategias utilizarían los estudiantes.

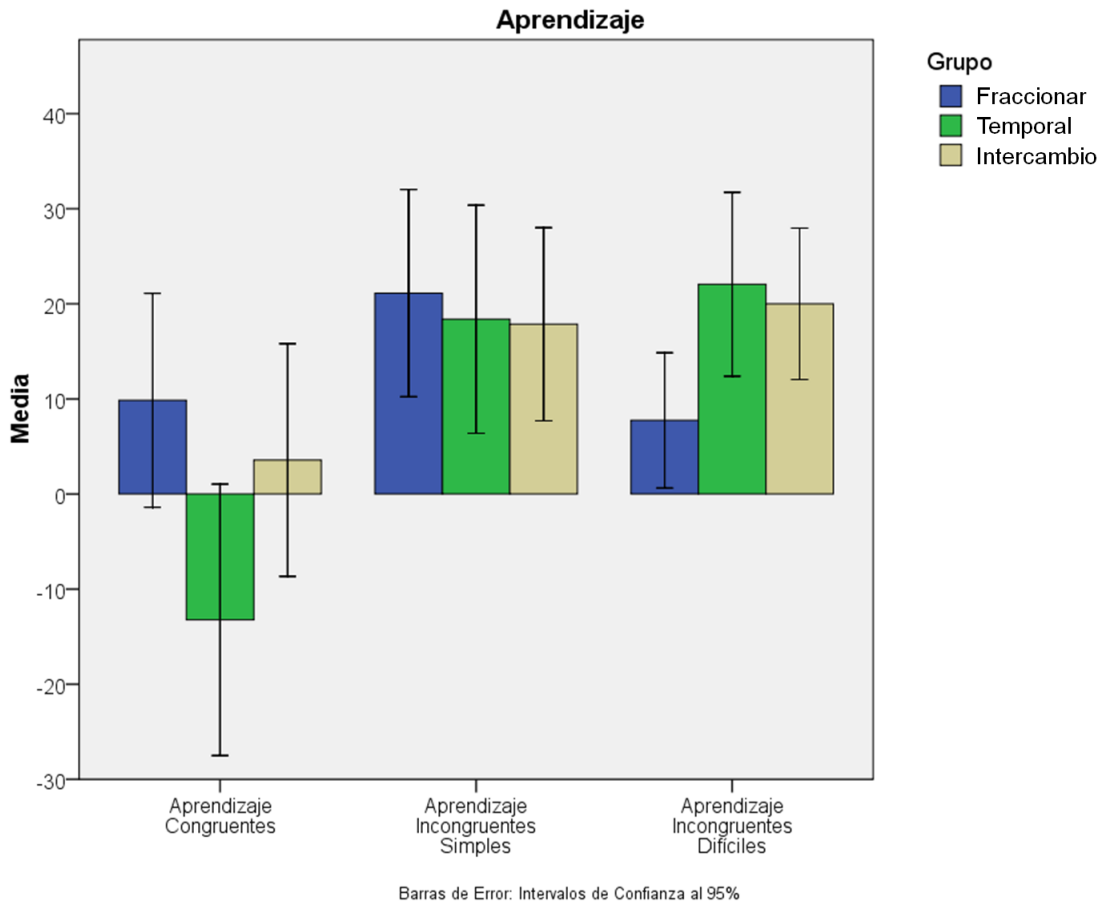


Figura 4.7: Nivel de Aprendizaje entre PreTest y PostTest

4.4. Escalamiento de Fracciones

Una pregunta recurrente en la tarea de comparación de magnitudes es si los niños tienen una estrategia con la cual ordenan las fracciones. Dicha estrategia debería verse manifiesta en las respuestas entregadas por los estudiantes al momento de decidir cuál es la mayor de las dos fracciones en cada pregunta del test.

La manera usual de abordar esta pregunta por parte de los investigadores es proponiendo diversas estrategias y comparar las respuestas que arrojaría la estrategia con las respuestas dadas por los alumnos. Este procedimiento culmina entregando un perfil de estrategia, indicando cuáles son las dominantes y en qué proporción se encuentran. Sin embargo, este modo de proceder tiene una dificultad intrínseca. Si se testean dos estrategia que son muy similares en sus respuestas, el método indicará que ambas estrategias son igualmente ocupadas por los estudiantes. Dado que en principio podrían existir infinitas estrategias a testear, esta metodología podría no ser tan útil, ya que impediría seleccionar con claridad una estrategia dominante.

En este estudio se tomó un enfoque diferente. En vez de testear distintas estrategias que podrían haber ocupado los alumnos, se procederá a ordenar las fracciones ocupadas en

el experimento a partir de un criterio dado y utilizando las respuestas de los alumnos. El objetivo es entonces, suponiendo que existe una estrategia con la que los niños ordenan las fracciones, encontrar dicho orden a partir de las respuestas y luego, inferir una estrategia. Esta estrategia debe ser coherente con el orden encontrado, ya que este último es el que debe ser la base para cualquier deducción y no al revés.

En los términos utilizados en el capítulo 2, lo que se quiso hacer es encontrar un escalamiento de las fracciones ocupadas en el experimento. El diseño experimental ocupado permite realizar este trabajo con los datos obtenidos del Post Test, ya que al seleccionar las 8 fracciones y preguntar por todos los pares que se pueden generar, se tiene la información precisa de cómo resulta ser la comparación entre cada uno de los estímulos. Para lograr el objetivo, a lo largo del desarrollo de este estudio, se plantearon 3 metodologías. A continuación se hace mención de cada una de ellas, destacando sus beneficios y desventajas. Es importante resaltar que las metodologías aquí expuestas son enfocadas hacia la realización un análisis global de los datos, y no pretenden, en principio, ser herramientas para hacer estudios individuales de los alumnos testeados.

4.4.1. Método de Escalamiento Multidimensional

Para simplificar la notación, a cada fracción se le asoció un número de 1 a 8, de forma correlativa, según el orden ascendente natural de las fracciones utilizadas.

Un supuesto que se hará sobre la tarea, es que ésta se ajusta a un modelo B.T.L. visto en el capítulo 2. Así, la probabilidad de declarar j mayor que i está dada por

$$P_{j,i} = \frac{v_j}{v_j + v_i} = \frac{\exp(\theta_j - \theta_i)}{1 + \exp(\theta_j - \theta_i)}$$

donde $\theta_j = \log(v_j)$. De este modo, si se define Z_{ji}^k como la respuesta del alumno k a la pregunta en que se le pide identificar si j es mayor que i , donde Z_{ji}^k es 1 si declaró que j es mayor y 0 si no, entonces, por la ley de los grandes números, se debería tener que

$$\overline{Z_{ji}} = \frac{\sum_{k=1}^N Z_{ji}^k}{N} \approx f(\theta_j - \theta_i)$$

donde $f(x) = \frac{\exp(x)}{1 + \exp(x)}$ y N la cantidad total de alumnos. Con esto, se podría inferir que

$$|f^{-1}(\overline{Z_{ji}})| \approx |\theta_j - \theta_i|$$

obteniendo de esta manera una matriz de distancias. Luego, lo que queda es aplicar el método de escalamiento multidimensional expuesto en el capítulo 2.

Una primera dificultad de este método surge cuando se plantea el escenario en donde todos contestan correctamente las preguntas. Si bien, puede ser poco probable, es un escenario

posible, sobre todo si se establece que existe alguna estrategia que utilizan los alumnos para contestar. Cuando lo anterior ocurre, se tiene

$$|f^{-1}(\overline{Z_{ji}})| = \infty \approx |\theta_j - \theta_i|$$

lo que impide hacer cualquier tipo de cálculo. Algo similar ocurre cuando todas las respuestas son incorrectas. Una manera de resolver esto es suavizando la estimación de las probabilidades, de modo que

$$\overline{Z_{ji}} = \frac{\sum_{k=1}^N (Z_{ji}^k) + 1}{N + 2} \approx f(\theta_j - \theta_i)$$

Con esto, siempre la estimación de las probabilidades será menor estricto que 1 y mayor estricto que 0, por lo que se podrá calcular la inversa de la función f , y proceder con el método de escalamiento multidimensional. Sin embargo, un nuevo problema aparece. Se debería esperar que si se tiene sólo respuestas correctas, el escalamiento de las fracciones sea el correcto. Esto no ocurre con esta metodología, lo que observa en la figura 4.8. Una de las razones de esto es la aplicación de la función valor absoluto, ya que al utilizarla se pierde la estructura de orden de los datos. Además, en el escenario en que todas las respuestas son correctas, se tiene que la distancia entre todos los estímulos es la misma, lo que implica que el buen espacio para representarlas debe tener dimensión 7 (la cantidad de estímulos menos 1), por lo que un espacio unidimensional no tiene poder representativo.

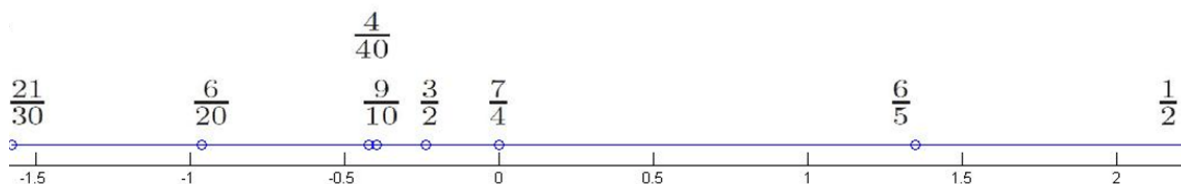


Figura 4.8: Problema del Método de Escalamiento Multidimensional

Este hecho hace de este método una herramienta poco útil, sin embargo, sus características iniciales servirán para dar pie al método de máxima verosimilitud.

4.4.2. Método de Máxima Verosimilitud

Siguiendo el supuesto sobre P_{ji} utilizado en el método anterior, en esta parte se asumirá que las respuestas entregadas por los alumnos al momento de comparar j con i , son realizaciones independientes de la variable aleatoria

$$Z_{ji} \sim \text{Bernoulli}(f[\theta_j - \theta_i])$$

con $f(x) = \frac{\exp(x)}{1+\exp(x)}$. La estructura de este método es similar a la utilizada en la Teoría de Respuesta al Ítem, en donde también se utiliza el supuesto de independencia entre las respuestas entregadas en un test [9].

Considerando, al igual que en el caso anterior, que $Z_{ji}^k = 1$ cuando el alumno k declara que j es mayor que i , y $Z_{ji}^k = 0$ si no, entonces, la verosimilitud de una muestra de tamaño N está dada por

$$\prod_{k=1}^N \prod_{i < j} f(\theta_j - \theta_i)^{Z_{ji}^k} \cdot [1 - f(\theta_j - \theta_i)]^{1 - Z_{ji}^k}$$

tomando logaritmo queda

$$\sum_{k=1}^N \sum_{i < j} Z_{ji}^k \cdot \log(f(\theta_j - \theta_i)) + (1 - Z_{ji}^k) \log(1 - f(\theta_j - \theta_i))$$

lo que es igual a

$$N \cdot \sum_{i < j} \bar{Z}_{ji} \cdot \log(f(\theta_j - \theta_i)) + (1 - \bar{Z}_{ji}) \log(1 - f(\theta_j - \theta_i))$$

donde $\bar{Z}_{ji} = \frac{\sum_{k=1}^N Z_{ji}^k}{N}$. Nótese además que

$$\begin{aligned} 1 - f(x) &= 1 - \frac{\exp(x)}{1 + \exp(x)} \\ &= \frac{1}{1 + \exp(x)} \\ &= \frac{\exp(-x)}{1 + \exp(-x)} \\ &= f(-x) \end{aligned}$$

Así se tiene maximizar la verosimilitud de la muestra equivale a maximizar

$$\sum_{i < j} \bar{Z}_{ji} \cdot \log(f(\theta_j - \theta_i)) + (1 - \bar{Z}_{ji}) \log(f(\theta_i - \theta_j))$$

Debe observarse que si se quiere maximizar con respecto a θ_i , $i = 1, \dots, p$ se tiene que el problema no tiene solución única, ya que si $(\theta_1, \dots, \theta_8)$ es solución, $(\theta_1 + K, \dots, \theta_8 + K)$ también lo es. Esto es condición para imponer la restricción:

$$\sum_i \theta_i = 0$$

Con esto, se tiene el problema de optimización

$$\begin{aligned} \text{Max}_{\theta} \quad & \sum_{i < j} \bar{Z}_{ji} \cdot \log(f(\theta_j - \theta_i)) + (1 - \bar{Z}_{ji}) \log(f(\theta_i - \theta_j)) \\ \text{s.a.} \quad & \sum_i \theta_i = 0 \end{aligned}$$

Ahora bien, la solución del problema anterior también será solución del problema

$$\begin{aligned} \text{Min}_{\theta} \quad & \sum_{i < j} F_{ji} (\theta_j - \theta_i) \\ \text{s.a.} \quad & \sum_i \theta_i = 0 \end{aligned}$$

donde

$$F_{ji} (\theta_j - \theta_i) = -\overline{Z_{ji}} \cdot \log (f (\theta_j - \theta_i)) - (1 - \overline{Z_{ji}}) \log (f (\theta_i - \theta_j))$$

Se tiene que el problema de minimización anterior es un problema convexo. En efecto, hay que observar que las funciones F_{ji} son convexas, pues

$$\frac{\partial F_{ji}}{\partial \theta_j} = -\overline{Z_{ji}} \frac{1}{f(\theta_j - \theta_i)} f'(\theta_j - \theta_j) + (1 - \overline{Z_{ji}}) \frac{1}{f(\theta_i - \theta_j)} f'(\theta_i - \theta_j)$$

Por simetría se tiene que

$$\frac{\partial F_{ji}}{\partial \theta_j} = -\frac{\partial F_{ji}}{\partial \theta_i}$$

lo que implica que

$$\frac{\partial^2 F_{ji}}{\partial \theta_j^2} = \frac{\partial^2 F_{ji}}{\partial \theta_i^2} = -\frac{\partial^2 F_{ji}}{\partial \theta_j \partial \theta_i} = -\frac{\partial^2 F_{ji}}{\partial \theta_i \partial \theta_j}$$

Además se tiene que

$$\frac{\partial^2 F_{ji}}{\partial \theta_j^2} = \overline{Z_{ji}} \frac{\exp(\theta_j - \theta_i)}{(1 + \exp(\theta_j - \theta_i))^2} + (1 - \overline{Z_{ji}}) \frac{\exp(\theta_i - \theta_j)}{(1 + \exp(\theta_i - \theta_j))^2}$$

Luego, a θ_j y θ_i fijos, la matriz hessiana de F_{ji} es de la forma

$$\begin{pmatrix} \alpha & -\alpha \\ -\alpha & \alpha \end{pmatrix}$$

con $\alpha > 0$. Luego, en vista que

$$(\theta_j \ \theta_i) \begin{pmatrix} \alpha & -\alpha \\ -\alpha & \alpha \end{pmatrix} \begin{pmatrix} \theta_j \\ \theta_i \end{pmatrix} = \alpha (\theta_j - \theta_i)^2$$

se tiene que la matriz Hessiana es semidefinida positiva. Luego, la función F_{ji} es convexa, y dado que la función objetivo es suma de funciones convexas y el conjunto en donde se restringe el problema es convexo, se tiene que el problema es convexo.

Sin embargo, esta situación no asegura existencia de soluciones. Es más, si $\overline{Z_{ji}} = 1$ para todo par j, i , se tiene que el problema es irrestricto. Con esto, se hace necesario, imponer un suavizamiento sobre las estimaciones de las probabilidades de declarar que j es mayor que i . Éste será el mismo que se ocupó en el método anterior,

$$\overline{Z_{ji}} = \frac{\sum_{k=1}^N (Z_{ji}^k) + 1}{N + 2}$$

Con esto, y simulando que todos los alumnos contestan correctamente, a diferencia del método anterior, sí se tiene como resultado un buen ordenamiento de las fracciones en estudio, como se muestra en la figura 4.9.

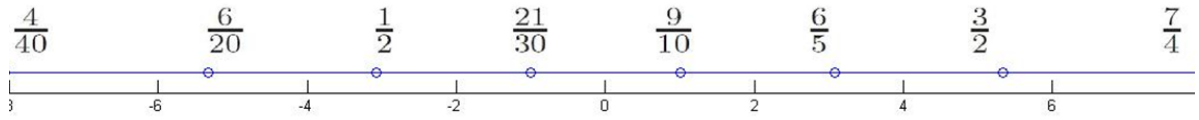


Figura 4.9: Consistencia del Método de Máxima Verosimilitud

4.4.3. Método de Camino Máximo

La última propuesta de ordenamiento hecho como parte de este trabajo, está basada en la idea de incremento psicométrico. Si se asume que existe un orden predominante en la tarea de comparación de fracciones, entonces se debería tener que si j se concibe mayor que i entonces, la probabilidad de declarar j mayor que i debe cumplir

$$P_{j,i} > \frac{1}{2}$$

De este modo, para cada par i, j de estímulos se tiene que, o bien $P_{j,i} > \frac{1}{2}$, o bien $P_{i,j} > \frac{1}{2}$. Si eventualmente los estímulos son ordenados de forma creciente bajo algún criterio, dígase i_1, i_2, \dots, i_8 , entonces una condición necesaria que se debería cumplir es que

$$P_{i_{k+1}, i_k} > \frac{1}{2} \quad (4.1)$$

$$\forall k = 1, \dots, 7$$

Una primera pregunta es si bajo cualquier configuración de los resultados entregados por los alumnos, existirá un orden de las fracciones que cumpla la condición anterior. Para responder esto, se ocupará el marco de teoría de grafos. El tipo de grafos con el que se trabajará está dado por la siguiente definición

Definición 4.1 *Se dirá que un grafo dirigido es de discriminación si su matriz de incidencia A , cumple*

$$A_{i,j} + A_{j,i} = \begin{cases} 1 & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases}$$

Es decir, para todo par de nodos existe un único arco que los une.

Recuérdese que las matrices de incidencias en grafos dirigidos son matrices cuadradas de orden n , donde n representa la cantidad de vértices del grafo. Estas matrices están compuestas sólo de ceros y unos, encontrando un 1 en la posición (i, j) si existe el arco que parte desde el vértice i y llega al vértice j .

A partir de los resultados entregados por parte de los alumnos, se puede construir un grafo de discriminación, en donde la matriz de incidencia queda dada por

$$A_{i,j} = \begin{cases} 1 & \text{si } P_{j,i} > \frac{1}{2} \\ 0 & \text{si no} \end{cases}$$

Con esto, encontrar una secuencia de vértices que cumpla la condición (4.1) equivale a encontrar un camino Hamiltoniano en el grafo de discriminación construido. La siguiente proposición muestra que esta condición siempre se cumple.

Proposición 4.2 *Todo grafo de discriminación posee un camino Hamiltoniano dirigido*

DEMOSTRACIÓN. Se procede por inducción sobre la cantidad de vértices. Para $n = 2$ la existencia del camino Hamiltoniano es evidente, ya que hay un único arco en el grafo, el cual es el que une a los dos vértices.

Para mostrar que si se tiene la propiedad para n vértices, entonces se tiene para $n + 1$, nótese lo siguiente. Si se toma un grafo de discriminación con $n + 1$ vértices, entonces, si se selecciona n vértices cualesquiera, y se restringe el grafo a esos nodos, entonces, por hipótesis de inducción, existe un camino Hamiltoniano que recorre a los n vértices seleccionados. Sin pérdida de generalidad, enumérese a dichos vértices desde 1 hasta n según el recorrido del camino.

Por otra parte, dado que el grafo con los $n + 1$ vértices es de discriminación, existe un arco que une al nodo $n + 1$ con el primer vértice del camino Hamiltoniano. Si este arco parte en el nodo $n + 1$ y llega a 1 entonces se tiene un camino Hamiltoniano para el grafo completo. A su vez, como el grafo es de discriminación, entonces existe un arco que une a n con $n + 1$. Si este arco parte en n y termina en $n + 1$, entonces se tiene un camino Hamiltoniano para el grafo completo. Estas dos situaciones se pueden apreciar en la figura 4.10.

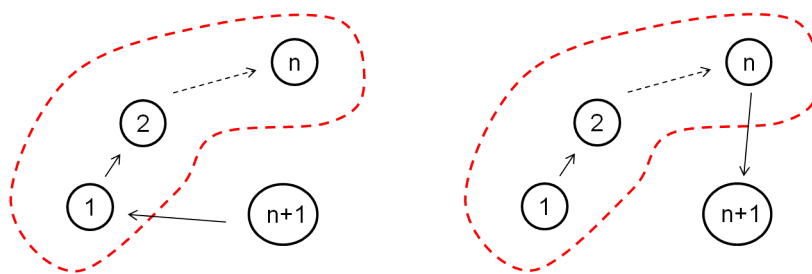


Figura 4.10: Situaciones simples de existencia de camino Hamiltoniano

De este modo, supóngase que no se tiene ninguna de las dos situaciones antes mencionadas, es decir el arco que une a $n + 1$ con 1 comienza en 1 y termina en $n + 1$, mientras que el arco que une a n con $n + 1$ comienza en $n + 1$ y termina en n .

Considérese i dado por

$$i = \inf_{j: A_{n+1,j}=1} j$$

es decir, i es el primer vértice en el camino Hamiltoniano desde 1 hasta n , en donde el arco que se conecta con $n + 1$ sale del nodo $n + 1$. Esto se intenta ilustrar en la figura 4.11.

Así, se tiene que un camino Hamiltoniano para el grafo completo sería el camino $(1, 2, \dots, i - 1, n + 1, i, i + 1, \dots, n - 1, n)$. Luego, siempre existe un camino Hamiltoniano para un grafo de discriminación.

□

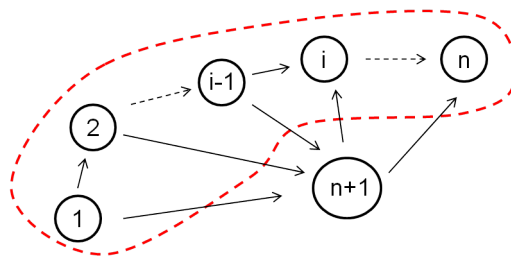


Figura 4.11: Situación General de existencia de camino Hamiltoniano

En vista de la proposición anterior, habría que determinar algún criterio para elegir un camino Hamiltoniano. Dado que se busca una escala, parece razonable proponer el camino que más discrimine, esto es, el que maximice

$$\sum_{k=1}^{n-1} P_{i_k, i_{k+1}} - \frac{1}{2}$$

donde n es la cantidad de estímulos, en este caso, fracciones.

Una condición necesaria que debe cumplir un camino para ser Hamiltoniano es que de todos los vértices puede salir a lo más un arco, que de todos los vértices puede incidir a lo más un arco, y que la cantidad de arcos que forman el camino es exactamente $n - 1$, donde n es la cantidad de vértices. Si se observa este camino como un subgrafo, se tiene que esta condición no es suficiente. La figura 4.12 muestra un contraejemplo de esto.

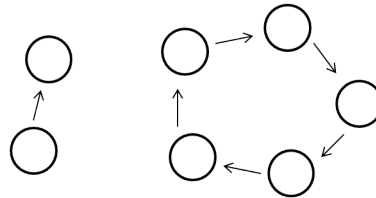


Figura 4.12: Contraejemplo sobre condición suficiente de Camino Hamiltoniano.

Sin embargo, trabajar con esta condición necesaria puede ser útil para encontrar el camino que se busca. Una manera para tratar de encontrar este camino, es formulando un problema de programación lineal entera en donde el conjunto de restricciones se asocie al cumplimiento de las condiciones descritas. Este problema se formula de la siguiente manera

$$\begin{aligned} \text{Max} \quad & \sum_{i,j} \left(P_{j,i} - \frac{1}{2} \right) \cdot X_{i,j} \\ \text{s.a.} \quad & X_{i,j} \leq A_{i,j} \\ & X \cdot \mathbf{1}_n \leq \mathbf{1}_n \\ & X^t \cdot \mathbf{1}_n \leq \mathbf{1}_n \\ & \sum_{i,j} X_{i,j} = n - 1 \\ & \forall i, j \quad X_{i,j} \in \{0, 1\} \end{aligned}$$

1_n es el vector de n componentes donde todas valen 1. La incógnita en este problema es la matriz X , la cual determinará el camino que se busca. La primera restricción alude a que el camino esté dentro del grafo, mientras que las restantes son las correspondientes a la condición necesaria señalada para que el camino sea Hamiltoniano. Este problema se puede trabajar como un problema vectorial de la siguiente forma

$$\begin{aligned}
 &Max \sum_{k=1}^{n^2} \tilde{P}_k \cdot Y_k \\
 &s.a. Y_k \leq \tilde{A}_k \\
 &U \cdot Y \leq 1_n \\
 &\tilde{U} \cdot Y \leq 1_n \\
 &\sum_k Y_k = n - 1 \\
 &\forall k Y_k \in \{0, 1\}
 \end{aligned}$$

Donde $Y_k = X_{i,j}$, $\tilde{P}_k = P_{j,i} - \frac{1}{2}$, $\tilde{A}_k = A_{i,j}$ con $k = (i-1)n + j$, mientras que $U, \tilde{U} \in \mathcal{M}_{n \times n^2}$, están dadas por

$$U = (U_1, U_2, \dots, U_n)$$

$$(U_k)_{ij} = \begin{cases} 1 & \text{si } i = k \\ 0 & \text{si no} \end{cases}$$

con $U_k \in \mathcal{M}_{n \times n}$, y

$$\tilde{U} = (I, I, \dots, I)$$

donde I es la matriz identidad de orden n . Esta formulación es útil, ya que para resolver estos problemas se utilizaron métodos numéricos cuyos inputs deben ser ingresados en formato vectorial.

El problema anterior no necesariamente tiene solución única ni tampoco es un camino Hamiltoniano. En caso que se encuentra una solución que sea un camino Hamiltoniano, se consigue un criterio de orden que parece ser razonable. Una observación importante es el que el método es consistente, pues si los alumnos contestaran de manera totalmente correcta, entonces el problema de optimización tiene solución única y el método entrega un orden correcto. En la figura 4.13 se observa que en cada arco aparece el valor del incremento $P_{j,i} - \frac{1}{2}$.

4.4.4. Comparación de los Métodos

Un primer ejercicio que se puede ejecutar para comparar los métodos es aplicarlos a nivel global con los resultados obtenidos del PostTest. Distinguiendo las respuestas de los

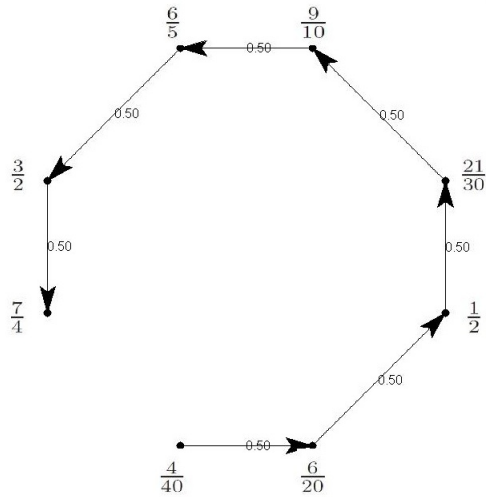


Figura 4.13: Consistencia del Método de Camino Máximo

estudiantes según el tipo de entrenamiento al que fueron sometidos, se puede crear tres tipos de escalamientos distintos para cada método.

En el caso del método de Máxima Verosimilitud los escalamientos para cada uno de los grupos de entrenamiento se puede ver en la figura 4.14.

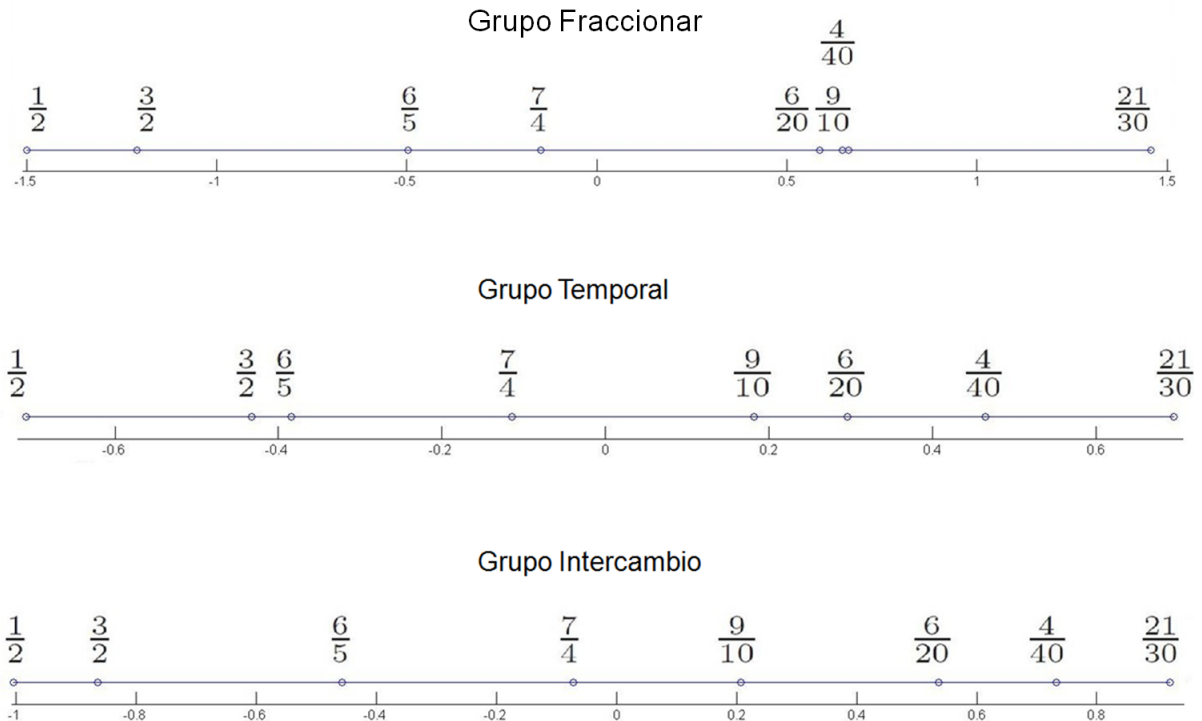


Figura 4.14: Escalamiento con Máxima Verosimilitud según Tipo de Entrenamiento

De las figuras anteriores se aprecia que en la mayor parte de los alumnos se tiende a considerar como las fracciones mayores aquellas que están formadas, tanto numerador como denominador, por números grandes. Así, según este tipo de análisis, es poco claro el aporte

que tienen los distintos métodos de enseñanza. Sin embargo, es importante notar que el entrenamiento fue bastante breve, por lo que si se quiere buscar diferencias entre los distintos métodos de enseñanza, se tendrá que identificar pequeños cambios de comportamiento en los alumnos. El método de máxima verosimilitud para estimar escalamiento se ve fuertemente influenciado por la gran cantidad de alumnos que no realizó correctamente la tarea.

A conclusiones parecidas llega el método de Camino Máximo, en donde se observa la misma tendencia que en el caso anterior, esto es, asociar como fracciones mayores aquellas formadas por números más grandes. Los caminos para cada grupo se muestran en la figura 4.15. Tanto el grupo Temporal como el Intercambio tienen el mismo ordenamiento, mientras que el grupo Fraccionar difiere solamente en las dos últimas fracciones que considera como mayores.

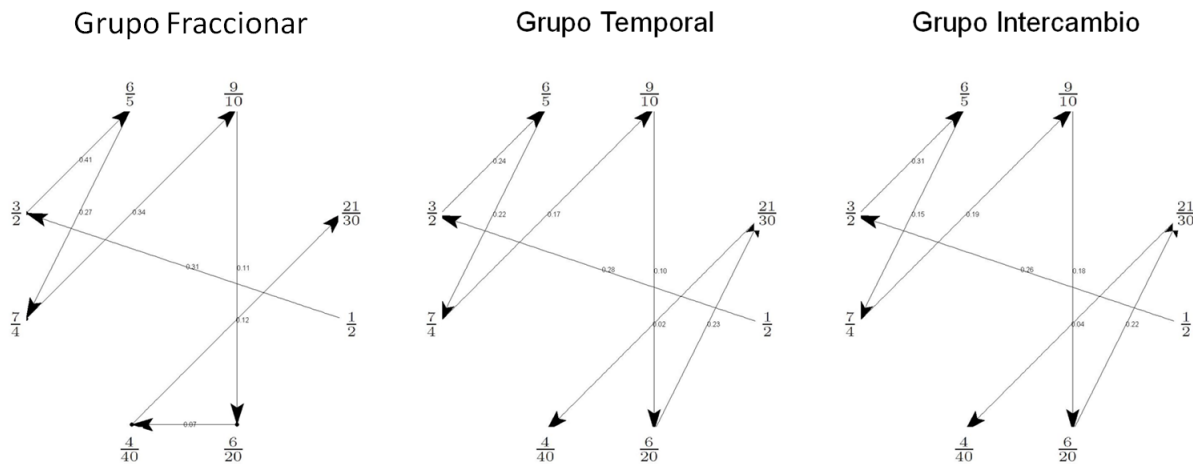


Figura 4.15: Camino Máximo según Tipo de Entrenamiento

Dado que ambos métodos entregan una visualización global del comportamiento, es bastante difícil identificar diferencias pequeñas entre los grupos. Una alternativa para resolver esto sería aplicar cada uno de los métodos con las respuestas individuales de cada alumno. Si bien esto podría dar luces de patrones en el comportamiento individual, los métodos terminan por entregar 8 parámetros construidos con 28 datos. Esto podría hacer que los métodos se vuelvan poco robustos y el análisis inválido. Otro problema que genera este tipo de métodos es el diseño experimental que involucra, ya que para poder aplicarlos es necesario testear todas las combinaciones entre estímulos analizados, como se dió en el Post Test. Sin embargo, los métodos no pueden ser aplicados en el Pre Test ya que, en el caso del Método de Máxima Verosimilitud, el tipo de estimaciones que podrían dar estaría sesgada de acuerdo a las preguntas con las que se construye el análisis, por ejemplo, si sólo hubiera una pregunta donde hay una fracción específica que se compara con otra, entonces el método entregaría sólo dos posibles relaciones, que una fracción es mayor a la otra, o que es menor a la otra. De este modo, no se puede establecer con claridad la relación entre dicha fracción con las que no se comparan directamente. En el caso del método de Camino Máximo, si no comparan directamente todas las preguntas, puede darse que el camino Hamiltoniano no exista.

Para intentar analizar los resultados en situaciones generales, en donde no sea necesario la comparación directa entre todos los estímulos involucrados, se propondrá un modelo ba-

sados en estrategias. Se supondrá que un alumno puede ocupar una de dos estrategias al momento de contestar una pregunta, dando la posibilidad que distintas estrategias generen distintas respuestas. Luego se estimará la probabilidad de utilizar una determinada estrategia al momento de comparar fracciones.

4.5. Estimación de Estrategias

En este modelo, se dará una estructura precisa a P_{ji}^k , esto es, la probabilidad de declarar j mayor que i por el alumno k . A diferencia de los modelos anteriores, aquí se incorpora un factor por alumno, el cual será especificado más adelante.

Al evento de declarar j mayor que i por el alumno k se denominará D_{ji}^k . Obviamente, se tiene que $\mathbb{P}(D_{ji}^k) = P_{ji}^k$

En principio, podrían existir múltiples estrategias que los alumnos serían capaces de utilizar. Para simplificar el modelo, se asumirá que los alumnos sólo pueden utilizar dos estrategias para comparar fracciones. La primera de ellas, que se denominó estrategia R , es aquella que al utilizarla, la probabilidad de encontrar la mayor de dos fracciones depende del cociente entre ellas. Por otra parte, la estrategia restante, la que se denominó por estrategia M , es aquella que al utilizarla, la probabilidad de encontrar la mayor de dos fracciones depende de la diferencia de los máximos entre el numerador y el denominador de cada fracción.

Así, el modelo que se propone supone que, si $i = \frac{a}{b}$, $j = \frac{c}{d}$, entonces

$$\mathbb{P}(D_{ji}^k | R) = f(\log(j) - \log(i))$$

y que

$$\mathbb{P}(D_{ji}^k | M) = f(\max(c, d) - \max(a, b))$$

donde

$$f(x) = \frac{\exp(x)}{1 + \exp(x)}$$

Con estas definiciones, se entiende que utilizar la estrategia R es más correcto que utilizar la estrategia M , ya que con la primera al momento de seleccionar la mayor de dos fracciones, es más probable elegir aquella que realmente es la mayor que elegir la menor.

A su vez, la probabilidad que el alumno k utilice la estrategia R se denominará α_k . Por consiguiente, la probabilidad que utilice la estrategia M es $1 - \alpha_k$.

Luego, con el modelo establecido, se tiene por el teorema de Probabilidades Totales que

$$P_{ji}^k = f(\log(j) - \log(i)) \cdot \alpha_k + f(\max(c, d) - \max(a, b)) \cdot (1 - \alpha_k)$$

Con esto, la tarea que queda es determinar para cada alumno cuanto vale α_k .

Se define Z_{ji}^k la respuesta que entrega el alumno k al comparar i con j , valiendo 1 si declara que j es mayor que i , 0 si no.

Para estimar los valores de α_k se utilizará el método de máxima verosimilitud. Así, la verosimilitud de la muestra, para ciertos valores de α_k con $k = 1, \dots, n$ es

$$\prod_{i < j} \prod_{k=1}^n (P_{ji}^k)^{Z_{ji}^k} \cdot (1 - P_{ji}^k)^{1 - Z_{ji}^k}$$

con lo que tomando logaritmo queda

$$\sum_{i < j} \sum_{k=1}^n Z_{ji}^k \log(P_{ji}^k) + (1 - Z_{ji}^k) \log(1 - P_{ji}^k)$$

es decir

$$\begin{aligned} & \sum_{i < j} \sum_{k=1}^n Z_{ji}^k \log(f(\log(j) - \log(i)) \cdot \alpha_k + f(\max(c, d) - \max(a, b)) \cdot (1 - \alpha_k)) \\ & + (1 - Z_{ji}^k) \log(1 - f(\log(j) - \log(i)) \cdot \alpha_k - f(\max(c, d) - \max(a, b)) \cdot (1 - \alpha_k)) \end{aligned}$$

Así, para estimar α_k se resolverá el problema de optimización de maximizar la función anterior, sujeto a que $\alpha_k \in [0, 1] \forall k = 1, \dots, n$

Hay que notar que, dado que la derivada parcial de la función objetivo con respecto a α_k sólo depende de α_k , resolver el problema anterior equivale a resolver n problemas en conjunto. Por otra parte, si el problema de maximización se transforma en un problema de minimización, resulta que este último es convexo. Para resolver este problema, se utilizaron métodos numéricos.

4.5.1. Resultados

Una de las ventajas de este método es que no se necesita que todas las fracciones que aparecen en las preguntas se comparen directamente, a diferencia de los métodos anteriores. Esto permite, hacer estimaciones no sólo en el Post Test, si no también en el Pre Test. De esta manera, en primera lugar se procedió a estimar el valor de α para cada alumno con las seis preguntas en formato simbólico que aparecieron en el Pre Test. Una vez calculados estos valores, se puede comparar la variación según grupo de entrenamiento. Como es de esperar, con los cálculos hechos con las preguntas del Pre Test, no se detectan diferencias estadísticamente significativas (Test ANOVA p-valor = 0,597).

Luego, se procedió a estimar los valores de α después del entrenamiento. Con las respuestas obtenidas en el Post Test, se estimaron las probabilidad de utilizar la estrategia R . Hecho este cálculo, se estudió el incremento de estos valores después del entrenamiento, es decir, se estudió la diferencia entre los valores de α calculados con los datos del Post Test con los calculados a partir del Pre Test. Para el caso del grupo Fraccionar, el incremento no registró una variación significativa (p-valor = 0,875). Sin embargo, tanto en los Grupos

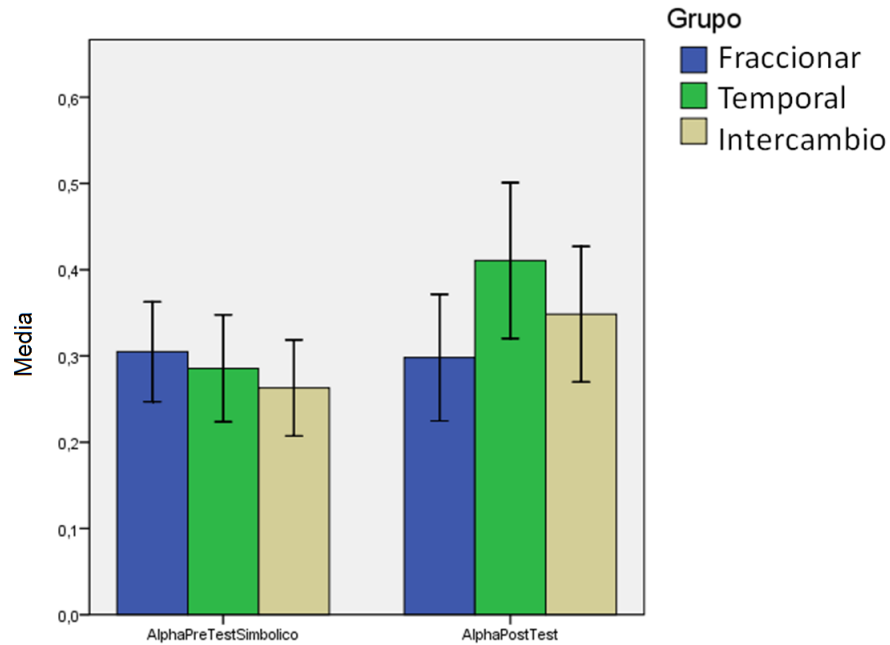


Figura 4.16: Variación de valores promedios de α según tipo de entrenamiento. A la izquierda se encuentran los valores promedios de la probabilidad de utilizar la estrategia R calculados con las respuestas a las preguntas en formato simbólico del Pre Test. A la derecha se encuentran los mismos estimadores pero calculados con los datos del Post Test.

Temporal como Intercambio se registran alzas en los valores de α después del entrenamiento que son estadísticamente significativas (p-valor = 0,013 y p-valor = 0,034 respectivamente). Lo descrito anteriormente se ilustra en la figura 4.16.

Este resultado es muy importante, ya que con esto se muestra que existe una diferencia entre los distintos tipos de entrenamiento. El hecho que los incrementos de los valores de α sean estadísticamente significativos en los grupos Temporal e Intercambio, da a entender que estas maneras de representar las fracciones podría ayudar a mejorar el desempeño en esta área. Es importante notar que estos hallazgos igualmente son pequeños, ya que no se puede desconocer, según los métodos de escalamiento, la prevalencia a ordenar las fracciones según el sesgo hacia los números enteros. De todos modos, es sumamente interesante haber encontrado una diferencia en los métodos de enseñanza con entrenamientos tan pequeños en su duración.

4.6. Conclusiones del Estudio

Sin duda, una de las principales conclusiones hace referencia al hallazgo en las diferencias de los métodos de enseñanza. Esto se observa tanto en el puntaje promedio de las Preguntas Incongruentes Difíciles del Post Test como en el cálculo de los estimadores de la probabili-

dad de utilizar la estrategia *R*. Declarar que los métodos Temporal e Intercambio son más ventajosos que el método tradicional de enseñanza de fracciones encaja en el marco teórico construido al principio de este estudio. Las representaciones Temporal e Intercambio podrían considerarse como privilegiadas respecto a la representación Geométrica de Fracción. Aparentemente, el ser humano está mejor diseñado para pensar en tasas temporales o en situaciones de intercambio, que en construcciones geométricas.

Es muy notable observar que estas diferencias se presentan aún cuando el entrenamiento es breve, ya que los alumnos no tardaban más de 20 minutos en ver los videos y contestar las preguntas respectivas. Esto daría indicios de las enormes ventajas que podría producir enseñar fracciones con las representaciones Temporal o Intercambio. Sin embargo, hay que ser cauto en este planteamiento, ya que en este estudio sólo se analizó la tarea de comparación de fracciones, dejando de lado otros aspectos como la suma, la multiplicación y división de fracciones. También es necesario estudiar cómo los alumnos retienen la información en el tiempo cuando se les enseña de manera sistemática con estos métodos. De este modo, es fundamental extender esta investigación, para así validar contundentemente los beneficios de las representaciones Temporal e Intercambio.

Respecto a las metodologías desarrolladas, se puede decir que se intentó abordar el problema desde un enfoque basado en los modelos de comparación de estímulos. Mediante este camino no se observaron diferencias claras entre los tres métodos de enseñanza. Es más, los resultados arrojados por el modelo indicarían que, en general, los alumnos tienen un evidente sesgo hacia los números enteros, esto sin hacer distinción entre tipos de entrenamiento. En principio, esto no debería sorprender, ya que los porcentajes de respuestas correctas de las preguntas Incongruentes Simples e Incongruentes Difíciles son bajo el 50 %, lo que indica la tendencia general a identificar como mayor aquellas fracciones que poseen números mayores. Pese a esto, la metodología no se invalida, si no todo lo contrario, ya que a nivel general, si un nuevo alumno pasa por cualquiera de los entrenamientos, es muy probable que haga sus declaraciones basado en el sesgo hacia los números enteros. Así, el análisis hecho con las estrategias indica que esas pequeñas probabilidades de mejorar aumentan de manera distinta según el tipo de entrenamiento.

Finalmente, es prudente indentificar las diferencias con otros estudios que aparentemente son muy similares. Se podría indicar que los métodos de enseñanza basados en los formatos Temporal e Intercambio se fundamentan en la noción de intensidad, como es tratada en [29]. Sin embargo, existen dos diferencias con dicho estudio. La primera es el tiempo de entrenamiento; en la investigación citada los alumnos fueron entrenados con 4 clases, mientras que en este estudio sólo estuvieron menos de 15 minutos en el proceso de aprendizaje. La otra diferencia hace alusión a el tipo de preguntas en el Post Test; en la otra investigación las preguntas hechas contextualizaban los problemas de fracciones directamente a un problema de intensidad, mientras que en este estudio se trabaja directamente con la comparación de fracciones en formato simbólico. De esta manera, este estudio revela nuevos hallazgos en relación a los tiempos necesarios para que se experimente aprendizaje en el trabajo con fracciones en lenguaje simbólico, y no solamente en el dominio de alguna metáfora, junto con el tipo de preguntas que los alumnos serían capaces de contestar correctamente.

Conclusión

La conclusión global, además de las conclusiones mencionadas en cada uno de los capítulos, es que este trabajo cumple su objetivo más importante, dar una perspectiva desde la modelación a aspectos fundamentales de la cognición matemática. En particular, en el capítulo 2, se logra dar un enfoque matemático para explicar los fenómenos básicos asociados al Sentido Numérico. Esto permitió hacer distinciones entre los distintos modelos de representación numérica que actualmente son propuestos por investigadores del área. A partir de eso, se consideró que el escalamiento más adecuado para los números en la recta numérica mental debería tener forma logarítmica.

Del capítulo 3, queda determinar experimentalmente si el supuesto sobre la cantidad de neuronas asociadas a los números tiene o no sentido, es decir, si efectivamente hay menos neuronas asociadas a los números grandes que asociadas a los números pequeños. Junto con esto, se propone estudiar el modelo con casos más generales que el aquí desarrollado, ya que la cantidad de neuronas específicas de numerosidad fue modelada por una función dada y no por una función arbitraria. Además, el modelo desarrollado hace predicciones sobre el impacto de la calibración que deben ser testeadas experimentalmente.

En relación a la investigación expuesta en el capítulo 4, es importante seguir examinando de qué manera funcionan las representaciones de fracciones propuestas. En este trabajo se analizaron resultados para una tarea específica, por lo que es primordial saber qué tipo de ventajas genera en otro tipo de tareas enseñar fracciones utilizando las metodologías aquí investigadas. La comparación de fracciones es lo primero que suele ser estudiado, sin embargo, se debe conocer el nivel de desempeño en tareas como la suma o la multiplicación de fracciones y el nivel de retención en el tiempo que tienen los estudiantes al momento de aprender estas representaciones.

Como se dijo en un principio, este trabajo de memoria intentó ser un aporte a la Educación Matemática a través de la modelación de procesos cognitivos asociados a la percepción numérica. Es fundamental profundizar las investigaciones aquí expuestas con el fin de comprender de mejor manera cómo aprendemos y comprendemos los elementos básicos de la matemática elemental. Existe la fuerte convicción que este tipo de estudios podría producir grandes beneficios en nuestro sistema educativo.

Bibliografía

- [1] John Adams and Graham Hitch. *Working memory and children's mental addition*. Journal of Experimental Child Psychology, 1997.
- [2] Julie Booth and Robert Siegler. *Numerical magnitude representations influence arithmetic learning*. Child Development, 2008.
- [3] Elizabeth Brannon, Courtney Wusthoff, C. R. Gallistel, and John Gibbon. *Numerical Subtraction in the Pigeon. Evidence for a Linear Subjective Number Scale*. Psychological Science, 2001.
- [4] Gary Brase. *Bugs built into the system: How privileged representations influence mathematical reasoning across the lifespan*. Learning and Individual Differences, 2002.
- [5] Rebecca Bull and Rhona. Johnston. *Children's Arithmetical Difficulties: Contributions from Processing Speed, Item Identification, and Short-Term Memory*. Journal of Experimental Child Psychology, 1997.
- [6] Jamie Campbell. *Handbook of Mathematical Cognition*. Psychology Press, 2005.
- [7] Jessica Cantlon and Elizabeth Brannon. *Shared system for ordering small and large numbers in monkeys and humans*. Psychological Science, 2006.
- [8] T. Carpenter, M. Corbitt, H. Kepner, M. Lindquist, and R Reys. *National assessment: Prospective of students' mastery of basic skills* . Selected issues in mathematics education, 1980.
- [9] R. J. de Ayala. *The Theory and Practice of Item Response Theory*. Guilford Press, 2008.
- [10] SIMCE Unidad de Curriculum y Evaluacion. *Resumen de Resultados PISA 2009 Chile*. Ministerio de Educacion de Chile.
- [11] Stanislas Dehaene. *The Number Sense, How The Mind Creates Mathematics*. Oxford University Press, New York, 1st edition, 1997.
- [12] Stanislas Dehaene. *The neural basis of the Weber-Fechner law: a logarithmic mental number line*. Trends in Cognitive Sciences, 2003.
- [13] Stanislas Dehaene. *Symbols and quantities in parietal cortex: elements of a mathematical*

theory of number representation and manipulation. Cognition, 2007.

- [14] Stanislas Dehaene and Elizabeth Brannon. *Space, Time and Number in the Brain, Searching for the Foundations of Mathematical Thought.* Academic Press, 2011.
- [15] Stanislas Dehaene, Veronique Izard, Elizabeth Spelke, and Pierre Pica. *Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures.* Science, 2008.
- [16] Ehtibar Dzhafarov and Hans Colonius. *Fechnerian metrics in unidimensional and multidimensional stimulus spaces.* Psychonomic Bulletin Review, 1999.
- [17] Ehtibar Dzhafarov and Hans Colonius. *Psychophysics without physics: a purely psychological theory of Fechnerian scaling in continuous stimulus spaces.* Journal of Mathematical Psychology, 2005.
- [18] Ehtibar Dzhafarov and Hans Colonius. *Psychophysics without physics: extension of Fechnerian scaling from continuous to discrete and discrete-continuous stimulus spaces.* Journal of Mathematical Psychology, 2005.
- [19] Lawrence Evans. *Partial Differential Equations.* American Mathematical Society, 1998.
- [20] Jean-Claude Falmagne. *Elements of Psychophysical Theory.* Oxford Psychology Series, 1985.
- [21] C. R. Gallistel and Rochel Gelman. *Preverbal and verbal counting and computation.* Cognition, 1992.
- [22] Dana Ganor-Stern. *Fractions but not negative numbers are represented on the mental number line.* Acta Psychologica, 2010.
- [23] Susan Gathercole and Susan Pickering. *Assessment of working memory in six- and seven-year-old children.* Journal of Educational Psychology, 2000.
- [24] David Geary, Christine Bow-Thomas, and Yuhong Yao. *Counting knowledge and skill in cognitive addition: a comparison of normal and mathematically disabled children.* Journal of Experimental Child Psychology, 1992.
- [25] David Gomez. *Estudios experimentales y de modelacion en aprendizaje y cognicion matematica.* Tesis para optar al grado de Doctor en Ciencias de la Ingenieria Mencion Modelacion Matematica. Universidad de Chile, 2010.
- [26] Michael Gumert. *Payment for sex in a macaque mating market.* Animal Behaviour, 2006.
- [27] Michael Gumert. *Grooming and infant handling interchange in Macaca fascicularis: The relationship between infant supply and grooming payment.* International Journal of Primatology, 2007.

- [28] S. Henzi and L. Barrett. *Infants as a commodity in a baboon market*. Animal Behaviour, 2002.
- [29] Christine Howe, Terezinha Nunes, and Peter Bryant. *Rational number and proportional reasoning: Using Intensive Quantities to Promote Achievement in Mathematics and Science*. International Journal of Science and Mathematics Education, 2011.
- [30] Tarow Indow and Masashi Ida. *Scaling of dot numerosity*. Perception Psychophysics, 1977.
- [31] Veronique Izard and Stanislas Dehaene. *Calibrating the mental number line*. Cognition, 2007.
- [32] Simon Jacob, Daniela Valletín, and Andreas Nieder. *Relating magnitudes: the brain's code for proportions*. Trends in Cognitive Sciences, 2012.
- [33] Peter Kloosterman. *Mathematics Skills of 17-Year-Olds in the United States: 1978 to 2004*. Journal for Research in Mathematics Education, 2010.
- [34] Lester Krueger. *Single judgments of numerosity*. Perception Psychophysics, 1982.
- [35] Lester Krueger. *Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments*. Perception Psychophysics, 1984.
- [36] Mario Lagos. *Análisis Factorial de Proximidades*. Memoria para Optar al Título de Ingeniero Matemático. Universidad de Chile, 1978.
- [37] Elida Laski and Robert Siegler. *Is 27 a Big Number? Correlational and Causal Connections Among Numerical Categorization, Number Line Estimation, and Numerical Magnitude Comparison*. Child Development, 2007.
- [38] John Lee and Michel Verleysen. *Nonlinear Dimensionality Reduction*. Springer, 2007.
- [39] T. Loetscher and P. Brugger. *Exploring number space by random digit generation*. Exp. Brain Res., 2007.
- [40] Liping Ma. *Conocimiento y Enseñanza de las Matemáticas Elementales*. Academia Chilena de Ciencias, 2010.
- [41] Jacques Mehler and Thomas Bever. *Cognitive Capacity of Very Young Children*. Science, 1967.
- [42] Jacques Mehler and Thomas Bever. *Cognitive Capacities of Young Children: A Replication*. Science, 1968.
- [43] Gonzalo Mena-Carrasco, David Gomez, Roberto Araya, and Pablo Dartnell. *Individual differences in inhibitory capacities and the learning of fractions in elementary school*. In Press, 2012.
- [44] Yujing Ni and Yong-Di Zhou. *Teaching and Learning Fraction and Rational Numbers:*

- The Origins and Implications of Whole Number Bias*. Educational Psychologist, 2005.
- [45] Andreas Nieder, David Freedman, and Earl Miller. *Representation of the Quantity of Visual Items in the Primate Prefrontal Cortex*. Science, 2002.
- [46] National Council of Teachers of Mathematics. *The Learning of Mathematics: 69th NCTM Yearbook*. National Council of Teachers of Mathematics. 1906 Association Drive, Reston, 2007.
- [47] National Mathematics Advisory Panel. *The Final Report of the National Mathematics Advisory Panel*. U.S. Department of Education, 2008.
- [48] Jean Piaget. *The Child's Conception of Number*. Routledge Kegan Paul, 1952.
- [49] Manuela Piazza, Veronique Izard, Philippe Pinel, Denis Le Bihan, and Stanislas Dehaene. *Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus*. Neuron, 2004.
- [50] Geetha Ramani and Robert Siegler. *Promoting Broad and Stable Improvements in Low-Income Children's Numerical Knowledge Through Playing Number Board Games*. Child Development, 2008.
- [51] Jamie Roitman, Elizabeth Brannon, and Michael Platt. *Monotonic Coding of Numerosity in Macaque Lateral Intraparietal Area*. Plos Biology, 2007.
- [52] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2010.
- [53] Michael Schneider and Robert Siegler. *Representations of Magnitudes of Fractions*. Journal of Experimental Psychology, 2010.
- [54] Robert Siegler, Thomas Carpenter, Francis Fennell, Favid Geary, James Lewis, Yukai Okamoto, Laurie Thompson, and Steven E. Shreve. *Developing Effective Fractions Instruction for Kindergarten Through 8th Grade*. Institute of Education Sciences, 2010.
- [55] Robert Siegler, Greg Ducan, Pamela Davis-Kean, Kathryn Ducjworth, Amy Claessens, Mimi Engel, Maria Ines Susperreguy, and Meichu Chen. *Early Predictors of High School Mathematics Achievement*. Psychological Science, 2012.
- [56] Robert Siegler, Clarissa Thompson, and Michael Schneider. *An integrated theory of whole number and fractions development*. Cognitive Psychology, 2011.
- [57] Bert De Smedt, Lieven Verschaffel, and Pol Ghesquiere. *The predictive value of numerical magnitude comparison for individual differences in mathematics achievement*. Journal of Experimental Child Psychology, 2009.
- [58] P Starkey and RG Cooper Jr. *Perception of numbers by human infants*. Science, 1980.
- [59] Mark Strauss and Lynne Curtis. *Infant Perception of Numerosity*. Child Development,

1981.

- [60] Xenia Vamvakoussi and Stella Vosniadou. *Understanding the structure of the set of rational numbers: a conceptual change approach*. Learning and Instruction, 2004.
- [61] Raffaella Ventura, Bonaventura Majolo, Nicola Kayama, Scott Hardie, and Gabriele Schino. *Reciprocation and interchange in wild Japanese macaques: grooming, cofeeding and agonistic support*. American Journal of Primatology, 2006.
- [62] Frans De Waal. *The Chimpanzee's service economy: Food for grooming*. Evolution and Human Behavior, 1997.
- [63] Matthias Wittlinger, Rudiger Wehner, and Harald Wolf. *The Ant Odometer: Stepping on stilts and stumps*. Science, 2006.
- [64] Matthias Wittlinger, Rudiger Wehner, and Harald Wolf. *The Desert Ant Odometer: A stride integrator that accounts for stride length and walking speed*. The Journal of Experimental Biology, 2007.
- [65] Monica Wong and David Evans. *Fractions as measure*. Proceedings of the 31st Annual Conference of the Mathematics Education Research Group of Australasia, 2008.
- [66] Hung-Hsi Wu. *Teaching Fractions According to the Common Core Standards*. 2011.
- [67] Karen Wynn. *Addition and subtraction by human infants*. Nature, 1992.
- [68] Liqi Zhu and Gerd Gigerenzer. *Children can solve Bayesian problems: the role of representation in mental computation*. Cognition, 2006.