



Grado en Psicología – Facultad de Psicología y Logopedia

**¿Es el Entremezclado Amigo de la Inducción de Categorías? Una
Replicación y Extensión de Kornell y Bjork (2008)**

**Is Interleaving a Friend of Category Induction? A Replication and Extension of Kornell
y Bjork (2008)**

Trabajo de Fin de Grado: Modalidad de Trabajo de Investigación

Autor:

Víctor Trillo Rodríguez

Bajo la tutorización de:

Joaquín Morís Fernández

Curso académico 2024-2025

Índice

Resumen	2
Introducción	4
Método	14
Experimento 1	14
<i>Participantes</i>	14
<i>Materiales</i>	15
<i>Procedimiento</i>	15
<i>Análisis de Datos</i>	18
<i>Resultados</i>	19
Experimento 2	22
<i>Participantes</i>	22
<i>Materiales</i>	23
<i>Procedimiento</i>	23
<i>Resultados</i>	24
Experimento 3	27
<i>Participantes</i>	27
<i>Materiales</i>	27
<i>Procedimiento</i>	27
<i>Resultados</i>	27
Experimento 4	29
<i>Participantes</i>	29
<i>Materiales</i>	30
<i>Procedimiento</i>	30
<i>Resultados</i>	30
Discusión	34
Conclusión	40
Referencias.....	41

Resumen

El aprendizaje inductivo es un proceso cognitivo mediante el cual podemos inferir reglas y patrones generales a partir de ejemplares particulares. Trabajos previos han investigado la influencia de la presentación entremezclada de ejemplares para mejorar el aprendizaje inductivo. En una condición de espaciado-entremezclado (SI), estímulos de diferentes categorías aparecen mezclados y separados en diferentes bloques. Por otra parte, en una condición de masivo-bloqueado (MB) ejemplares de una misma categoría aparecen de forma secuencial y cercana en el tiempo en un mismo bloque. En 2008, Kornell y Bjork realizaron dos experimentos utilizando pinturas de diferentes artistas. En el primero encontraron que SI mejoraba la categorización frente a MB en una tarea a corto plazo con retroalimentación. En el segundo se mostró que SI era mejor que MB en una tarea de reconocimiento en el que los participantes debían reconocer elementos nuevos y viejos de diferentes categorías. El presente trabajo replica y extiende el trabajo de Kornell y Bjork (2008) con cuatro experimentos. En el Experimento 1 y 2 se replicaron los experimentos originales, añadiendo el uso de análisis derivados de la Teoría de Detección de Señales. En el Experimento 3 se realizó una tarea de categorización eliminando la retroalimentación para evitar el reaprendizaje durante la prueba. En el Experimento 4 se probó el mismo fenómeno, pero a largo plazo, introduciendo una semana de demora entre el entrenamiento y prueba. Consistentes con resultados previos, las categorías presentadas en SI obtuvieron mayor proporción de aciertos comparado con MB en todos los experimentos.

Palabras Clave: Entremezclado, aprendizaje inductivo, aprendizaje de categorías, contraste discriminativo, memoria a largo plazo.

Abstract

Inductive learning is a cognitive process by which we can infer general rules and patterns from particular exemplars. Previous works have investigated the influence of interleaved exemplar presentation to improve inductive learning. In a spaced-interleaved (SI) condition, stimuli from different categories appear mixed and spaced apart across different blocks. On the other hand, in a massed-blocked (MB) condition, exemplars from the same category appear sequentially and closely spaced within a single block. In 2008, Kornell and Bjork conducted two experiments using paintings by different artists. In the first one, they found that SI improved categorization compared to MB in a short-term feedback task. In the second one, they found that SI was better than MB in a recognition task in which participants had to recognize new and old exemplars from different categories. The present work replicates and extends the work of Kornell and Bjork (2008) based on four experiments. Experiments 1 and 2 replicated the original experiments including analyses derived from Signal Detection Theory. In Experiment 3, a categorization task was performed without feedback to prevent relearning during the test. Experiment 4 examined whether the same effect could take place including a one-week delay between training and testing. Consistent with previous results, categories presented in SI obtained a higher proportion of correct answers compared to MB in all experiments.

Keywords: Interleaving, inductive learning, category learning, discriminative contrast, long term memory.

Introducción

La forma en la que se presenta la información puede afectar sustancialmente al aprendizaje, no solo determinando que es lo que se aprende, sino como de bien se hace (Carvalho y Goldstone, 2015a). Tanto es así, que en los últimos años han surgido gran cantidad de guías que exploran las diferentes técnicas de estudio y enseñanza (Dulonsky et al., 2013).

Una de las técnicas más recomendadas y con mayor recorrido en la literatura, es el espaciado o práctica distribuida. Estudios previos han mostrado que cuando la práctica se distribuye a lo largo del tiempo, es más efectiva a largo plazo que cuando se agrupa en un momento concreto, incluso cuando el tiempo total de estudio es equivalente (Cepeda et al., 2006). Este hallazgo, denominado comúnmente como efecto de espaciado, lleva siendo investigado desde hace más de 100 años y ha sido demostrado en cientos de estudios, por lo que es uno de los efectos más robustos en la psicología del aprendizaje (Carpenter et al., 2022).

Sin embargo, pese al gran número de estudios que han investigado el efecto de espaciado, la mayoría de los trabajos se han centrado en tareas de memorización, es decir, tareas que prueban el efecto de retención de información estudiada previamente, ya fuese de forma espaciada o agrupada (Zulkipli et al., 2012). Y pese a que el efecto de espaciado en memoria es robusto y muy importante en el contexto académico, en la vida cotidiana es poco probable encontrarse con circunstancias exactamente iguales a las experimentadas en el pasado. Por esto, podría ser más útil identificar patrones y principios generales a partir de experiencias previas y ser capaz de trasladar dichos conocimientos a situaciones nuevas (Kang y Pashler, 2012). Este tipo de aprendizaje es denominado inducción y se trata de un proceso cognitivo que involucra la abstracción de información general a partir de características generales de varios ejemplares específicos (Brunmair y Richter, 2019). El

aprendizaje inductivo es, por tanto, uno de los elementos clave de los que se sirven humanos y animales para aprender sobre lo que les rodea, estando presente en prácticamente todos los aspectos de su experiencia (Brunmair y Richter, 2019).

De hecho, la mayoría del conocimiento que adquieren las personas no llega a través de la instrucción formal, si no de generalizar ejemplos observados a lo largo de la vida (Kornell et al., 2010). Dicho esto, parece razonable pensar que potenciar la memorización de hechos específicos a menudo no es tan relevante como potenciar la inducción (Carvalho y Goldstone, 2014a). Mientras que la memorización pura nos permite retener cuales son los principales ríos de España y los nombres de los elementos de la tabla periódica, ser capaz de generalizar características distintivas a partir de una serie de ejemplares, permite que un bebé aprenda lo que es una silla mientras observa a sus padres hablar sobre sillas (Kornell y Bjork, 2008) o que un estudiante de historia del arte aprenda a identificar arquitectura colonial española a partir de características comunes entre edificios diferentes (Kornell et al., 2010).

En consecuencia, dado el rango de situaciones en las que influye, la investigación sobre aprendizaje inductivo se ha convertido en objeto de estudio en disciplinas como la neurociencia, la psicología cognitiva, la psicología educacional o la psicología social, entre otras (Brumair y Ritcher, 2019). Esto se da incluso en el desarrollo de la inteligencia artificial, especialmente con el aprendizaje automático o “machine learning”, donde la inducción es uno de los pilares sobre los que se sustenta esta tecnología (Brumair y Richter, 2019).

Dado que el aprendizaje inductivo se produce a lo largo del tiempo, es esperable que la secuencia y forma en la que se estudia la información afecte al aprendizaje (Kornell y Bjork, 2008). Kurtz y Hovland (1956) demostraron que la inducción de categorías no se beneficiaba del espaciado. Es más, afirmaban que el estudio agrupado de ejemplares

beneficiaría la inducción ya que, permitiría distinguir las similitudes de los diferentes elementos, mejorando la generalización. Además, algunos estudios sugirieron que, si bien el espaciado podía mejorar la memoria, podría ser perjudicial para el aprendizaje inductivo (Kornell y Bork, 2008).

Ante tales hallazgos, Kornell y Bjork (2008) hipotetizaron que los efectos de espaciado desaparecerían en tareas de inducción, puesto que este tipo de tareas requieren la exposición a estímulos diferentes y no repetición exacta. Para poner a prueba esta hipótesis llevaron a cabo una serie de experimentos. En su Experimento 1a, los participantes estudiaban cuadros de diferentes artistas bajo diferentes condiciones de aprendizaje. Cabe destacar que, en el trabajo original, los autores hablaban de condición de espaciado, pero debido a que introducían cuadros de diferentes autores de forma entremezclada, en la condición se está produciendo a la vez espaciado y entremezclado, frente a la condición en la que la presentación era masiva y en bloque. En adelante se hará referencia a las condiciones de espaciado-entremezclado y masivo-bloqueado.

En la condición de masivo-bloqueado se presentaban bloques de seis cuadros diferentes de un mismo autor. En el caso de la condición de espaciado-entremezclado cada bloque de seis cuadros consistía en una mezcla de cuadros de seis artistas diferentes, un cuadro por artista y bloque. Los participantes debían aprender las características generales del estilo de cada autor y se presentaban los nombres de los autores a la vez que veían cada uno de los cuadros, con el objetivo de que aprendieran a relacionar cada estilo con su autor.

Tras la fase de aprendizaje, los participantes realizaban una tarea de categorización con retroalimentación en la cual se les presentaban nuevas pinturas de los artistas que habían estudiado previamente. En cada ensayo de prueba, debían identificar el autor de cada cuadro. Tras su respuesta, los participantes recibían retroalimentación, que les indicaba si acertaban o

les corregía si se equivocaban. En contra de lo esperado, los resultados del experimento mostraron que el rendimiento en la fase de prueba era mejor para la condición de espaciado-entremezclado que para la condición de masivo-bloqueado. Este resultado se replicó con un diseño intrasujeto en el Experimento 1b.

Tras estos resultados, Kornell y Bjork (2008) argumentaron que la condición de espaciado-entremezclado podría haber obtenido mejores resultados puesto que la prueba requería recordar el nombre de cada autor además del estilo artístico. De esta forma, consideraron que la condición de masivo-bloqueado podría haber sido mejor para inducir el estilo artístico, pero que los resultados habían sido opacados por el efecto de espaciado en la memorización de los nombres de los artistas. Para probar esta posibilidad, llevaron a cabo el Experimento 2, que solamente difería del Experimento 1 en que empleaba en este caso una prueba de reconocimiento en vez de categorización. Durante la fase de prueba aparecían cuadros de los artistas que habían estudiado previamente y otros que funcionaban como distractores, con aspectos comunes, pero de artistas no estudiados. Los participantes debían responder indicando si creían que los cuadros eran de artistas ya estudiados o de nuevos artistas. De esta forma, la tarea solo requería inducir el estilo artístico de los autores, no siendo relevante para la prueba recordar la relación con el nombre del artista. En contra de lo hipotetizado, la proporción de aciertos fue superior en la condición de espaciado-entremezclado que en la condición de masivo-bloqueado.

Los hallazgos de Kornell y Bjork (2008) demostraron que los efectos de espaciado-entremezclado no se limitan a tareas de memorización, sino que aplica a condiciones de inducción. Dada la importancia de la discriminación de categorías en la vida cotidiana, los autores hipotetizaron que el efecto de espaciado-entremezclado en tareas de inducción podía darse de manera general (Kornell et al., 2010). Además, se empezaron a desarrollar varias

propuestas teóricas para el fenómeno del espaciado-entremezclado en inducción que pasamos a describir.

En un trabajo posterior, Kornell et al. (2010) propusieron la Hipótesis de la Atenuación de la Atención. Según esta hipótesis, los resultados no se deberían al beneficio por el efecto de espaciado-entremezclado, sino que se darían por un peor rendimiento en la condición de masivo-bloqueado. En un bloque la condición de masivo-bloqueado, tras estudiar varios ejemplares similares, los participantes podrían pensar que ya han aprendido lo suficiente sobre la categoría, por lo que reducirían su atención a siguientes elementos del bloque. Por el contrario, en el espaciado-entremezclado, estudian ejemplares diferentes entre sí, por lo que mantendrían la atención a lo largo del bloque (Kornell et al., 2010).

Como se comentó anteriormente, en diseños de entremezclado, los elementos de estudio son necesariamente espaciados, algunos trabajos han estudiado si el aumento del intervalo de tiempo entre categorías entremezcladas conlleva beneficios adicionales (Brunmair y Richter, 2019). Estos estudios han concluido que el espaciado entre exposiciones no necesariamente mejora los efectos de entremezclado (Kang y Pashler, 2012; Zulkiply y Burt, 2013b).

A raíz de estos argumentos, una pregunta potencial es si los beneficios del espaciado-entremezclado en este paradigma se deben precisamente al espaciado o al entremezclado, ya que en el experimento original ambos fenómenos son inseparables. Kornell y Bjork (2008) ya argumentaron que los beneficios del espaciado podrían deberse al hecho de que el espaciado implique necesariamente entremezclado. Es plausible que el entremezclado produzca una mejora del aprendizaje al ayudar a los participantes a discriminar las diferencias y similitudes de los estímulos (Noh et al. 2016). Al resaltar las diferencias entre los elementos de una

categoría, entremezclar elementos podría ser útil para discriminar entre diferentes categorías, especialmente cuando sean difíciles de diferenciar (Kornell et al., 2010).

Esta idea va en la línea de Hatala et al. (2003), que hipotetizaron que una de las claves en el aprendizaje de categorías está en la discriminación entre categorías. En un experimento, un grupo estudió ejemplos de diagnósticos de electrocardiogramas diferentes, entremezclados entre sí (lo que denominaron práctica contrastiva), mientras que otro grupo estudió ejemplos de diagnósticos del mismo tipo de forma agrupada (práctica no contrastiva). Los resultados indicaron que la práctica contrastiva fue superior a la práctica no contrastiva. Dicho de otra forma, entremezclar ejemplares fue más efectivo para discriminar entre categorías.

Kang y Pashler (2012), realizaron una serie de experimentos en los que intentaron aislar los efectos de entremezclado de los efectos de espaciado, encontrando que el entremezclado de cuadros producía mejor resultado por sí solo que el espaciado. Estos resultados fueron replicados por estudios posteriores (Carvalho y Goldstone, 2014b; Zulkiply y Burt, 2013b), que afirmaron que el factor clave en tareas de inducción sería el entremezclado y no el espaciado. Birnbaum et al. (2013) encontraron que el espaciado no solo no interfiere en las condiciones de entremezclado, sino que parece mejorar el rendimiento cuando se introduce espaciado en las condiciones de masivo.

La revisión de Chen et al. (2021) se centró en estas cuestiones, mostrando que a pesar de que frecuentemente son confundidos, el espaciado y el entremezclado son fenómenos con mecanismos cognitivos diferentes. De acuerdo con este trabajo, el espaciado potencia el aprendizaje permitiendo que se recupere información de la memoria de trabajo durante momentos de descanso, mientras que el entremezclado mejoraría el aprendizaje mediante la discriminación (Chen et al., 2021).

De forma paralela al desarrollo de las Hipótesis de Contraste Discriminativo, Carvalho y Goldstone (2015b) propusieron la Teoría de la Atención Secuencial como explicación complementaria. Según esta idea, durante la inducción de categorías, los participantes tienden a centrar su atención en las diferencias entre elementos de diferentes categorías y en las similitudes de elementos de una misma categoría (Carvalho y Goldstone, 2015a). Resultados posteriores han apoyado esta propuesta. Carvalho y Goldstone, (2017) realizaron un estudio con *eye-tracking* en el que describieron variaciones en la atención visual dependiendo del tipo de presentación. Según su explicación, durante el aprendizaje, los participantes comparan el ejemplar actual con el que apareció justo antes y, dependiendo de la condición de estudio, eso significará prestar atención a diferencias o similitudes entre los dos ejemplares. Si los elementos se presentan espaciados-entremezclados la atención de los participantes se dirigirá hacia las diferencias, fomentando la discriminación. Si, por el contrario, los elementos se presentan en una condición masiva-bloqueada los participantes atenderán a las similitudes entre elementos (Carvalho y Goldstone, 2015a). Además, cuanto más difícil sea diferenciar entre los elementos de una categoría, más se potenciaría el efecto de entremezclado (Carvalho y Goldstone, 2014b).

A medida que ha avanzado la investigación y que han aflorado estas teorías explicativas, el efecto de entremezclado ha ganado preponderancia, sustituyendo en la mayoría de los trabajos al efecto de espaciado en tareas de aprendizaje inductivo. Esta nueva perspectiva, ha desembocado en una gran cantidad de estudios y replicaciones estudiando los diferentes moduladores del efecto observado por Kornell y Bjork (2008). Una de las cuestiones que ha suscitado más interés es la generalización del efecto de entremezclado en inducción en diferentes tipos de materiales de estudio. Se han obtenido efectos de espaciado-entremezclado en inducción en tareas con estilos artísticos (Guzman-Munoz, 2017; Kang y Pashler, 2012; Verkoeijen y Bouwmeester, 2014; Zulkipli y Burt, 2013a), fotografías de

pájaros y mariposas (Birnbaum et al., 2013; Wahlheim et al., 2011), imágenes de radiología (Rozenstein et al., 2016; Thompson y Hughes, 2023), imágenes de tipos de rocas (Do y Thomas, 2023; Whitehead et al., 2022), categorías de estructuras químicas (Eglington y Kang, 2017), materiales de texto, tanto por vía visual como auditiva (Zulkipli et al., 2012; Zulkipli, 2013), aprendizaje de verbos de lenguas extranjeras (Pan et al., 2018, 2019; Schweppe et al., 2025), operaciones matemáticas y estadísticas, con resultados mixtos (Nemeth et al., 2019; Rowlandson y Simpson, 2023; Sana et al., 2017; Taylor y Rohrer, 2010), habilidades musicales (Simmons, 2012; Wong et al., 2020, 2021) y secuencias motoras (Lin et al., 2016).

El entremezclado ha obtenido resultados mixtos en materiales matemáticos, por lo que es conveniente ser prudentes con la generalización del efecto (Braithwaite y Hall, 2024). Algunos estudios sugieren que hay materiales y contenidos que no solo son más difíciles de aprender a partir del espaciado-entremezclado, sino que algunos se benefician más del masivo-bloqueado (Brunmair y Richter, 2019; Yan y Sana, 2021).

Otro factor modulador que ha despertado interés es la edad de los participantes. El efecto de entremezclado se ha replicado con diferentes materiales en niños de entre 10 a 14 años (Nemeth et al., 2019; Rohrer et al., 2020; Rowlandson y Simpson, 2023; Taylor y Rohrer, 2010). Vlach et al. (2008) no encontraron efectos de entremezclado en niños de tres años. Varios estudios replicaron el efecto en poblaciones de adultos mayores (Kornell et al., 2010; Lin et al., 2016; Wahlheim et al., 2011), aunque con resultados poco concluyentes.

El aprendizaje obtenido a partir de este paradigma puede ser probado inmediatamente tras el entrenamiento o también puede introducirse una demora entre el entrenamiento y la prueba. Hay pocos estudios que hayan estudiado el intervalo de retención en este tipo de tareas. Se han realizado experimentos con demoras de 1 a 2 días con diferentes materiales

visuales (Carvalho y Goldstone, 2014b; Eglinton y Kang, 2017; Whitehead et al., 2021). También se han realizado experimentos con una semana de demora, usando materiales de texto, algunos de los cuales proporcionaba retroalimentación durante las fases de prueba (Pan et al., 2018; Schweppe et al., 2025; Zulkipli, 2013). En 2013, Zulkipli y Burt realizaron una extensión del Experimento 1a de Kornell y Bjork (2008) usando cuadros de artistas e introduciendo una semana de demora entre el entrenamiento y la prueba. Durante la prueba se les proporcionaba retroalimentación a los participantes, tal y como ocurría en el trabajo original. Hasta el momento, el estudio de Zulkipli y Burt (2013a) es el único trabajo conocido que ha extendido el Experimento 1a de Kornell y Bjork (2008) a largo plazo.

Este paradigma, pese a haber recibido mucha atención en los últimos años, tiene algunas limitaciones. Los Experimentos 1a y 1b de Kornell y Bjork (2008) realizaron una fase de prueba en la cual los ensayos eran seguidos de retroalimentación. Esta estrategia ha sido empleada por otros estudios posteriores (Li et al., 2025; Zulkipli et al., 2012). Sin embargo, esto plantea el problema de que la fase de prueba se convierte también en una oportunidad de aprendizaje. Por ello, puede darse el caso de que el efecto de espaciado-entremezclado quede contaminado por la retroalimentación. Realizar una fase de prueba con ensayos de categorización sin retroalimentación facilitaría caracterizar de manera más precisa el efecto de las condiciones en la fase de entrenamiento.

En el Experimento 2 del mismo artículo se realizó una prueba de reconocimiento con distractores para evaluar si los participantes aprendían el estilo de los artistas presentados durante el entrenamiento. Las medidas recogidas eran la tasa de acierto y las falsas alarmas, es decir, se registraba cuando los participantes reconocían correctamente las pinturas de los autores estudiados y cuando los participantes clasificaban erróneamente un distractor como un artista familiar. Este tipo de medidas en tareas de discriminación Sí-No a menudo fallan al medir la discriminabilidad cuando hay tasas elevadas de acierto y no emplean toda la

información disponible (Green y Swets, 1966). Para paliar este problema, puede calcularse un $\log d$, basada en transformaciones de matrices de respuesta de Teoría de Detección de Señales, que son medidas menos contaminadas por sesgos de respuesta y funcionan mejor en tareas con pocos ensayos (Brown y White, 2005). Otra limitación del Experimento 2 de Kornell y Bjork (2008), es que sólo hay constancia de una replicación (Verkoeijen y Bouwmeester, 2014) por lo que los resultados podrían no ser fiables y son necesarias más repeticiones.

En el presente trabajo se describen los resultados realizados con el objetivo de replicar los experimentos originales y además paliar estas posibles limitaciones. En el Experimento 1 se realizó una replicación del Experimento 1a de Kornell y Bjork (2008) manteniendo la retroalimentación durante la fase de prueba. Esta replicación permite confirmar si el efecto reportado en el experimento original aparece con una muestra diferente, aportando evidencia convergente. Se hipotetiza que la condición de espaciado-entremezclado obtendrá mejores resultados que la condición de masivo-bloqueado y que el rendimiento de los participantes aumentará a lo largo de los bloques de la fase de prueba debido a la retroalimentación.

En el Experimento 2, se realizó una replicación del Experimento 2 de Kornell y Bjork (2008). El objetivo es encontrar evidencia convergente y mejorar la precisión de la discriminabilidad usando medidas como $\log d$. Se hipotetiza que la condición de espaciado-entremezclado obtendrá mejores resultados que la condición de masivo-bloqueado. Replicar los Experimentos 1 y 2, nos permitió garantizar que el funcionamiento de las tareas era correcto y el efecto era robusto, lo que permitió realizar los Experimentos 3 y 4.

El Experimento 3, extendió el experimento 1a original, eliminando la retroalimentación en la fase de prueba. Eliminar la retroalimentación es fundamental para aislar el efecto de entremezclado durante la prueba, sin que la propia prueba se convierta en

una oportunidad para aprender. Esta replicación nos permite comparar los resultados con el Experimento 1 y evaluar hasta qué punto se daría el efecto en una prueba más directa de inducción. Se hipotetizó que la condición espaciado-entremezclado obtendría mejores resultados también en este caso.

El Experimento 4 consistió en una extensión del Experimento 1a de Kornell y Bjork (2008) y del propio Experimento 3 de este trabajo, realizando una tarea de categorización sin retroalimentación con una semana de demora entre la fase de entrenamiento y la fase de prueba. El objetivo era observar los efectos de entremezclado-espaciado a largo plazo eliminando la retroalimentación. Hasta la fecha solo Zulkipli y Burt (2013a) han extendido la tarea original a largo plazo, pero lo hicieron manteniendo la retroalimentación durante la prueba. El presente experimento sería el primero en extender el Experimento 1a de Kornell y Bjork (2008) a largo plazo sin retroalimentación. Se hipotetizó que la condición espaciado-entremezclado obtendría mejores resultados que la condición masivo-bloqueado.

Método

Experimento 1

Participantes

Un total de 20 estudiantes del Grado de Psicología de la Universidad de Málaga (15 mujeres y 5 hombres; Edad media = 25.9, $SD = 9.37$, rango de 21-62 años) que participaron a cambio de puntuación en una asignatura. La selección de los participantes se realizó mediante un muestreo de conveniencia. Para establecer el tamaño de la muestra se realizó un análisis a priori a partir de los resultados del Experimento 1a de Kornell y Bjork (2008), en el que se reportó un tamaño del efecto de $d = 0.99$. Con un nivel de significancia de $\alpha = 0.05$ y una potencia del 95%, se estimaron 16 participantes (Faul et al., 2009). En todos los experimentos, los participantes dieron el consentimiento informado previo y el proyecto

recibió un informe positivo del Comité de Ética de la Universidad de Málaga, con código 186-2023-H.

Materiales

Los materiales fueron los empleados por Kornell y Bjork (2008), los cuales se encuentran disponibles en su página web¹. Se componen de 10 cuadros paisajísticos de 12 artistas poco conocidos para el público general (Georges Braque, Henri-Edmond Cross, Judy Hawkins, Philip Juras, Ryan Lewis, Marilyn Mylrea, Bruno Pessani, Ron Schlorff, Georges Seurat, Ciprian Stratulat, George Wexler y Yie Mei). Las imágenes de los cuadros se presentaron centradas en pantallas LED de resolución 1920 x 1080. Las imágenes aparecieron en tamaño original. Las dimensiones variaron entre 328 y 622 píxeles de ancho y entre 311 y 498 píxeles de alto.

Procedimiento

La tarea se componía de tres fases, una de entrenamiento, una tarea distractora y una fase de prueba. La segunda fase perseguía asegurar que la información de la fase de entrenamiento no se mantuviese en memoria de trabajo y no se produjera un repaso de los materiales presentados antes de la prueba, eliminando así también posibles efectos de recencia (Bjork y Whitten, 1974).

Los artistas se dividieron en dos condiciones de presentación. La mitad de los artistas fueron asignados a la condición de masivo-bloqueado (MB² por sus siglas en inglés). La otra mitad fue asignada a la condición de espaciado-entremezclado (SI³ por sus siglas en inglés). Esta fase se dividió en 12 bloques, de los que seis fueron bloques de presentación de artistas de la condición MB y la otra mitad de artistas de la condición SI. En un bloque MB los seis

¹ <http://sites.williams.edu/nk2/stimuli/>

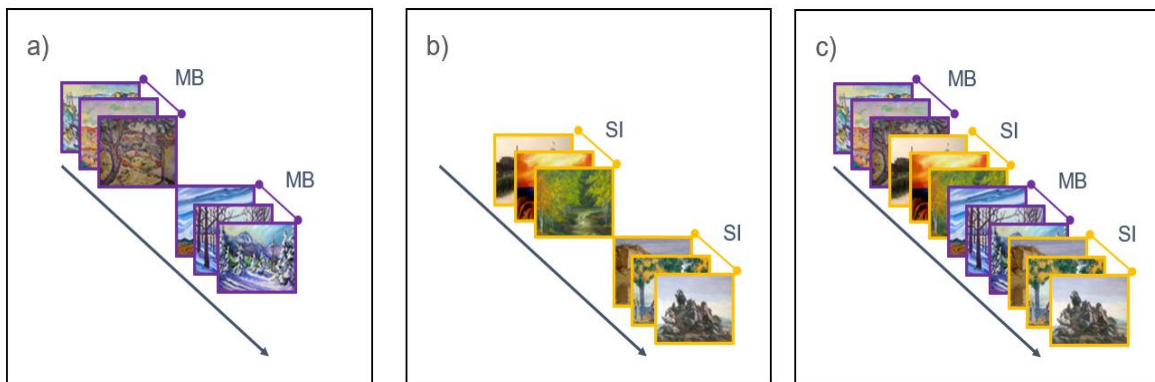
² Del Inglés Massed-Blocked

³ Del Inglés Spaced-Interleaved

cuadros pertenecían al mismo artista. En cambio, en un bloque SI cada uno de los seis cuadros pertenecía a un artista diferente. De esta manera, los cuadros de cada artista MB solo aparecían en un bloque, muy cerca en el tiempo y sin cuadros de otros artistas entre ellos. Por el contrario, un cuadro de cada artista de la condición SI se presentaba en cada bloque SI, mezclados con los cuadros de otros artistas y más separados temporalmente, ya que hasta el siguiente bloque SI no verían otra aparición de un cuadro del mismo artista. El orden de aparición de los bloques fue SMMSSMMSSMMS o MSSMMSSMMSSM contrabalanceado entre participantes. En la Figura 1 se puede ver una representación gráfica de una versión simplificada de la fase de entrenamiento.

Figura 1

Ejemplos de bloques MB (a), bloques SI (b) y con ambos bloques (c) en la fase de entrenamiento



Nota: En (a) pueden observarse ejemplos de cuadros agrupados en la condición MB, todos los cuadros que aparecen en cada bloque de dicha condición pertenecen a un autor determinado. En (b) se observan ejemplares entremezclados en la condición SI, los cuadros que aparecen en dicha condición pertenecen a autores diferentes. En (c) se aprecia un ejemplo del diseño completo, en el que los bloques MB y SI se suceden. Cabe aclarar que los colores usados la figura tienen como fin facilitar al lector la interpretación del diseño y en ningún

caso aparecieron durante la presentación a los participantes. Además, durante la presentación, se mostraron seis cuadros en cada bloque y no tres, como se ejemplifica en la figura.

Los participantes se situaron frente a un ordenador y se les informó de la naturaleza y el procedimiento del experimento. La fase de entrenamiento consistió en 72 ensayos de tres segundos de duración, mostrándose en la pantalla un solo cuadro por ensayo. Los ensayos se dividían en 12 bloques de seis cuadros. Seis de esos bloques fueron MB, mientras que los seis bloques restantes fueron SI. El orden de aparición de los cuadros dentro del bloque y los artistas asignados a cada condición fueron aleatorizados para cada participante.

Tras finalizar la fase de entrenamiento, se introdujo una tarea distractora de 15 segundos de duración, durante la cual los participantes debían contar hacia atrás de tres en tres desde 547 cada vez que aparecía un círculo blanco en pantalla. Al finalizar el conteo, se les mostraba en pantalla una serie de cifras y los participantes debían indicar cuál de ellas era la que más se acercaba a la que habían calculado mentalmente. El objetivo de esta tarea fue generar una alta carga cognitiva y evitar el repaso controlado de la información adquirida durante la fase de entrenamiento.

Tras la finalización de la tarea distractora, se inició la fase de prueba. El objetivo de la fase fue realizar una prueba de inducción usando cuadros nuevos de los artistas estudiados en la fase de entrenamiento. Es decir, se evaluó la capacidad de los participantes de clasificar cuadros nuevos de los artistas previamente estudiados. La fase de prueba consistió en 48 ensayos divididos en cuatro bloques de 12 cuadros. Durante cada ensayo, se presentaron cuadros nuevos que no habían aparecido durante la fase de entrenamiento de cada uno de los artistas. Junto a cada cuadro, aparecía una lista con los nombres de todos los autores, etiquetados junto a una tecla. Es necesario destacar que se eliminó la opción de “No lo sé”, presente en el diseño original, con el objetivo de forzar a los participantes a seleccionar una

opción de respuesta. El orden de aparición en la lista de los autores, así como la tecla asignada a cada uno fue aleatorizado.

La tarea de los participantes fue seleccionar al autor que pensaban que había realizado el cuadro que aparecía en pantalla a partir del conocimiento obtenido en la fase de entrenamiento. Tras cada respuesta se proporcionó retroalimentación al participante. La palabra “Correcto” apareció cada vez que se seleccionaba al autor correcto, mientras que cada error se seguía del nombre correcto del autor del cuadro. La retroalimentación se mantuvo en pantalla durante un segundo y el tamaño de la fuente fue de 30. Los experimentos del presente trabajo fueron diseñados utilizando PsychoPy (versión v2024.2.1post4) (Pierce et al., 2019).

Análisis de Datos

Los análisis estadísticos fueron realizados con JASP (versión 0.19.3) (JASP Team, 2025). Previo a las pruebas paramétricas, se verificó la normalidad de la distribución de los residuos mediante pruebas de Shapiro-Wilk y la esfericidad con la prueba de Mauchly. Además, se revisaron las gráficas Q-Q de la distribución de los residuos. Para comparar las medias de las condiciones se usaron pruebas t para muestras relacionadas y ANOVA de medidas repetidas. El nivel de significancia elegido para todas las pruebas estadísticas fue de 0.05. En cuanto al tamaño del efecto, se optó por usar g_{av} de Hedges para las pruebas t, puesto que se trata de una versión corregida y en general más precisa que d de Cohen, mientras que para el ANOVA se hizo uso de ω^2 , dado que su utilización ha sido recomendada para corregir los posibles sesgos que presenta η^2 , especialmente en muestras pequeñas (Lakens, 2013). Por último, se realizaron análisis bayesianos equivalentes a la pruebas t, calculándose el factor Bayes a favor de la hipótesis alternativa (BF_{10}) empleando una

distribución a priori de Cauchy con una $r = 1/\sqrt{2}$ y comprobando además la robustez a través de distintos valores de r (Van Doorn et al., 2021).

Resultados

La Figura 2 muestra las proporciones medias de acierto obtenidas en las condiciones SI y MB en la fase de prueba. La condición SI obtuvo mejores resultados ($M = 0.49$, $SD = 0.23$, 95% IC [0.39, 0.59]) que la condición MB ($M = 0.34$, $DT = 0.13$, 95% IC [0.28, 0.40]). La Figura 3 muestra la distribución de las diferencias individuales entre las condiciones SI y MB.

Para evaluar si las diferencias entre las medias eran significativas se realizó una prueba t para muestras relacionadas. Los resultados indicaron una diferencia significativa, $t(19) = 3.80$, $p = .001$, M_{diff} 95% IC [0.07, 0.23], y un tamaño del efecto de g_{av} de Hedges = 0.79, lo que sugiere un tamaño del efecto de moderado a grande. Se realizaron análisis bayesianos adicionales (Figura 11, panel a).

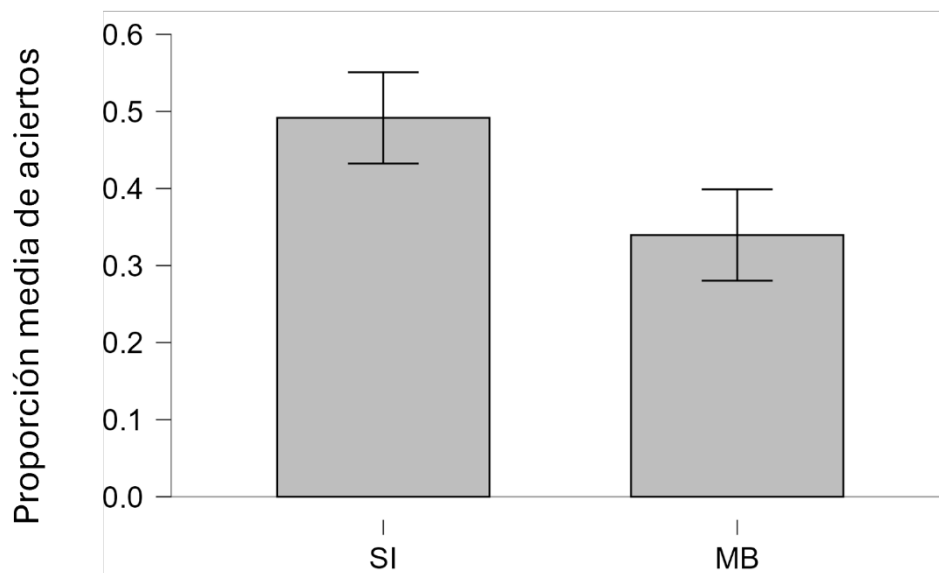
Los resultados obtenidos en el Experimento 1a de Kornell y Bjork (2008) mostraron proporciones de acierto más altas en la condición de espaciado-entremezclado durante la prueba, pero observaron que los participantes mejoraban a lo largo de los bloques de entrenamiento debido al efecto de la retroalimentación durante la tarea. Para comprobar el efecto de la retroalimentación en el presente experimento, se realizó un ANOVA de medidas repetidas de dos factores para examinar la interacción entre las condiciones de entrenamiento (SI vs MB) y los cuatro bloques que componían la fase de prueba.

Se encontró un efecto principal en las condiciones de presentación, $F(1, 19) = 14.44$, $p = .001$, $\omega^2 = 0.137$, no se encontraron efectos principales en los bloques de la tarea, $F(3, 57) = 0.907$, $p = .443$, $\omega^2 = .000$. Pese a que se proporcionó retroalimentación, el rendimiento no mejoró a lo largo de los bloques, $F(3, 57) = 0.78$, $p = .508$, $\omega^2 = .000$ (Figura 4).

Los análisis bayesianos correspondientes se resumen en la Figura 11 (panel a). De manera general se ha replicado el efecto principal descrito por Kornell y Bjork (2008), pero no así el efecto de la retroalimentación durante la fase de prueba. En la discusión se propone una posible explicación de esta diferencia.

Figura 2

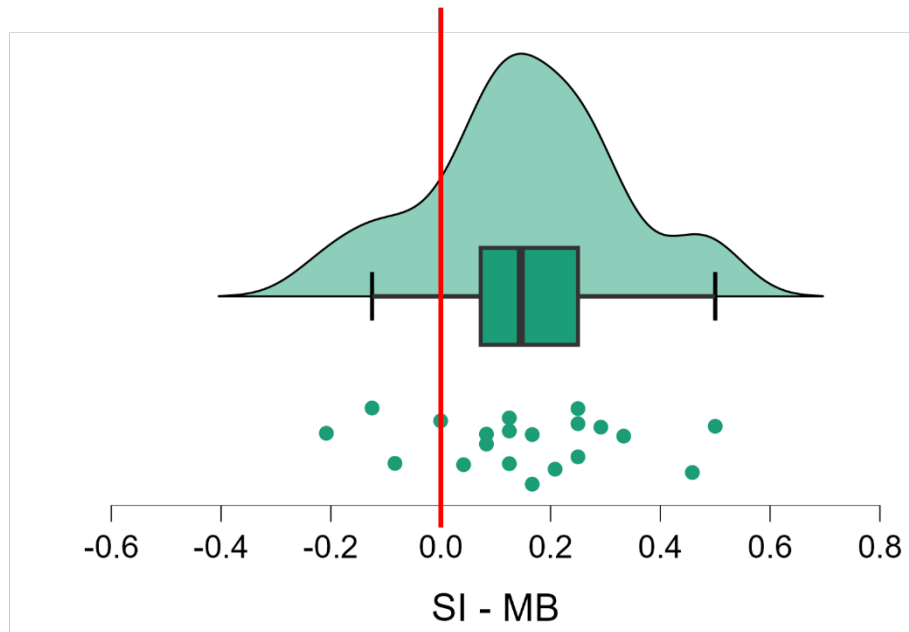
Proporción media de aciertos en la fase de prueba en función de las condiciones SI y MB en el Experimento 1.



Nota: Proporción media de aciertos obtenida en la fase de prueba en función de las condiciones de estudio (SI vs MB). Las barras de error indican intervalos de confianza del 95%

Figura 3

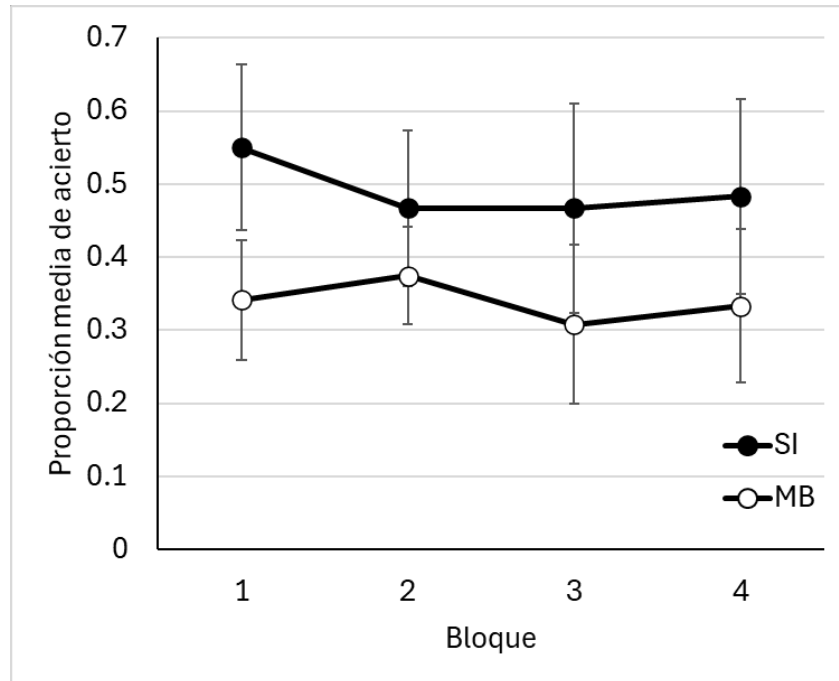
Distribución de las diferencias individuales entre las condiciones SI y MB durante fase de prueba en el Experimento 1



Nota: Distribución de las diferencias entre los participantes. Cada punto individual corresponde a un participante. Se obtuvieron restando las puntuaciones obtenidas en la condición MB a las obtenidas en la condición SI. Puesto que la mayoría de los puntos están situados por encima de 0, se observa mejor desempeño en la condición SI que en MB. La curva superior representa la distribución de las diferencias entre SI y MB, el diagrama de caja representa la mediana y los cuartiles (Q1 y Q3) y las líneas negras indican el rango intercuartílico de dichas puntuaciones.

Figura 4

Proporción media de aciertos durante los cuatro bloques de la fase de prueba en las condiciones SI y MB en el Experimento 1



Nota: Rendimiento promedio a lo largo de los bloques de la tarea de categorización con retroalimentación. Las barras de error representan intervalos de confianza del 95%.

Experimento 2

Participantes

Un total de 94 estudiantes del Grado de Psicología de la Universidad de Málaga (81 mujeres, 10 hombres y 3 no binarios; Edad media = 21, $SD = 5.74$, rango de 18-59 años) que participaron a cambio de puntuación en una asignatura. La selección de los participantes se realizó a partir de un muestreo de conveniencia. Para determinar el tamaño muestral se tuvieron en cuenta varias consideraciones. Se observó que el tamaño del efecto obtenido por Kornell y Bjork (2008) en su Experimento 2 fue menor al descrito para el Experimento 1. Además, en el presente experimento se realizaron análisis diferentes a los reportados en el

trabajo original. Se usó un log d , que se esperaba que fuese más sensible que el análisis original. Por ello, ante la falta de seguridad del tamaño muestral necesario, se optó por reclutar un número similar de participantes al reportado en el Experimento 2 de Kornell y Bjork (2008).

Materiales

Los materiales consistieron en el mismo conjunto de cuadros que se usaron en el Experimento 1. Además, se emplearon cuatro cuadros distractores adicionales por cada artista en la fase de prueba. Los cuadros distractores, de distintos artistas, se eligieron en base a su similitud con los cuadros de los 12 artistas presentados en la fase de entrenamiento y fueron los empleados por Kornell y Bjork (2008).

Procedimiento

La fase de entrenamiento y tarea distractora fueron la mismas que en el Experimento 1. Tras la tarea distractora, los participantes realizaron la fase de prueba, donde se empleó una tarea de reconocimiento idéntica a la del Experimento 2 de Kornell y Bjork (2008). Se compuso de cuatro bloques de 24 cuadros. Cada uno de los bloques contuvo 12 cuadros objetivo (es decir, cuadros nuevos de artistas estudiados en la fase de entrenamiento) y 12 cuadros distractores, de artistas no presentados en la primera fase. Los cuadros se presentaron sin etiquetas verbales, para que los participantes solo tuvieran que aprender los estilos de los artistas y no sus nombres. En cada ensayo de la fase de prueba se presentó un cuadro sin la etiqueta del nombre del autor. Los participantes debían pulsar la tecla asociada a la respuesta de estímulo familiar cuando se presentaran cuadros de artistas que hubieran aparecido en la fase de entrenamiento y pulsar la tecla de respuesta no familiar cuando se presentaran cuadros de artistas que no hubieran estudiado. Los participantes no recibieron retroalimentación durante la tarea.

Resultados

Siguiendo el protocolo de análisis del Experimento 2 de Kornell y Bjork (2008), solo se realizó el análisis de los dos primeros bloques, puesto que los ensayos de una tarea de reconocimiento son también oportunidades de aprendizaje, y la exposición a nuevos distractores puede contaminar el recuerdo de la fase de entrenamiento durante el paso de los bloques y empeorar progresivamente la precisión de los participantes durante el reconocimiento.

En este tipo de tareas, las medidas habituales basadas en la proporción de aciertos pueden fallar para estimar la capacidad de discriminación. Para abordar este problema, se usan medidas como d' y $\log d$. Estas medidas transforman las frecuencias de los cuatro tipos de respuestas (aciertos, errores, rechazos correctos y falsas alarmas) en una matriz de detección de señales que no está contaminada por sesgos de respuesta, no tienen efectos de techo y facilitan la interpretación de las diferencias entre condiciones (Brown y White, 2005). El uso de d' está muy extendido, pero $\log d$ es más fácil de calcular, su base teórica trata el sesgo de respuesta de manera más realista y es mejor en caso de tener pocas medidas (Brown y White, 2005).

Por todo ello, se usó $\log d$ como variable dependiente. Los valores para cada participante y condición se obtuvieron a partir de la siguiente ecuación:

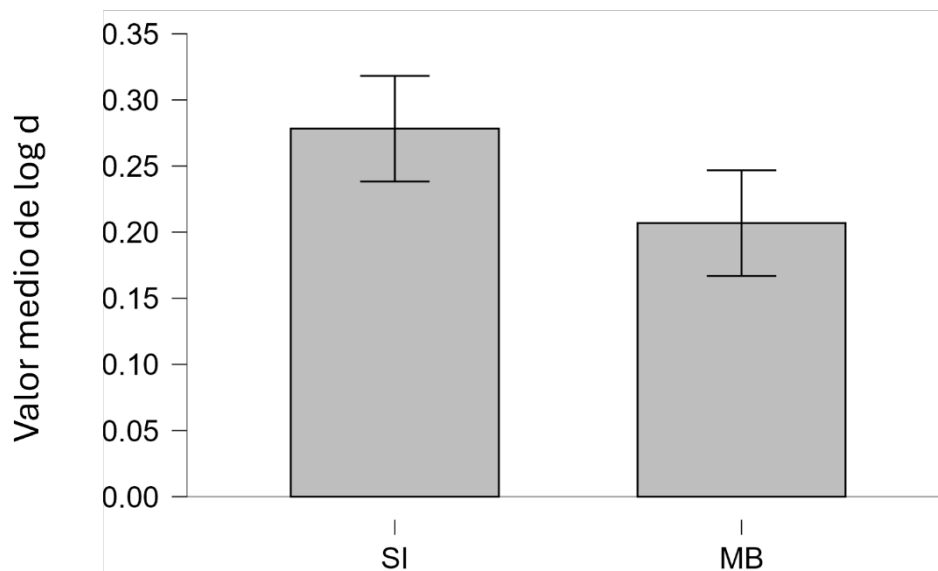
$$\log d = \frac{1}{2} * \log_{10} \left(\frac{\text{aciertos}}{\text{errores}} * \frac{\text{rechazos correctos}}{\text{falsas alarmas}} \right).$$

El problema de este tipo de medidas es que pueden volverse infinitas si algunos de los valores de la matriz, alcanza valores de 0, esto puede ocurrir si los participantes son muy precisos o cuando hay pocos ensayos. Para solucionar esto, se añade una constante de corrección de 0.5 a las cuatro celdas de la matriz, lo que evita la sobreestimación y que se alcancen valores anómalos (Brown y White, 2005).

La Figura 5 muestra los valores medios de $\log d$ obtenidos en las condiciones SI y MB en la fase de prueba. La condición SI obtuvo mejores resultados ($M = 0.28$, $SD = 0.18$, 95% IC [0.24, 0.32]) que la condición MB ($M = 0.21$, $DT = 0.21$, 95% IC [0.16, 0.25]). La Figura 6 muestra la distribución de las diferencias individuales entre las condiciones SI y MB.

Figura 5

Valor medio de $\log d$ en la fase de prueba en función de las condiciones SI y MB en el Experimento 2



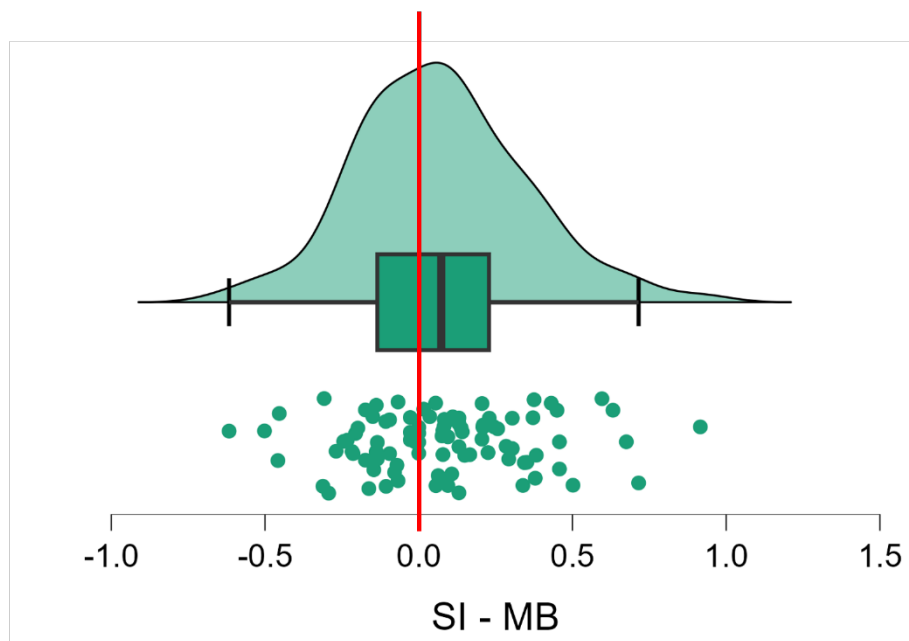
Nota: Valor medio de $\log d$ obtenido en la fase de prueba en función de las condiciones de estudio (SI vs MB). Las barras de error indican intervalos de confianza del 95%.

Para comparar el $\log d$ entre la condición SI y la condición MB se realizó una prueba t para muestras relacionadas. Los resultados indicaron diferencias significativas, $t(93) = 2.51$, $p = .007$, M_{diff} 95% IC [0.01, 0.12], y un tamaño del efecto de g_{av} de Hedges = 0.35, lo que

sugiere un tamaño del efecto de pequeño a mediano. Los resultados de los análisis bayesianos del experimento se encuentran resumidos en la Figura 11 (panel b).

Figura 6

Distribución de las diferencias individuales entre las condiciones SI y MB durante fase de prueba en el Experimento 2



Nota: Distribución de las diferencias entre los participantes. Cada punto individual corresponde a un participante. Se obtuvieron restando las puntuaciones obtenidas en la condición MB a las obtenidas en la condición SI. Puesto que la mayoría de los puntos están situados por encima de 0, se observa mejor desempeño en la condición SI que en MB. La curva superior representa la distribución de las diferencias entre SI y MB, el diagrama de caja representa la mediana y los cuartiles (Q1 y Q3) y las líneas negras indican el rango intercuartílico de dichas puntuaciones.

Experimento 3

Participantes

Un total de 47 estudiantes del Grado de Psicología de la Universidad de Málaga (42 mujeres y 5 hombres; Edad media = 20.7, $SD = 0.83$, rango de 20-24 años) que participaron a cambio de puntuación en una asignatura. El muestreo se realizó por conveniencia. Para obtener el tamaño del efecto deseado se buscaban al menos a 16 participantes, tal y como se estimó en el Experimento 1. Sin embargo, puesto que la participación se ofreció a estudiantes a cambio de puntuación adicional en una asignatura, la muestra superó la estimación prevista.

Materiales

Los materiales consisten en el mismo conjunto de cuadros se usaron en el Experimento 1.

Procedimiento

El procedimiento fue el mismo que se siguió en el Experimento 1, salvo por la eliminación de la retroalimentación durante la fase de prueba.

Resultados

Las proporciones medias de acierto obtenidas en las condiciones SI y MB se muestran en la Figura 7. Los participantes obtuvieron mayores proporciones medias de acierto en la condición SI ($M = 0.46$, $SD = 0.21$ 95% IC [0.40, 0.53]) que en la condición MB ($M = 0.31$, $SD = 0.16$, 95% IC [0.264, 0.36]). La Figura 8 muestra la distribución de las diferencias individuales entre las condiciones SI y MB.

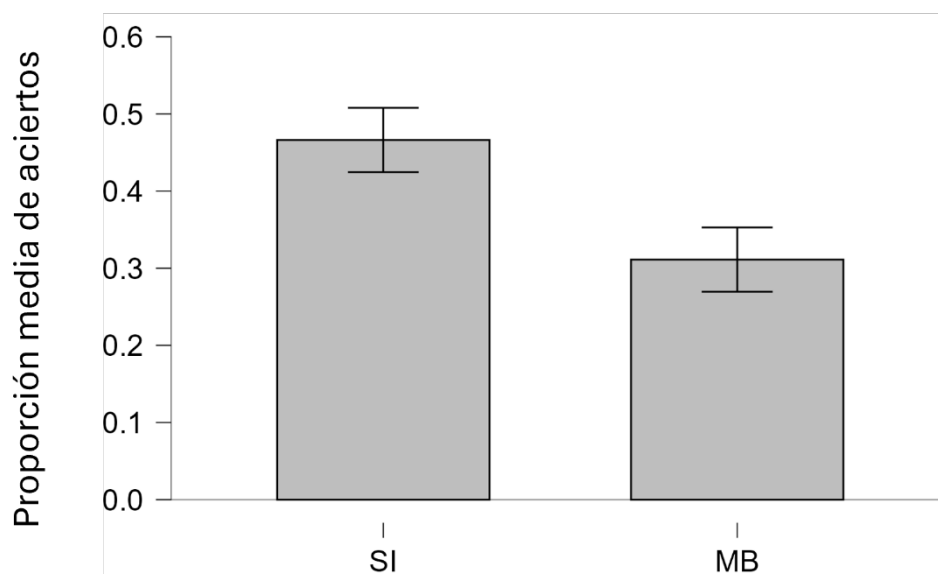
Los resultados indican diferencias significativas entre condiciones, $t(46) = 5.294$, $p = .001$, M_{diff} 95% IC [0.09, 0.21], g_{av} de Hedges = 0.82, lo que sugiere un tamaño del efecto de

moderado a grande. Los análisis bayesianos mostraron resultados equivalentes y están resumidos en la Figura 11 (panel c).

Los resultados extienden los hallazgos del Experimento 1. Tras la eliminación de la retroalimentación a la respuesta de los participantes sigue observándose un efecto robusto. Las medias del Experimento 1 son ligeramente superiores en ambas condiciones, lo que podría atribuirse al efecto de la retroalimentación.

Figura 7

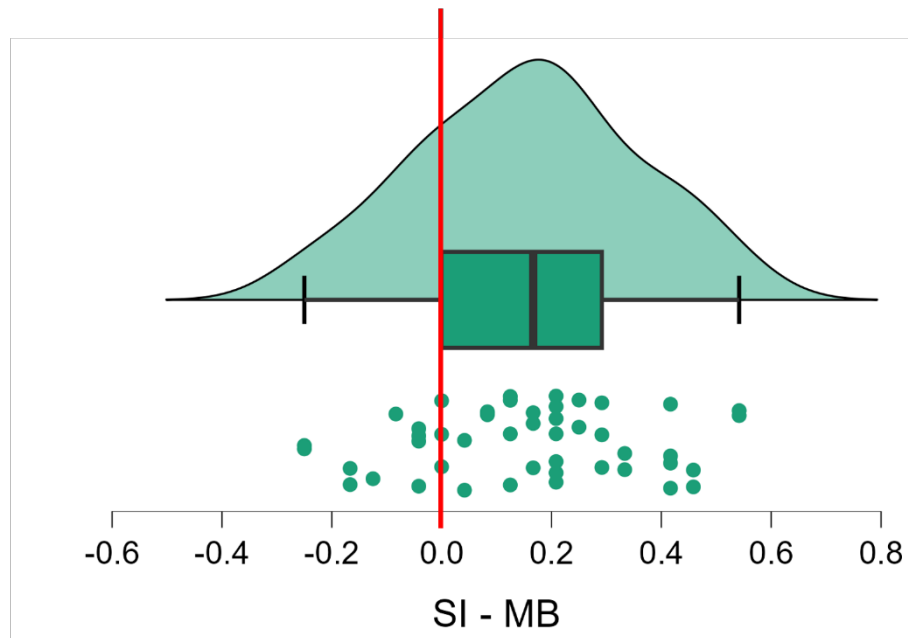
Proporción media de aciertos en la fase de prueba en función de las condiciones SI y MB en el Experimento 3



Nota: Rendimiento promedio obtenido en la fase de prueba en función de las condiciones de estudio (SI vs MB). Las barras de error indican intervalos de confianza del 95%

Figura 8

Distribución de las diferencias individuales entre las condiciones SI y MB durante fase de prueba en el Experimento 3



Nota: Distribución de las diferencias entre los participantes. Cada punto individual corresponde a un participante. Se obtuvieron restando las puntuaciones obtenidas en la condición MB a las obtenidas en la condición SI. Puesto que la mayoría de los puntos están situados por encima de 0, se observa mejor desempeño en la condición SI que en MB. La curva superior representa la distribución de las diferencias entre SI y MB, el diagrama de caja representa la mediana y los cuartiles (Q1 y Q3) y las líneas negras indican el rango intercuartílico de dichas puntuaciones.

Experimento 4

Participantes

Un total de 39 estudiantes del Grado de Logopedia de la Universidad de Málaga (34 mujeres y 5 hombres; Edad media = 19.9, $SD = 1.71$, rango de 18-24 años) que participaron a

cambio de puntuación en una asignatura. Para obtener el tamaño del efecto deseado se buscaban al menos a 16 participantes, tal y como se estimó en el Experimento 1. Sin embargo, puesto que la participación se ofreció a estudiantes a cambio de puntuación adicional en una asignatura, la muestra superó la estimación prevista.

Materiales

Los materiales consisten en el mismo conjunto de cuadros se usaron en el Experimento 1.

Procedimiento

El procedimiento fue el mismo que se siguió en el Experimento 1, salvo por la inclusión de una demora de una semana entre la fase de entrenamiento y la fase de prueba. De esta forma, el experimento tuvo dos sesiones. Una primera sesión en la que los participantes realizaron la fase de entrenamiento y la tarea distractora, y una segunda sesión siete días después, en la que tuvo lugar la fase de prueba con una tarea de categorización sin retroalimentación, idéntica a la realizada en el Experimento 3.

Resultados

Las proporciones de acierto en las condiciones SI y MB a largo plazo se muestran en la Figura 9. Los resultados obtenidos van en la línea de los resultados anteriores, los participantes obtuvieron mayores proporciones de acierto en la condición SI ($M = 0.24$, $SD = 0.15$, 95% IC [0.19, 0.30]) que en la condición MB ($M = 0.13$, $SD = 0.09$, 95% IC [0.09, 0.16]). La Figura 10 muestra la distribución de las diferencias individuales entre las condiciones SI y MB.

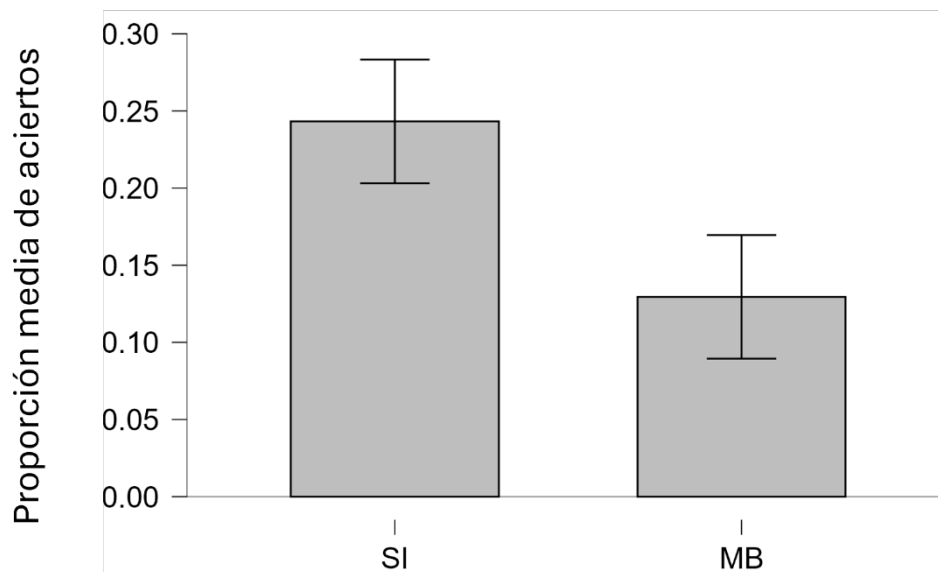
Los resultados indican diferencias significativas entre condiciones, $t(36) = 4.07$, $p = .001$, M_{diff} 95% IC [0.05, 0.17], g_{av} de Hedges = 0.88, lo que sugiere un tamaño del efecto de

moderado a grande. Se obtuvo resultados consistentes en los análisis bayesianos realizados (Figura 11, panel d).

Los resultados replican y extienden los hallazgos de Zulkipli y Burt (2013a). Este trabajo extendió el diseño original de Kornell y Bjork (2008) y encontró mejores resultados en la condición SI frente a MB tras una semana de demora entre la fase de entrenamiento y la fase de prueba. El presente experimento es el primero en realizar la tarea original a largo plazo y sin retroalimentación.

Figura 9

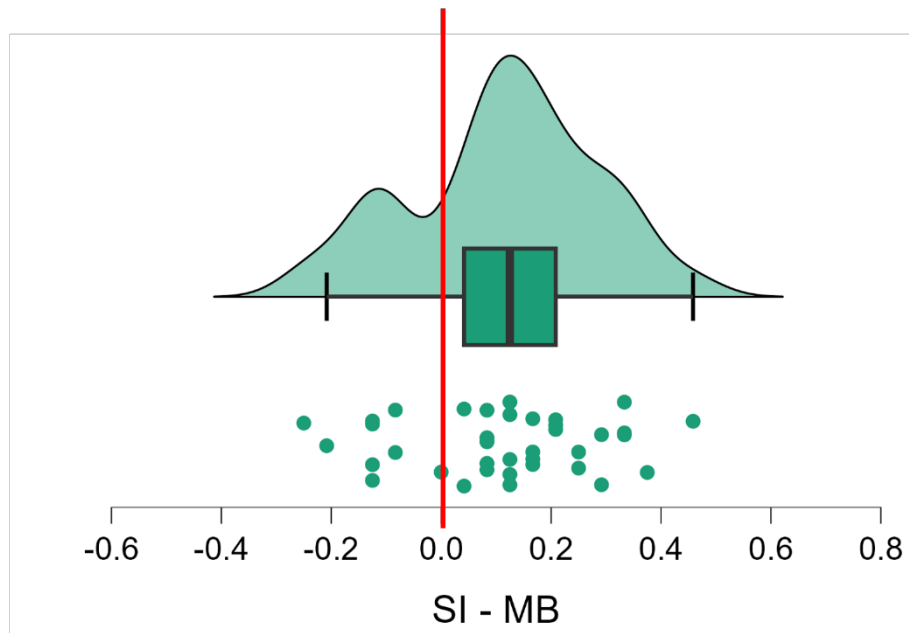
Proporción media de aciertos en las condiciones SI y MB con una demora de una semana entre la fase de entrenamiento y la fase de prueba en el Experimento 4



Nota: Proporción media de aciertos obtenidos en la fase de prueba en función de las condiciones de estudio (SI vs MB). Las barras de error indican intervalos de confianza del 95%

Figura 10

Distribución de las diferencias individuales entre las condiciones SI y MB durante fase de prueba en el Experimento 4

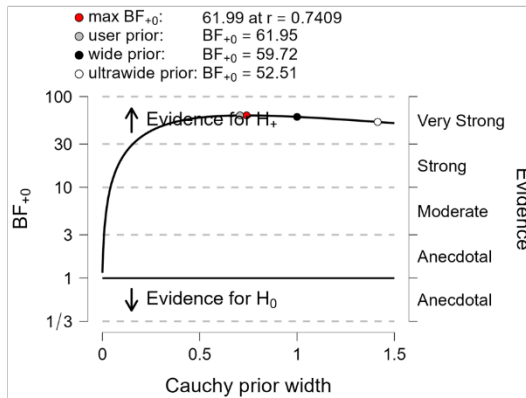


Nota: Distribución de las diferencias entre los participantes. Cada punto individual corresponde a un participante. Se obtuvieron restando las puntuaciones obtenidas en la condición MB a las obtenidas en la condición SI. Puesto que la mayoría de los puntos están situados por encima de 0, se observa mejor desempeño en la condición SI que en MB. La curva superior representa la distribución de las diferencias entre SI y MB, el diagrama de caja representa la mediana y los cuartiles (Q1 y Q3) y las líneas negras indican el rango intercuartílico de dichas puntuaciones.

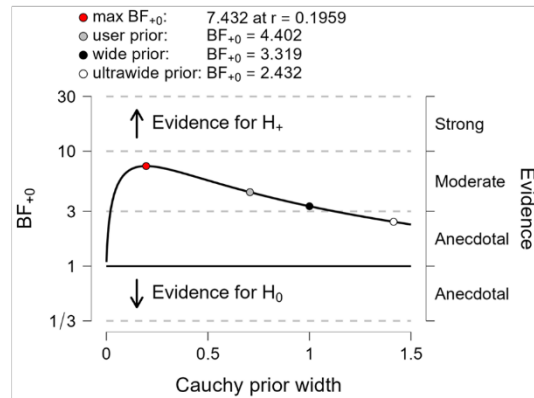
Figura 11

Pruebas de robustez del Factor de Bayes en pruebas t bayesianas confirmatorias en los Experimentos 1 (a), 2 (b), 3 (c) y 4 (d)

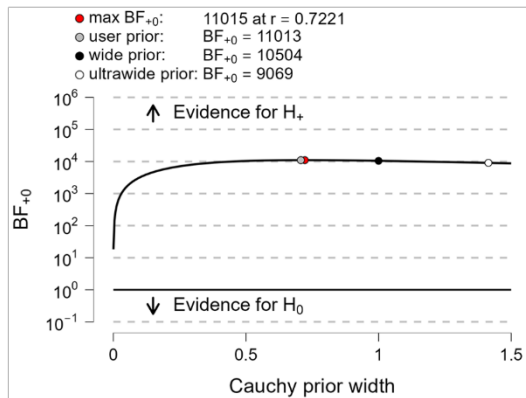
(a) Experimento 1



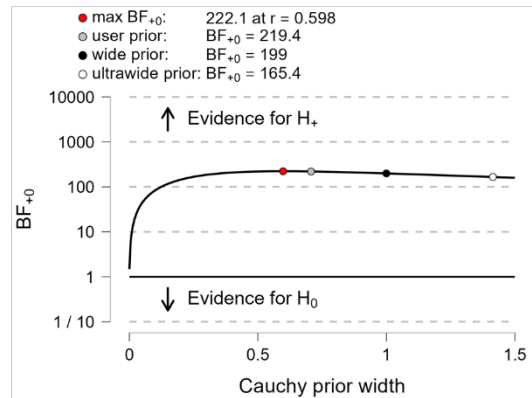
(b) Experimento 2



(c) Experimento 3



(d) Experimento 4



Nota: Análisis bayesianos de pruebas t para muestras relacionadas y sus pruebas de robustez para los Experimentos 1 a 4. Se presentan gráficamente los BF_{10} de cada experimento en el intervalo de valores de r entre 0 y 1.5. Se muestran numéricamente también los BF_{10} con el valor de $r = 1/\sqrt{2}$ (*user prior*). En el Experimento 1 (panel a) el Factor de Bayes fue de $BF_{10} = 61.65$ lo que sugiere evidencia muy fuerte a favor de la hipótesis alternativa. En el Experimento 2 (panel b) el Factor de Bayes resultante fue de $BF_{10} = 4.40$ lo que sugiere evidencia moderada a favor de la hipótesis alternativa. En el Experimento 3 (c) el Factor de Bayes resultante fue de $BF_{10} = 11015$, lo que sugiere evidencia extrema a favor de la hipótesis

alternativa. En el Experimento 4 (d) el Factor de Bayes resultante fue de $BF_{10} = 222.1$ lo que sugiere evidencia extrema a favor de la hipótesis alternativa. En todos los paneles se observa que el nivel de evidencia permanece estable en todo el rango a partir de $r = 0.2$, indicando que los resultados son robustos a la selección de distribuciones a priori. Sin embargo, en el caso del Experimento 2 se observa que hay disminución en los valores mayores de r , lo que sería consistente con un tamaño del efecto menor al del resto de experimentos, un dato congruente con los reportados por Kornell y Bjork (2008).

Discusión

Kornell y colaboradores (Birnbaum et al., 2013; Kornell et al., 2010; Kornell & Bjork, 2008) encontraron que el entremezclado de ejemplares ofrecía mayores beneficios en el aprendizaje inductivo de estilos artísticos que la presentación masiva de ejemplares. El presente trabajo replicó y extendió estos hallazgos, explorando cuestiones que habían recibido poca atención previamente.

El Experimento 1 consistió en una replicación directa del Experimento 1a de Kornell y Bjork (2008) en el que los participantes aprendieron el estilo artístico de diferentes pintores. Se obtuvieron resultados similares al observado en el trabajo original. Cuando se estudian cuadros de artistas de manera espaciada-entremezclada, se obtienen mayores tasas de acierto en tareas de categorización que cuando los cuadros aparecen de forma masiva-bloqueada. Sin embargo, se esperaba observar una mejora en el rendimiento a medida que se sucedían los bloques de la fase de prueba debido a la retroalimentación, tal y como ocurría en el trabajo original. Inesperadamente, en el presente trabajo no se observaron estos resultados pese a que se proporcionó retroalimentación. Una posible explicación es la eliminación la opción de “No lo sé” que aparecía junto a los nombres de los autores en la tarea de categorización original. Es posible que en el experimento original de Kornell y Bjork (2008) los participantes

eligiesen dicha opción cuando tenían poca confianza en su decisión y que ello llevase a un menor número de aciertos al principio de la tarea. En el presente experimento, los participantes no contaban con dicha opción, por lo que una interpretación plausible es que, pese a una falta de confianza en su respuesta, la representación generada de las categorías les hiciese elegir las respuestas correctas desde el principio. En definitiva, los resultados del Experimento 1, respaldaron parcialmente las hipótesis propuestas.

En el experimento 2, se realizó una replicación del Experimento 2 de Kornell y Bjork (2008) usando la Teoría de Detección de Señales para evaluar la discriminación de los participantes. Durante la fase de prueba, los participantes debían discernir si una serie de cuadros novedosos les resultaban familiares o no. Para ello, durante la tarea se presentaban cuadros nuevos de artistas previamente estudiados y cuadros de autores desconocidos que servían como distractores. Un problema encontrado con este experimento es que la fase de prueba podía actuar en sí misma como un evento de aprendizaje, por lo que tantos aciertos como errores podían incorporar información a las representaciones que los participantes tenían de las categorías, lo que contaminaba la prueba y reducía la precisión de la discriminación a lo largo de los bloques (Verkoeijen y Bouwmeester, 2014). Para paliar esta contaminación y siguiendo el protocolo de Kornell y Bjork (2008) solo se tuvieron en cuenta los resultados de los dos primeros bloques de la tarea. Además, siguiendo las recomendaciones de Brown y White (2005) se utilizó un $\log d$ como medida de la discriminación de los participantes, aportando más precisión y sensibilidad a la prueba. Los resultados obtenidos apoyaron la hipótesis inicial y replican con éxito el Experimento 2 de Kornell y Bjork (2008). Estos hallazgos son importantes para encontrar evidencia convergente, pues hasta la fecha solo el trabajo de Verkoeijen y Bouwmeester (2014) había replicado el experimento original.

En el Experimento 3 se siguió un procedimiento idéntico al reportado en el Experimento 1, pero realizando una tarea de categorización sin retroalimentación durante la fase de prueba. Este experimento se llevó a cabo para abordar la escasez de trabajos con este tipo de prueba (Brunmair y Richter, 2019). Si bien la retroalimentación durante el estudio puede ser deseable para aumentar el aprendizaje de los estudiantes (Schweppe et al., 2025), puede constituir un obstáculo para aislar el efecto del tipo de presentación. Tal y como se hipotetizó, la condición de espaciado-entremezclado obtuvo mejores proporciones de acierto que la condición de masivo-bloqueado pese a la eliminación de la retroalimentación.

Por último, el Experimento 4 consistió en una extensión del Experimento 3, añadiendo una semana de demora entre la fase de entrenamiento y la fase de prueba. El objetivo era probar los efectos de entremezclado en la retención de categorías a largo plazo. Existen varios trabajos que han realizado experimentos probando la inducción de categorías a largo plazo con demoras similares (Pan et al., 2018; Schweppe et al., 2025), pero hasta la fecha solo Zulkipli y Burt (2013a) habían usado cuadros de artistas en tareas de categorización a largo plazo, aunque ofrecieron retroalimentación durante la fase de prueba. Por lo tanto, el presente experimento podría ser el primero en probar con este paradigma el efecto a largo plazo y sin retroalimentación. Los resultados mostraron de nuevo ventaja a favor del espaciado-entremezclado, consistentemente con las hipótesis iniciales y con el resto de los experimentos.

En general, estos hallazgos van en la línea de trabajos que defienden la Hipótesis de Contraste Discriminativo. La idea central de esta hipótesis es que el beneficio del entremezclado en la inducción de categorías viene dado por la comparación entre elementos de diferentes categorías que se presentan secuencialmente (Birnbaum et al., 2013). Birnbaum et al. (2013) realizaron una serie de experimentos en los que se interrumpía el proceso de contraste entre categorías y encontraron que cuando se eliminaba la posibilidad de contrastar

ejemplares se perjudicaba notablemente el aprendizaje de categorías. Otros trabajos manipularon la posibilidad de realizar contrastes discriminativos en tareas de inducción de estilos artísticos y observaron que cuando los participantes podían hacer contrastes entre ejemplares obtenían mejores resultados que cuando se dificultaba el contraste (Ge et al., 2021). Además, múltiples trabajos han demostrado que el entremezclado se beneficia de situaciones en la que las categorías son muy similares y los participantes requieren extraer diferencias entre categorías (Carvalho y Goldstone, 2014a, b). Esta evidencia, sumada a la falta de apoyo que han recibido otras hipótesis ha hecho que la Hipótesis de Contraste Discriminativo se postule como explicación principal al efecto de entremezclado (Chen et al., 2021; Firth et al., 2021).

Esto va en la línea de los hallazgos del presente trabajo, ya que, dada la complejidad y variedad de dimensiones estímulares que ofrecen los cuadros, cabe esperar que la comparación entre ejemplares de diferentes categorías haya favorecido la discriminación de sus diferencias, lo que permitiría extraer las características definitorias para generar una representación del estilo artístico de cada autor. Pero las implicaciones prácticas de estos hallazgos van más allá del aprendizaje de estilos artísticos. Las pinturas, al igual que otros materiales naturales, tienen características sutiles que los hacen difíciles de discriminar, lo que las hace adecuadas para estudiar el efecto de entremezclado en tareas donde la discriminación entre categorías es más complicada (Zulkipli y Burt, 2013b).

Esto lleva a pensar que la generalización del entremezclado puede ir más allá del estudio de estilos artísticos y que puede ser una estrategia relevante en el mundo real. Por ejemplo, si extrapolamos su utilidad al mundo académico, durante el desarrollo de un libro de texto de biología se podrían introducir secciones donde se hace una presentación de distintos tipos de plantas aparentemente similares, favoreciendo extracción de diferencias entre ejemplares. Esto se ve apoyado por el conjunto de estudios mencionados en la introducción.

No puede descartarse que, al efecto de entremezclado en la inducción de categorías, se sume el efecto de espaciado en el aprendizaje de etiquetas asociadas a las categorías. Kornell y Bjork (2008) buscaron aislar el efecto de entremezclado del efecto de espaciado, eliminando las etiquetas verbales. Pero desde un punto de vista práctico, tendría interés maximizar el aprendizaje y servirse para ello de todas las estrategias y mecanismos disponibles. Por ejemplo, si bien en el presente trabajo se extendieron los resultados de Kornell y Bjork (2008) eliminando el uso de retroalimentación durante las fases de prueba, algunos estudios recientes afirman que el uso de retroalimentación ofrece resultados más claros a favor de las condiciones de entremezclado (Schweppe et al., 2025), por lo no hay razones para eliminar la retroalimentación en contextos donde se busque favorecer el aprendizaje.

Por otra parte, varios estudios han demostrado que la presentación masiva-bloqueada de ejemplares puede ser efectiva en algunos contextos, como durante estudios de pronunciación extranjera (Carpenter y Mueller, 2013) o cuando las categorías son muy fáciles de discriminar (Carvalho y Goldstone, 2017). Estos resultados abogan por tener en cuenta la especificidad de los materiales educativos y el contexto en el que se usan al aplicar estos fenómenos en la práctica pedagógica, por consiguiente, es necesario ser cauteloso y explorar detenidamente cada caso de estudio particular (Rowlandson y Simpson, 2023).

El presente trabajo no está exento de limitaciones que deben ser consideradas para interpretar sus aportaciones. La primera limitación es la caracterización de la muestra, pues fue recopilada a través de un muestreo por conveniencia reclutando a alumnado de la Facultad de Psicología y Logopedia de la Universidad de Málaga. Esto quiere decir que no se seleccionaron de forma aleatoria y que no conforman una muestra representativa de la población general, por lo que no se pueden generalizar los datos a la misma de manera directa.

El uso de cuadros de artistas ha sido considerado como un conjunto estimular con validez ecológica (Kang y Pashler, 2012). Pese a ello, se podría considerar que hay una falta de aplicabilidad a un contexto cotidiano, donde hay gran cantidad de estímulos que difieren en contenido y forma, que pueden verse beneficiados o perjudicados por el fenómeno de entremezclado. Es poco probable que los participantes vean estímulos similares de forma secuencial fuera de un contexto académico o experimental. Además, el presente trabajo se llevó a cabo en un entorno controlado de laboratorio, donde se reduce la aplicabilidad de los resultados en situaciones en las que los distractores, la ansiedad o las características idiosincráticas del contexto puede cambiar los efectos de estos fenómenos. Replicas conceptuales llevadas a cabo en aulas a veces no han podido replicar el efecto de entremezclado en el aprendizaje de categorías (Brunmair y Richter, 2019).

Estas limitaciones no solo marcan el alcance del presente trabajo, sino que ofrecen varias líneas prometedoras para futuros estudios. Se hace necesario extender el corpus de investigaciones de estos fenómenos a poblaciones diferentes y heterogéneas. Si bien algunos estudios han estudiado muestras de diferentes edades, muy pocos han realizado muestreos aleatorios y representativos (Brunmair y Richter, 2019).

Trabajos previos han intentado aislar los procesos cognitivos influyentes en este tipo de experimentos (Birnbbaum et al., 2013; Carvalho y Goldstone, 2015a, b; Kang y Pashler, 2012). Pese a que ofrecen resultados muy prometedores, la falta de consenso indica que la evidencia actual es insuficiente. Se hace evidente la necesidad de llevar a cabo más investigaciones que apunten a aislar el fenómeno de entremezclado para ofrecer una explicación más precisa y completa.

En esta línea, hasta la fecha pocos estudios han indagado en los correlatos fisiológicos implicados en el entremezclado y la inducción. Algunos estudios han utilizado *eye-tracking*

para investigar la atención de los participantes durante el aprendizaje (Carvalho y Goldstone, 2017; Ge et al., 2021). Un estudio de Lin et al. (2016) usó resonancia magnética funcional (fMRI) para investigar correlatos neurales asociados con el entremezclado durante tareas de aprendizaje motor. Pero no se conocen estudios de electroencefalografía (EEG) que aborden el correlato de estos fenómenos. Es evidente que la memoria es fundamental para el aprendizaje inductivo, poco sentido tendría inducir categorías si no somos capaces de recordarlas (Vlach et al., 2008). Puesto que entonces, el procesamiento de la información durante la inducción está estrechamente ligado a la memoria, una pregunta potencial es en que punto del procesamiento divergen estos procesos. Futuros trabajos podrían centrarse en utilizar el EEG para estudiar el momento en que estos dos procesos cognitivos se diferencian.

Conclusión

En síntesis, este trabajo ha demostrado que el entremezclado de ejemplares mejora el aprendizaje inductivo de categorías. Los resultados mostraron que se trata de un efecto robusto, que se replica en diferentes tareas, con beneficios tanto a corto como a largo plazo. Estos hallazgos subrayan el potencial de esta estrategia para mejorar los resultados en contextos académicos y abren la puerta hacia futuras investigaciones que exploren las complejidades de este fenómeno y nos ayuden a refinar su comprensión.

Referencias

- Birnbaum, M. S., Kornell, N., Bjork, E. L., y Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, 41(3), 392-402. <https://doi.org/10.3758/s13421-012-0272-7>
- Bjork, R. A., y Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, 6(2), 173-189. [https://doi.org/10.1016/0010-0285\(74\)90009-7](https://doi.org/10.1016/0010-0285(74)90009-7)
- Braithwaite, D. W., y Hall, G. J. (2024). Explaining procedures and interleaving practice in fraction arithmetic. *Learning and Instruction*, 90, 101854. <https://doi.org/10.1016/j.learninstruc.2023.101854>
- Brown, G. S., y White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods*, 37(3), 436-449. <https://doi.org/10.3758/bf03192712>
- Brunmair, M., y Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, 145(11), 1029-1052. <https://doi.org/10.1037/bul0000209>
- Carpenter, S. K., y Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, 41(5), 671-682. <https://doi.org/10.3758/s13421-012-0291-4>
- Carpenter, S. K., Pan, S. C., y Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, 1(9), 496-511. <https://doi.org/10.1038/s44159-022-00089-1>

- Carvalho, P. F., y Goldstone, R. L. (2014a). Effects of interleaved and blocked study on delayed test of category learning generalization. *Frontiers in Psychology*, 5, 936. <https://doi.org/10.3389/fpsyg.2014.00936>
- Carvalho, P. F., y Goldstone, R. L. (2014b). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, 42(3), 481-495. <https://doi.org/10.3758/s13421-013-0371-0>
- Carvalho, P. F., y Goldstone, R. L. (2015a). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, 22(1), 281-288. <https://doi.org/10.3758/s13423-014-0676-4>
- Carvalho, P. F., y Goldstone, R. L. (2015b). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, 6, 505. <https://doi.org/10.3389/fpsyg.2015.00505>
- Carvalho, P. F., y Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1699-1719. <https://doi.org/10.1037/xlm0000406>
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., y Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354-380. <https://doi.org/10.1037/0033-2909.132.3.354>
- Chen, O., Paas, F., y Sweller, J. (2021). Spacing and Interleaving Effects Require Distinct Theoretical Bases: A Systematic Review Testing the Cognitive Load and Discriminative-Contrast Hypotheses. *Educational Psychology Review*, 33(4), 1499-1522. <https://doi.org/10.1007/s10648-021-09613-w>

- Do, L. A., y Thomas, A. K. (2023). The Underappreciated Benefits of Interleaving for Category Learning. *Journal of Intelligence*, 11(8), 153.
<https://doi.org/10.3390/jintelligence11080153>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., y Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. <https://doi.org/10.1177/1529100612453266>
- Eglington, L. G., y Kang, S. H. K. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition*, 6(4), 475-485.
<https://doi.org/10.1016/j.jarmac.2017.07.005>
- Faul, F., Erdfelder, E., Buchner, A., y Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Firth, J., Rivers, I., y Boyle, J. (2021). A systematic review of interleaving as a concept learning strategy. *Review of Education*, 9(2), 642-684.
<https://doi.org/10.1002/rev3.3266>
- Ge, Y., Li, F., Li, X., y Li, W. (2021). What Is the Mechanism Underlying the Interleaving Effect in Category Induction: An Eye-Tracking and Behavioral Study. *Frontiers in Psychology*, 12, 770885. <https://doi.org/10.3389/fpsyg.2021.770885>
- Green, D. M., y Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.
- Guzman-Munoz, F. J. (2017). The advantage of mixing examples in inductive learning: A comparison of three hypotheses. *Educational Psychology*, 37(4), 421-437.
<https://doi.org/10.1080/01443410.2015.1127331>

- Hatala, R. M., Brooks, L. R., y Norman, G. R. (2003). Practice Makes Perfect: The Critical Role of Mixed Practice in the Acquisition of ECG Interpretation Skills. *Advances in Health Sciences Education*, 8(1), 17-26. <https://doi.org/10.1023/A:1022687404380>
- JASP Team (2025). JASP (Version 0.19.3) [Software].
- Kang, S. H. K., y Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. *Applied Cognitive Psychology*, 26(1), 97-103. <https://doi.org/10.1002/acp.1801>
- Kornell, N., y Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the “Enemy of Induction”? *Psychological Science*, 19(6), 585-592. <https://doi.org/10.1111/j.1467-9280.2008.02127.x>
- Kornell, N., Castel, A. D., Eich, T. S., y Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, 25(2), 498-503. <https://doi.org/10.1037/a0017807>
- Kurtz, K. H., y Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology*, 51(4), 239-243. <https://doi.org/10.1037/h0040295>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Li, A., Hu, M., Xu, A., Zhao, W., Hu, X., Shanks, D. R., Luo, L., y Yang, C. (2025). Instructional intervention effects on interleaving preference and distance during self-regulated inductive learning. *Journal of Educational Psychology*, 117(2), 206-227. <https://doi.org/10.1037/edu0000909>

- Lin, C.-H. (Janice), Knowlton, B. J., Wu, A. D., Iacoboni, M., Yang, H.-C., Ye, Y.-L., Liu, K.-H., y Chiang, M.-C. (2016). Benefit of interleaved practice of motor skills is associated with changes in functional brain network topology that differ between younger and older adults. *Neurobiology of Aging*, *42*, 189-198.
<https://doi.org/10.1016/j.neurobiolaging.2016.03.010>
- Nemeth, L., Werker, K., Arend, J., Vogel, S., y Lipowsky, F. (2019). Interleaved Learning in Elementary School Mathematics: Effects on the Flexible and Adaptive Use of Subtraction Strategies. *Frontiers in Psychology*, *10*, 86.
<https://doi.org/10.3389/fpsyg.2019.00086>
- Noh, S. M., Yan, V. X., Bjork, R. A., y Maddox, W. T. (2016). Optimal sequencing during category learning: Testing a dual-learning systems perspective. *Cognition*, *155*, 23-29.
<https://doi.org/10.1016/j.cognition.2016.06.007>
- Pan, S. C., Lovelett, J. T., Phun, V., y Rickard, T. C. (2019). The synergistic benefits of systematic and random interleaving for second language grammar learning. *Journal of Applied Research in Memory and Cognition*, *8*(4), 450-462.
<https://doi.org/10.1016/j.jarmac.2019.07.004>
- Pan, S. C., Tajran, J., Lovelett, J., Osuna, J., y Rickard, T. C. (2018). Does interleaved practice enhance foreign language learning? The effects of training schedule on Spanish verb conjugation skills. *Journal of Educational Psychology*, *111*(7), 1172-1188. <https://doi.org/10.1037/edu0000336>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., y Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203. <https://doi.org/10.3758/s13428-018-01193-y>

- Rohrer, D., Dedrick, R. F., Hartwig, M. K., y Cheung, C.-N. (2020). A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology*, 112(1), 40-52. <https://doi.org/10.1037/edu0000367>
- Rowlandson, P., y Simpson, A. (2023). *Interleaving in mathematical category learning*. PsyArXiv. <https://doi.org/10.31234/osf.io/gz5r7>
- Rozenstein, A., Pearson, G. D. N., Yan, S. X., Liu, A. Z., y Toy, D. (2016). Effect of Massed Versus Interleaved Teaching Method on Performance of Students in Radiology. *Journal of the American College of Radiology*, 13(8), 979-984. <https://doi.org/10.1016/j.jacr.2016.03.031>
- Sana, F., Yan, V. X., y Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology*, 109(1), 84-98. <https://doi.org/10.1037/edu0000119>
- Schweppe, J., Lenk-Blochowitz, A., Pucher, M., y Ketzer-Nöltge, A. (2025). Interleaved practice in foreign language grammar learning: A field study. *Journal of Educational Psychology*. Advance online publication. <https://doi.org/10.1037/edu0000917>
- Simmons, A. L. (2012). Distributed Practice and Procedural Memory Consolidation in Musicians' Skill Learning. *Journal of Research in Music Education*, 59(4), 357-368. <https://doi.org/10.1177/0022429411424798>
- Taylor, K., y Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, 24(6), 837-848. <https://doi.org/10.1002/acp.1598>
- Thompson, C. P., y Hughes, M. A. (2023). The Effectiveness of Spaced Learning, Interleaving, and Retrieval Practice in Radiology Education: A Systematic Review.

Journal of the American College of Radiology, 20(11), 1092-1101.

<https://doi.org/10.1016/j.jacr.2023.08.028>

Van Doorn, J., Van Den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A.,

Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman,

M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., y

Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a

Bayesian analysis. *Psychonomic Bulletin & Review*, 28(3), 813-826.

<https://doi.org/10.3758/s13423-020-01798-5>

Verkoeijen, P. P. J. L., y Bouwmeester, S. (2014). Is spacing really the «friend of induction»?

Frontiers in Psychology, 5, 259. <https://doi.org/10.3389/fpsyg.2014.00259>

Vlach, H. A., Sandhofer, C. M., y Kornell, N. (2008). The spacing effect in children's

memory and category induction. *Cognition*, 109(1), 163-167.

<https://doi.org/10.1016/j.cognition.2008.07.013>

Wahlheim, C. N., Dunlosky, J., y Jacoby, L. L. (2011). Spacing enhances the learning of

natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory*

& *Cognition*, 39(5), 750-763. <https://doi.org/10.3758/s13421-010-0063-y>

Whitehead, P. S., Zarny, A., y Marsh, E. J. (2022). Transfer of category learning to

impoverished contexts. *Psychonomic Bulletin & Review*, 29(3), 1035-1044.

<https://doi.org/10.3758/s13423-021-02031-7>

Wong, S. S. H., Chen, S., y Lim, S. W. H. (2021). Learning melodic musical intervals: To

block or to interleave? *Psychology of Music*, 49(4), 1027-1046.

<https://doi.org/10.1177/0305735620922595>

- Wong, S. S. H., Low, A. C. M., Kang, S. H. K., y Lim, S. W. H. (2020). Learning Music Composers' Styles: To Block or to Interleave? *Journal of Research in Music Education*, 68(2), 156-174. <https://doi.org/10.1177/0022429420908312>
- Yan, V. X., y Sana, F. (2021). The robustness of the interleaving benefit. *Journal of Applied Research in Memory and Cognition*, 10(4), 589-602. <https://doi.org/10.1037/h0101863>
- Zulkipli, N. (2013). Effect of Interleaving Exemplars Presented as Auditory Text on Long-term Retention in Inductive Learning. *Procedia - Social and Behavioral Sciences*, 97, 238-245. <https://doi.org/10.1016/j.sbspro.2013.10.228>
- Zulkipli, N., y Burt, J. S. (2013a). Inductive Learning: Does Interleaving Exemplars Affect Long-Term Retention? *Malaysian Journal of Learning and Instruction*, 10, 133-155. <https://doi.org/10.32890/mjli.10.2013.7655>
- Zulkipli, N., y Burt, J. S. (2013b). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, 41(1), 16-27. <https://doi.org/10.3758/s13421-012-0238-9>
- Zulkipli, N., McLean, J., Burt, J. S., y Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, 22(3), 215-221. <https://doi.org/10.1016/j.learninstruc.2011.11.002>