

Sesgos: ¿Vemos según estos son o son estos según vemos?

Libertad Campassi, Matías Grinberg y Romina D'Alessandro

Introducción a la neurociencia cognitiva y computacional

Diciembre 2018

Resumen

En esta investigación se estudia el sesgo representacional en humanos en función de la acumulación de pericia (o "expertise"). Particularmente, se evalúa la hipótesis de si quienes tuvieron formación considerable en Ciencias Exactas demuestran menores niveles de empatía y cognición social.

Se realiza una prueba que incluye la descripción libre de escenas naturales y el recuerdo de un relato, y mediante la aplicación de técnicas de NLP, específicamente vectores de palabra *GloVe*, se comparan tendencias atencionales sistemáticas relacionadas con los sesgos predichos por la hipótesis propuesta, mientras se contempla la validez externa mediante la prueba *Reading the Mind in the Eyes*.

1. Introducción

El aprendizaje, definido como la mejoría en alguna tarea y la acumulación de información cognitiva mediante plasticidad cerebral en algún tipo de memoria declarativa o no declarativa, conlleva un aumento en la eficiencia del procesamiento de la información [1]. La memoria procedural motora por ejemplo, permite que aprendamos a andar en bicicleta codificando una serie compleja de movimientos corporales independientes en una representación organizada. Este hecho se fundamenta en fenómenos conocidos a través de distintos dominios, como en los distintos tipos de apraxias, la afasia anómica específica a grupos léxicos, o la organización jerárquica de la corteza visual. Esta frontera entre lo subjetivo y lo objetivo remite a un debate ubicuo en la filosofía, incluyendo el empirismo y racionalismo, y la imposibilidad de demostrar el qualia [2], pero excede el alcance de este trabajo.

El mecanismo cognitivo de *chunking*, donde distintas estrategias median la retención en la memoria de trabajo, la codificación, el almacenamiento y la evocación [3], es otro ejemplo que exhibe estos mecanismos llamados "*top-down*" [4]. Este modo de procesamiento se puede observar en las ilusiones ópticas como *The Checker Board Shadow Illusion* de Edward H. Adelson [5], donde las expectativas dadas por los patrones visuales aprendidos por las capas tempranas de la corteza visual distorsionan una característica en la percepción. Desde un enfoque más moderno, podemos identificar la curiosidad, como exploración activa del mundo para aprender información útil y reducir la incertidumbre generando representaciones parsimoniosas. Estas devienen en expectativas, que nos facilitan una percepción unificada de la consciencia, y nos dan una ventaja adaptativa permitiéndonos adelantar posibles amenazas y economizar el sensorio [6]. Esta curiosidad activa puede ser modelada desde una perspectiva Bayesiana [1][7] [8], donde se busca reducir el error de predicción y el grado de incertidumbre, minimizando la entropía de los priors (variational free energy).

Para estudiar esta interacción entre sesgo y aprendizaje, el

lenguaje ofrece una buena ventana. La elección de palabras, y de todos los elementos dentro del continuo entre forma y contenido [9], transmite en sí misma una gran cantidad de información, sobre el contexto de la comunicación, la relación entre los hablantes, y otra meta-información. Mediante el lenguaje, se pueden contemplar tantas características del hablante como nivel de alerta, años de educación, profesión, o detectar trastornos de la salud como condiciones psiquiátricas [10]. Podemos notar si el hablante o escritor es emocionalmente cercano o distante, reflexivo o superficial, y posiblemente extrovertido o abierto a nuevas experiencias. El uso semántico de la palabra es un marcador significativo y un mediador ocasional de procesos sociales y de personalidad. [11]

La capacidad de leer las emociones de los demás está vinculada a la "inteligencia social". Dado que los ojos son una señal destacada en la comunicación social humana y categorización social [12], sirven como un estímulo adecuado para el presente estudio. Particularmente, la prueba *Reading the mind in the eyes* [13] toma una medida de la capacidad de leer emociones de los demás. Se desarrolló en Gran Bretaña y las imágenes fueron tomadas de revistas británicas en los años 90. Como era de esperar, la prueba no funciona perfectamente para personas que no son hablantes nativos de inglés o para personas que provienen de culturas que son muy diferentes a las de Gran Bretaña, sin embargo, muestra validez convergente con pruebas socio-cognitivas de la teoría de la mente. En el presente trabajo se usará como referencia para tomar una medida de la capacidad de leer emociones.

Estudios anteriores han demostrado que los ingenieros tienen bajos grados de competencia social, particularmente estudiantes de ingeniería mostraron menos empatía que los estudiantes de psicología y trabajadores sociales en las subescalas de fantasía y toma de perspectiva [14]. Esto motiva la propuesta de estudiar la incidencia del tópico de empatía en dos grupos, estudiantes o egresados de carreras exactas (se toman estudios en Ciencias de la computación, Matemáticas, Físicas e Ingeniería y serán definidos como: exactas) versus

estudiantes o egresados de otras carreras o sin formación específica (no exactas).

La empatía generalmente se define como un rasgo afectivo (por ejemplo, la capacidad de experimentar las emociones de otro [15]) y/o una capacidad cognitiva (por ejemplo, la capacidad de comprender las emociones de otro[16]).

2. Diseño experimental

La prueba se distribuyó mediante una interfaz web, hecha con HTML, y Javascript con la librería JsPsych para experimentos psicológicos. Para programas utilitarios, el servidor o "back-end" el análisis, se usó el lenguaje Python. Las librerías utilizadas fueron: Flask para crear el servidor, SpaCy para hacer el procesamiento de lenguaje natural, y varias otras como Pandas, NumPy, sci-kit learn y matplotlib para el análisis. El link del experimento se encuentra disponible aún.[24] y el análisis se realizó predominantemente en la plataforma abierta Google Colaboratory y también se encuentra disponible para revisión.[25]

El diseño procede de modo secuencial. Comienza con una historia para leer y recordar. Luego se proyectan dos videos de 20 segundos de duración por separado, donde el sujeto, al terminar de visualizar cada uno y debe describirlo con el mayor detalle posible por escrito con un mínimo de 120 palabras.

Luego se evalúa una versión reducida de la prueba *Reading the mind in the eyes* de 10 imágenes. Después se hace las preguntas de si estudia o estudió alguna carrera de ciencias exactas, su edad y genero. Para responder esto último se propuso una barra continua en la cual el sujeto podía poner una medida de su autopercepción.

También se pide resolver un ejercicio matemático [17] con respuestas en multiple choice.

Por ultimo se pide que el sujeto describa lo que recuerde de la historia que leyó al inicio, en este caso no se exige un mínimo de palabras. Finalmente se pide que distribuyan el link del experimento para así crear un efecto tipo "bola de nieve".

La prueba fue realizada por 86 sujetos, de los cuales 16 pertenecen al grupo de exactas y 70 al grupo de no exactas.

3. Análisis

Para medir la similitud semántica entre las descripciones, utilizamos *Word Embeddings* de tipo GloVe [18], pre-entrenados en la Universidad de Chile [19], para comparar la distancia con respecto a "empatía" entre los 2 grupos de sujetos.

Existen muchas maneras de generar embeddings, como análisis semántico latente (LSA), de uso extendido, o Word2vec. Es un método que permite representar lenguaje natural en forma de vectores, con el objetivo de que las palabras de significado semántico similar se encuentren cercanos en ese espacio vectorial (hipótesis distribucional) [20]. Para evaluar los embeddings, se aplicó la reducción de dimensionalidad UMAP [21] para generar un gráfico 3D interactivo [22] y observar su adecuación. El mismo fue exhibido al final del experimento como refuerzo lúdico, con las palabras más cercanas para las respuestas dadas por el sujeto, ver Figura (1). También, se exploraron las palabras más cercanas a "empatía",

que resultaron ser acordes a lo esperado (ej. "generosidad", "solidaridad", "sentimiento")

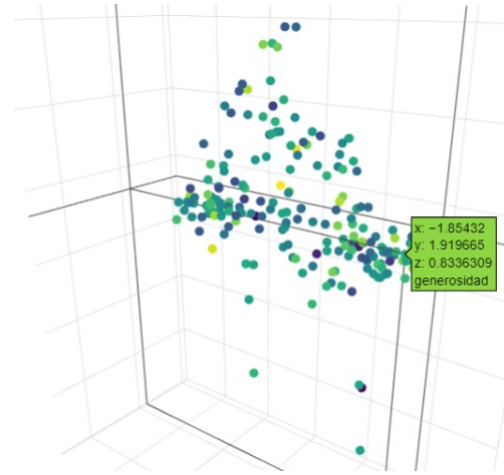


Figura 1: ScatterPlot 3D, cada punto representa una palabra en el espacio semántico siendo estos los vectores más cercanos a empatía.

Un sujeto fue asignado al grupo exactas luego de haber respondido de manera afirmativa a la pregunta "¿Estudias o estudiaste en exactas?", y haber resuelto el ejercicio matemático correctamente. Esto fue para asegurar aprendizaje procedural en esa tarea, asumiendo que seguir carreras con prácticas de matemáticas se asocia a mejor habilidades numéricas [23]. Para el testeo, se siguen los siguientes pasos:

- Concatenación de los textos de cada sujeto.
- Remoción de stop words y signos de puntuación.
- Cómputo de distancia coseno entre cada palabra de cada sujeto y empatía (con y sin tilde)
- Binarización de las distancias usando umbral definido a priori (0.15) [26]
- Recuento de palabras positivas para cada sujeto
- Prueba rank-sum o U de Mann-Whitney entre grupos

4. Resultados

Se separó en dos grupos, los que declararon estudiar exactas (A) y los que no (B), y se evaluó la respuesta para el ejercicio de matemática como se ve en la Figura (2) a continuación:

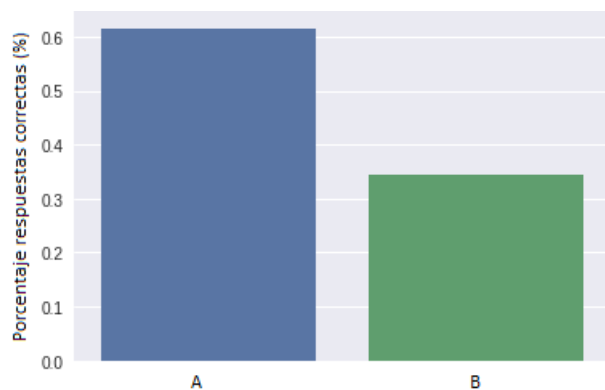


Figura 2: Gráfico en el cual se muestran los sujetos de ambos grupos que respondieron correctamente el problema de matemáticas

Donde el 61,5 % del grupo que afirmó que estudiaba ciencias exactas respondió correctamente. Mientras que en el grupo que afirmó no estudiar nada de esa índole, un 34,4 % respondió correctamente. Que el grupo de exactas haya resuelto el ejercicio matemático significativamente mejor ($r = 0.56$, $p < 0.001$), refuerza la hipótesis sobre que el aprendizaje procedural se asocia a mejores habilidades en pensamiento matemático.

Luego se separó en los grupos a estudiar (exactas versus no exactas) con la condición antes mencionada (haber respondido de manera afirmativa a la pregunta "¿Estudias o estudiaste en exactas?", y haber resuelto el ejercicio matemático correctamente).

Los resultados para la versión reducida de el test *Reading the mind in the eyes* se muestran a continuación, Figura (3):

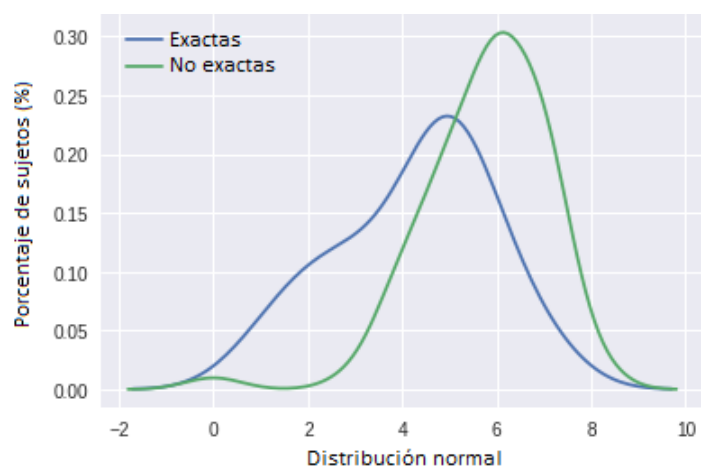


Figura 3: Resultados de versión reducida de *Reading the mind in the eyes*.

Este gráfico nos indica que el grupo de exactas muestra menos empatía que el grupo de no exactas con la versión reducida tal que $T = 3.69$ y $p < 0.001$

En cuanto a los resultados de la prueba U de Mann-Whitney entre los grupos, el valor del estadístico fue $U = 48.0$ con $p < 0.001$, por lo tanto se rechaza la hipótesis nula en favor de la alternativa. Las distribuciones obtenidas se muestran en la Figura (4):

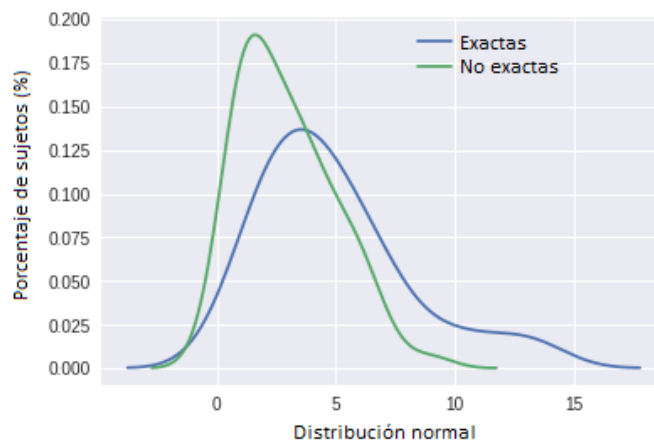


Figura 4: Resultados obtenidos de la prueba U Mann-Whitney.

Información adicional sobre los sujetos:

La autopercepción del género, se grafica en la Figura (5)

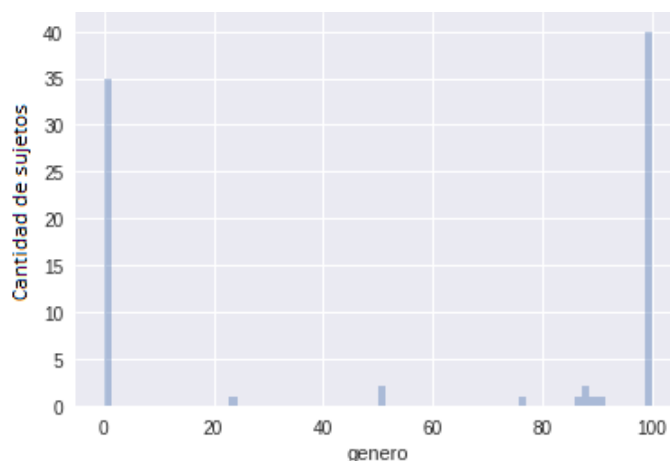


Figura 5: Gráfico obtenido para la barra de género, donde 0 indica mujer y 100 indica varón.

Del total, 35 sujetos se autopercebieron completamente mujeres y 40 completamente varones.

La edad se grafica en la Figura (6)

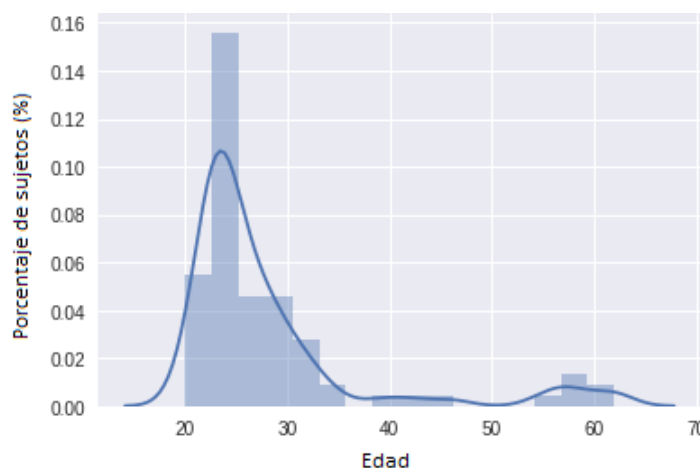


Figura 6: Gráfico obtenido para la edad.

Se puede observar que la gran mayoría de los sujetos que hicieron el experimento tienen una edad mayor a los 20 años y menor que 35 años. Entonces se podría tomar para analizar ese rango de edad, se cree que eso daría una mejor respuesta en la prueba de embeddings.

5. Conclusiones

- La versión reducida del test *Reading the mind in the eyes* arrojó resultados satisfactorios que apoyan la hipótesis de que el grupo exactas muestra menos empatía que el resto.
- La metodología online de recolección es una gran ventaja para el muestreo de datos, a pesar de relegar parcialmente el control del entorno durante su ejecución.
- Si bien los resultados fueron acordes a lo esperado, es probable que los estadísticos calculados por ejemplo para la prueba U de Mann-Whitney otorguen mayor nivel de significación y tamaño de efecto si se toman más su-

jetos, específicamente del grupo de exactas.

- El grupo de exactas resolvió el ejercicio matemático significativamente mejor que el otro, lo que refuerza la hipótesis sobre que el aprendizaje procedural se asocia a mejores habilidades en pensamiento matemático.
- Encontramos una diferencia en las distribuciones, pero no una tendencia significativa hacia la empatía.
- Los embeddings resultaron ser una herramienta robusta para medir la similitud semántica y comparar la distancia con respecto a "empatía" entre los dos grupos.
- Sería preferible la recolección de más datos, para poder controlar posibles fuentes espurias de variabilidad como horas de sueño, distracciones, o factores latentes ruidosos.
- Dado que los embeddings resultantes dependen del corpus de texto con el que fueron entrenados, y que el "transfer learning" aparenta ser esencial para el buen rendimiento, sería preferible computar embeddings *ex profeso* para la tarea, utilizando textos regionales.

Referencias

- [1] Hesselmann, G., Sadaghiani, S., Friston, K. J., Kleinschmidt, A. (2010). Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PloS One*, 5(3), e9926.
- [2] Dennett, D. C. (2014). Why and how does consciousness seem the way it seems?. Open MIND. Frankfurt am Main: MIND Group.
- [3] Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology*, 19(2), 183-204.
- [4] Kveraga, K., Ghuman, A. S., Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and cognition*, 65(2), 145-168.
- [5] <http://www.cogsci.rpi.edu/~heuveb/teaching/CriticalThinking/Web/Presentations/PerceptionMemory.pdf>
- [6] Hesselmann, G., Sadaghiani, S., Friston, K. J., Kleinschmidt, A. (2010). Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PloS One*, 5(3), e9926.
- [7] Mandel, D. R. (2014). The psychology of Bayesian reasoning. *Frontiers in psychology*, 5, 1144.
- [8] https://scholar.harvard.edu/files/xiaoshengmu/files/paying_attention.161129.pdf
- [9] Ashual, O., Jurgenson, T., Grinberg, D. (2017). Deep Text Style Transfer.
- [10] Altszyler, E., Berenstein, A. J., Milne, D., Calvo, R., Fernandez-Slezak, D. (2018, November). Detección automática de casos urgentes en foro de salud mental. In XIX Simposio Argentino de Inteligencia Artificial (ASAI)-JAIIO 47 (CABA, 2018).
- [11] Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1), 547-577.
- [12] Zebrowitz, L. A. (2006). Finally, faces find favor. *Social Cognition*, 24(5), 657-701.
- [13] Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of child psychology and psychiatry*, 42(2), 241-251.
- [14] Chato Rasool, Henrik Danielsson and Tomas Jungert (2012) Empathy among students in engineering programmes, *European Journal of Engineering Education*, 37:5, 427-435, DOI: 10.1080/03043797.2012.708720
- [15] Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child development*, 413-425.)
- [16] Hogan, R. (1969). Development of an empathy scale. *Journal of consulting and clinical psychology*, 33(3), 307.)

- [17] Kahneman, D., Egan, P. (2011). Thinking, fast and slow (Vol. 1). New York: Farrar, Straus and Giroux.
- [18] <https://nlp.stanford.edu/pubs/glove.pdf>
- [19] <https://github.com/uchile-nlp/spanish-word-embeddings>
- [20] Naili, M., Chaibi, A. H., Ghezala, H. H. B. (2017). Comparative study of word embedding methods in topic segmentation. *Procedia Computer Science*, 112, 340-349.
- [21] UMAP <https://arxiv.org/abs/1802.03426>
- [22] <https://plot.ly/>.
- [23] Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition*, 93(2), 99-125.
- [24] <https://cerebrock.github.io/mindspace/>
- [25] <https://colab.research.google.com/drive/1AsVS4pFXIba38-XxQ1IIS6Eq-h5Tf3Ap>
- [26] Diuk, C., Fernandez Slezak, D., Raskovsky, I., Sigman, M., and Cecchi, G. A. (2012). A quantitative philology of introspection. *Frontiers in integrative neuroscience*, 6, 80.

Algorithm 1 Data Generation Pipeline

```

Initialize blank image
for random layers do
    Sample  $p$  polygons
    Add augmentations(polygons) to image

```
