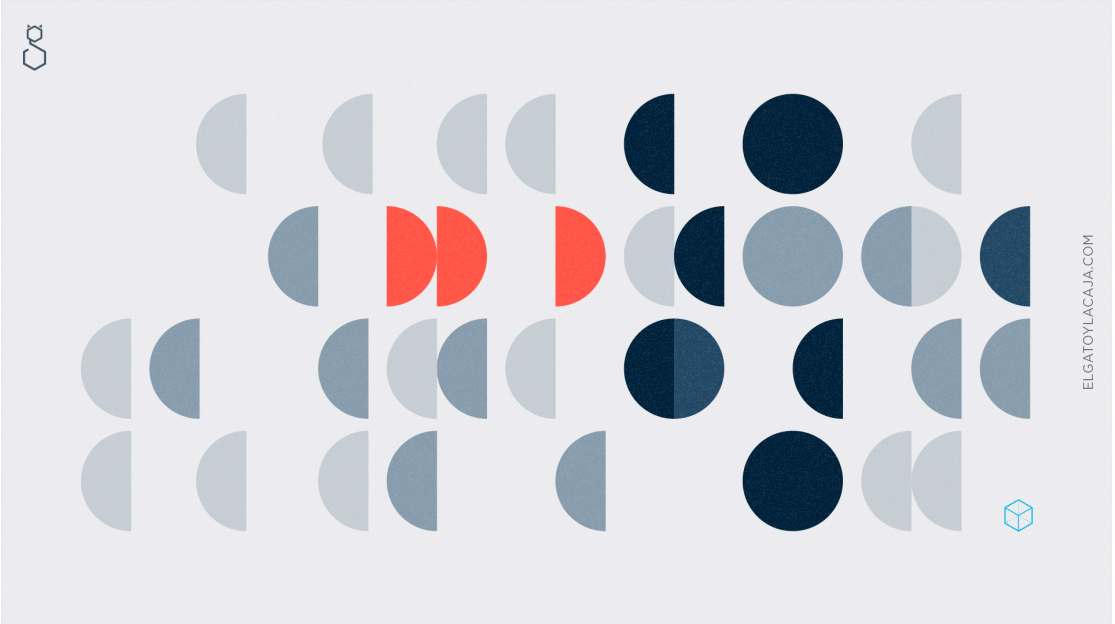


02/10/2017

El Revisor

TXT PABLO A. GONZÁLEZ, ANDRÉS RIEZNIK IMG JUAMA

¿Existen mecanismos que revisan que no justifiquemos como propia una opinión que deviene de un engaño? ¿Podemos mapear los mecanismos mentales detrás de la sensación de 'acabo de decir una estupidez'?

"A nuestra comprensión de por qué el cerebro parece inmutable ante desconexiones no sólo se le agregó la noción de que, en un sentido, envía la mitad de sus decisiones al reino del inconsciente; sino también el descubrimiento de 'El Intérprete'. Este sistema especial del hemisferio cerebral izquierdo tomaba nota de los comportamientos que resultaban de muchos sistemas mentales. Parecía ser la cámara de vigilancia de nuestro comportamiento que, por supuesto, era la evidencia de que había ocurrido un acto mental o cognitivo. El intérprete no sólo tomaba nota; trataba de darle 'sentido' a esa conducta, manteniendo una narrativa sobre por qué una cadena de actos

ocurría. Es un dispositivo precioso y, probablemente, únicamente humano. Está trabajando en nosotros constantemente mientras tratamos de explicar por qué nos gusta algo o por qué tenemos una opinión en particular, o cuando intentamos racionalizar algo que hemos hecho. Es éste dispositivo de interpretación el que toma las entradas del cerebro masivamente modularizado y automático, y crea orden a partir de caos; inventa la explicación que 'tiene sentido', que nos lleva a creer en una cierta forma de esencialismo, es decir, que somos un agente consciente unificado. ¡Buen intento, Intérprete!"

– *Michael S. Gazzaniga, Relatos desde los dos lados del cerebro: una vida dedicada a la neurociencia.*

Referirnos a nosotros mismos como 'uno' esconde ahí, de forma evidente y gritona, no sólo lo reflexivo (ese '*self*' del '*one self*', que no contiene la misma cualidad reflexiva que el 'uno mismo' del castellano), sino también **lo numeroso**. Uno es —o dice que es, o entiende que es, o capaz espera ser— uno (1). Pero ¿qué pasa si es menos sencillo? ¿Qué pasa si Walt Whitman tenía razón, si no pasa nada si se contradice porque sí, porque *es* inmenso, porque sí contiene multitudes?

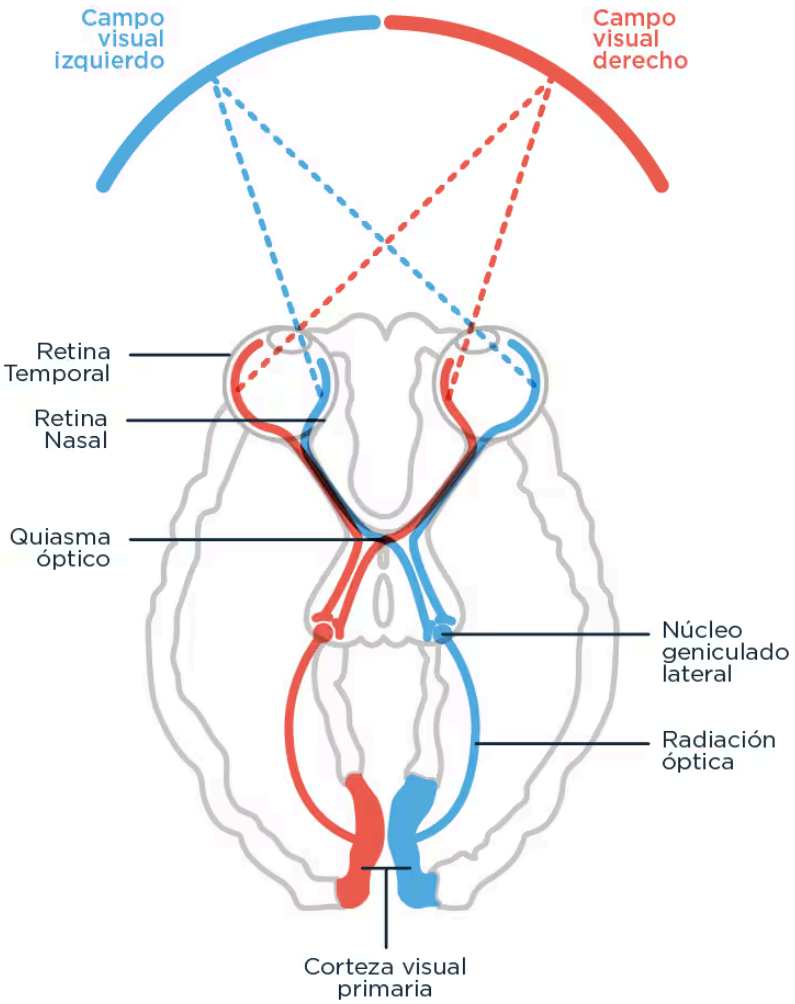
Durante la primera mitad del siglo XX, un grupo de neurocirujanos decidió que una forma de abordar la epilepsia severa podía incluir detener la tormenta de actividad eléctrica a medida que se expandía por el cerebro en un lugar clave, una estructura que conecta ambos hemisferios cerebrales y que permite que enormes cantidades de información fluyan entre ambos: **el cuerpo calloso**. Romper este puente permitiría contener la actividad en un solo hemisferio y brindarle al paciente la posibilidad de no perder el control de todo su cuerpo, conteniendo el ataque en una mitad y reduciendo significativamente uno de los daños más frecuentes en un ataque de epilepsia (cerca del 37%): las heridas a la cabeza como consecuencias de esta pérdida del control corporal. Si sólo la mitad del cuerpo está afectado, el paciente retiene un grado de control en el episodio, minimizando golpes y caídas. La técnica resultó enormemente efectiva para contener los ataques, y las personas que habían sido sometidas a este procedimiento fueron conocidas

por la literatura clínica con el nombre de **pacientes *split-brain*** (de cerebro dividido).

Siguiendo una larga tradición de las neurociencias, ahí donde hay una anormalidad, los científicos también vemos una oportunidad, y así fue que **estos individuos con cerebros divididos se convirtieron en una oportunidad de observar la mente humana en una condición extremadamente atípica**. Si bien es cierto que estos pacientes eventualmente desarrollan formas de compensar la pérdida de comunicación interhemisférica (a veces involucrando, por ejemplo, estructuras subcorticales que no son cortadas en la callostomía), el cerebro nunca vuelve a ser como era, y eso nos pone de frente a la pregunta de **qué pasa con 'uno' cuando el soporte que lo sostiene ya no es igual**.

Roger Sperry fue uno de los primeros investigadores en abordar la idea (absolutamente radical) de que, en estas situaciones, **ya no estábamos frente a una conciencia, sino dos**. En sus propias palabras *"Cada hemisferio es un sistema consciente en sí mismo. Percibiendo, pensando, recordando, razonando, decidiendo y emocionándose (...) tanto el hemisferio derecho como el izquierdo pueden ser conscientes en simultáneo de experiencias mentales diferentes, a veces mutuamente en conflicto, que se desarrollan en paralelo".*

Para llegar a esta idea, Sperry no había trabajado originalmente con estos pacientes, sino que su trabajo en comunicación interhemisférica venía de trabajar en gatos a los que les había cortado nervios del sistema visual de manera que el ojo izquierdo sólo mandara información al hemisferio derecho, y el derecho solamente al izquierdo (recordemos que los hemisferios cerebrales reciben y envían información predominantemente al lado opuesto del cuerpo). Además, había cortado el cuerpo calloso, cosa que la información que entraba por un ojo quedaba contenida en un solo hemisferio. Una vez hecho esto, los entrenaba para diferenciar figuras geométricas y, cuando ensayaba esa capacidad de aprender, ambos hemisferios recordaban y se comportaban de maneras completamente diferentes.



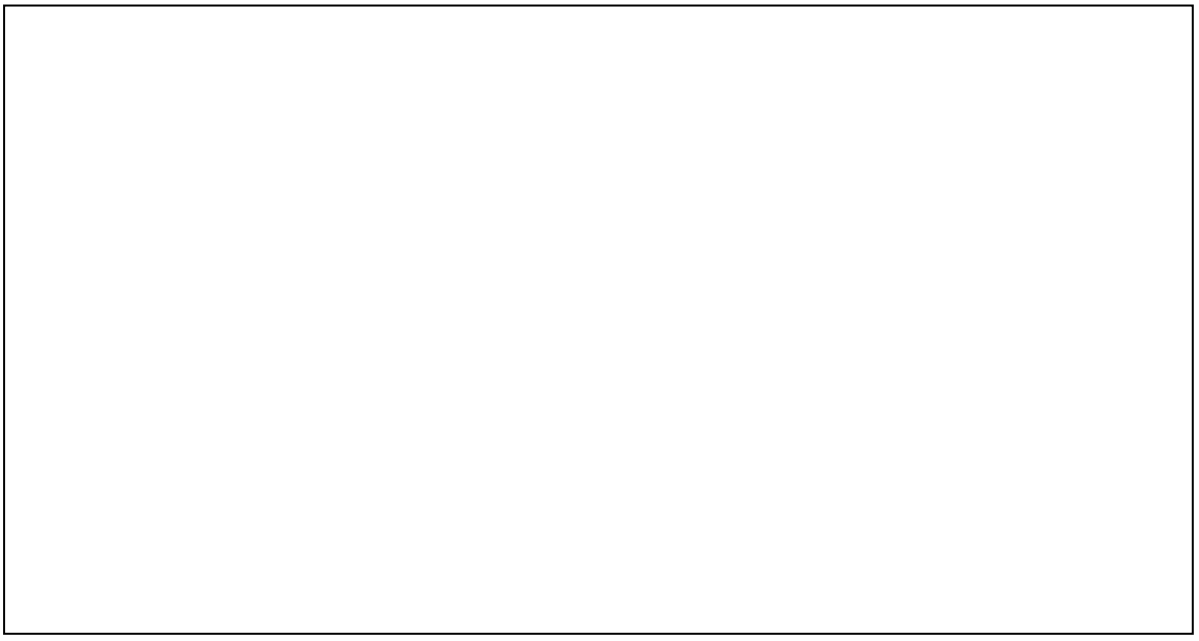
Así funcionan los ojos. Por un lado, este esquema aclara pero complejiza la idea de que 'cada ojo manda información al hemisferio opuesto'. Esto no es exactamente así, sino que cada ojo tiene dos mitades que mandan información por dos vías. El ojo derecho recoge información que viene desde SU derecha y SU izquierda, y lo mismo el izquierdo. Esa información va a recorrer un camino que puede ir directo al mismo hemisferio (ipsilateral) o atravesar el quiasma óptico y pasar al otro hemisferio (contralateral). La clave está en que toda la información visual sobre lo que pasa en nuestro 'lado derecho', llega al hemisferio izquierdo, independientemente de por qué ojo entre, y viceversa.

El esquema también sirve para entender qué se corta y qué no en un paciente split brain. Mientras que el cuerpo calloso es completamente seccionado, no así el quiasma óptico. Esto permite que la información que viene de los ojos sí llegue al hemisferio opuesto, no así la información que ya está procesándose en cada hemisferio.

Con esta información, la idea y la oportunidad de trabajar con humanos, uno de sus estudiantes (Michael Gazzaniga) comenzó a hacer experimentos que cambiarían para siempre la forma en la que entendemos la mente. En 1978

Gazzaniga, junto a **Joseph DeLoux**, empezó a hacer experimentos de percepción sobre pacientes de cerebros divididos a los que les pedían ejecutar una tarea perceptual. Exponían a los pacientes a una pantalla con 2 imágenes: una sólo podía ser vista por su lado derecho y la otra por el izquierdo (sí, a veces espiaban y, sí, con el tiempo y el *eye tracker* pudieron repetirse y mejorarse estos experimentos, sacando la información de pantalla si el paciente usaba el ojo indebido para observar la imagen). Cuando presentaban información al lado derecho (una palabra escrita o una imagen) y les preguntaban qué habían visto, podían reportarlo sin problemas. Esto adquiere mucho sentido cuando se toma en cuenta que los dos hemisferios no hacen exactamente lo mismo, sino que contienen funciones diferentes (o sea, **existe *lateralización* en el cerebro**), y justamente **en el hemisferio izquierdo** (a donde llegaba la información que entraba por el lado derecho de los pacientes *split brain*) **es donde encontramos dominancia para el lenguaje** (que, si quisiéramos localizar todavía más, podríamos encontrar en las áreas de Wernicke y Broca).

Por otro lado, **cuando exponían el lado izquierdo a palabras o imágenes y le preguntaban al paciente qué había visto, éste no lograba articular una respuesta. Lo sorprendente era que, al pedirle que dibujara lo que sintiese utilizando la mano derecha, lo hacía correctamente.**



Así como el lado izquierdo contenía una dominancia sobre el lenguaje, hoy sabemos que **el derecho tiene una dominancia fuerte en tareas que incluyen**

la integración entre un todo y sus partes y en muchas tareas visuoespaciales, como relaciones espaciales, movimiento aparente, rotación mental y relaciones especulares (si algo es el reflejo de otro algo), por nombrar algunas. **Que los pacientes no pudieran reportar oralmente lo que veían, pero sí dibujarlo, en sí mismo ya es impresionante**. Gazzaniga publicó decenas de investigaciones sobre estos fenómenos, pero cuenta (en su biografía) que tardó 15 años desde que empezó a investigar a estos pacientes hasta que se dio cuenta de que debía hacer una pregunta clave.

En un experimento, se exponía al paciente a dos imágenes (cada una expuesta de manera de fluir hacia hemisferios diferentes) y se le pedía que reportara lo que veía eligiendo una opción de entre una serie de dibujos que le presentaban. En una oportunidad, a uno **le presentaron una casa nevada (al lado izquierdo) y una pata de gallina (al lado derecho)**. De entre las imágenes, **el paciente rápidamente relacionó la pata de gallina con una imagen de una gallina, y para la otra serie de imágenes (en la que le habían mostrado la casa nevada), eligió una pala para nieve**. Hasta acá no había sorpresa, momento en el que apareció esa pregunta clave: **Gazzaniga decidió consultarle al paciente por qué había elegido la imagen de la pala**, a lo cual el paciente respondió *"Es simple. La pata de gallina va con la gallina, y la pala es para limpiar el gallinero"*. **El hemisferio izquierdo, viendo (literalmente) la respuesta del lado derecho, había contextualizado esa respuesta y le había dado una narrativa.** En un escenario donde no podía más que observar una acción sobre el pasado, y sin tener forma de acceder a la información de por qué esa decisión había sido tomada, algo en su hemisferio izquierdo contaba una historia que le daba al presente cohesión con el pasado. **En algún lugar de ese lado izquierdo del cerebro, algún tipo de mecanismo, programa, agente o como sea que queramos denominarlo minimizando los peligros de crear un Teatro Cartesiano, contaba una historia con la información que podía y le daba sentido**. Gazzaniga llamó a este mecanismo ***El Intérprete*** y, desafiante, dijo que *"Al estudiar cómo el cerebro permite una sensación de totalidad, también podemos obtener pistas sobre cómo nuestro sentido de la identidad personal es provocado por el cerebro, una cuestión que*

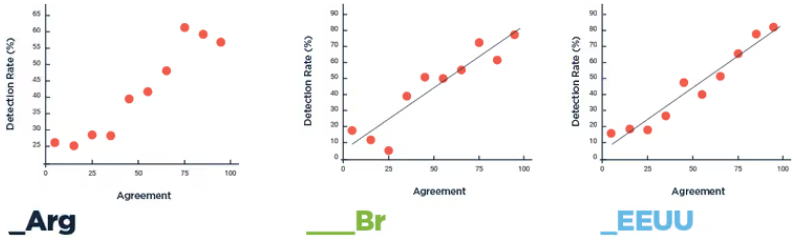
se ha discutido durante mucho tiempo en la filosofía y ahora también en el campo de la neurociencia". Es entonces esta sensación subjetiva de unicidad, esta completitud que experimentamos, la que transito ahora, escribiendo, y que reconoce en sí mismo el que lee, esa de ser un solo agente unificado, consciente y en control, con un solo propósito unificado, no más que eso: **una sensación. Una narrativa. La última estación de una integración de procesos perceptivos e introspectivos que van sumándose y generando emergentes hasta coalescer en un único espacio atencional unificado, sobre el que después contamos una historia con un solo protagonista indivisible y monolítico**.

Pero estos son pacientes a los que les partieron el cerebro en dos (casi casi literalmente, aunque sigue unido por otros puentes). Esto no es generalizable a todas las personas, por lo menos hasta que un experimento bien planteado lo demuestre. Pasaron años hasta que a alguien se le ocurriera la forma de fotografiar a este Intérprete en acción en un cerebro normal. **En 2005, un grupo sueco decidió usar magia para medir si las personas eran o no capaces de identificar diferencias entre decisiones que habían tomado e información que se les presentaba** ('magia' entendida no como expresión de algo genial, aunque también lo fuera, sino tomando la prestidigitación como técnica indispensable del procedimiento experimental). La pregunta de fondo era **¿podemos engañar a alguien para que justifique como propia una opinión que deviene de un engaño?**, ¿podemos ver la narrativa cubriendo la distancia entre el mundo como *debería ser* dada mi participación y como me dicen que *es*? El experimento era tan hermoso como simple: le mostraban a cada sujeto dos fotos de personas y le hacían elegir cuál de las dos le resultaba más atractiva. Inmediatamente después, truco mediante, les entregaban la foto *equivocada* (la que no habían elegido) y les pedían que justificaran su respuesta. El resultado: **más de la mitad de las veces los voluntarios, en lugar de identificar la discrepancia, generaban una narrativa para argumentar una decisión que no era suya.** Contaban cómo habían elegido la foto (la manipulada, la falsa, el gato por liebre) por su sonrisa, por sus ojos o por lo que fuera, pero siempre con la idea de fondo de que esa era la foto que habían elegido. **Los sujetos habían**

tomado la información con la que contaban y generado una narrativa retrospectiva que justificaba esa idea. Ahí, en un cerebro normal, también estaba el Intérprete.

Algo que hoy (retrospectivamente, y con varios experimentos de Ceguera a la elección encima) nos resultó particularmente extraño en el trabajo de los suecos es que ellos informan que *'la tasa de detección no fue influenciada por la similaridad de las caras'*. Es decir, **para ellos no importaba qué tan distante estuviera la realidad de la expectativa, el Intérprete funcionaba con igual firmeza.** Nosotros decidimos ver si esto era así o si, tal vez, ya fuera la naturaleza del experimento, el tamaño de la muestra o algún otro factor, podíamos encontrar en nuestro intento un resultado diferente. Intuitivamente, recorríamos la idea de que no debería poder engañarse a una persona de la misma manera con dos cosas parecidas que con dos muy diferentes, y menos con algo en lo que siente gran confianza: uno no esperaría que lo engañen con una foto de su propia pareja, sobre la que tiene altísima confianza de reconocer. La mejor manera de ensayar esa intuición era, por supuesto, haciendo un experimento.

En 2015 aprovechamos las elecciones presidenciales (el *ballotage*, estrictamente) **para 'engañar' a personas, o más bien para comprender el proceso de autoengaño con la ayuda de muchísimos voluntarios.** Nuestra ventaja fue que, además de preguntarles *'qué foto te gusta más'* (lo que en nuestro caso era *'estás a favor o en contra de esta propuesta'*), les podíamos preguntar **cuánto**. Los resultados nos permitieron no sólo observar y confirmar que el fenómeno de ceguera de la elección (la justificación retrospectiva vía *Intérprete* de decisiones que no tomamos) estaba fuertemente presente, sino que además pudimos ver que **la detección dependía directa y linealmente del nivel de acuerdo.** Qué tanto te gusta o disgusta la foto (o qué tan de acuerdo o en desacuerdo estás respecto de una idea) hacía que fuese muy distinta la tasa de engaño.



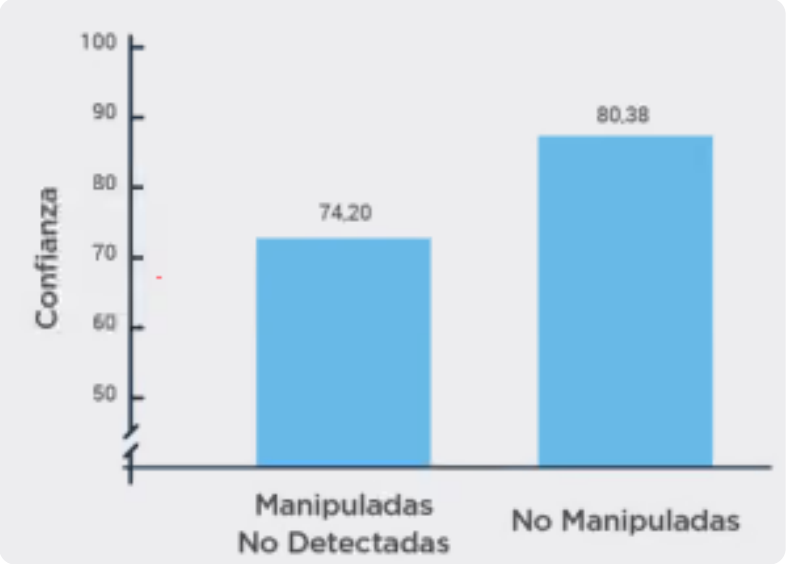
Podemos ver cómo la tasa de detección (que el sujeto elija la opción de 'quiero cambiar mi respuesta' por sobre 'mantengo mi respuesta', es para los 3 experimentos independientes positivamente correlacionado con el grado de acuerdo (la distancia entre 5 y los extremos, ya sea 0 o 10). ¿Barras de error? Están, pero no son visibles (son tan chicas que el mismo punto es visualmente más grande que las barras de error). Aguanten los datasets grandes.

Habíamos encontrado diferencias con respecto a los resultado de los suecos, y ese intérprete absoluto e independiente ahora estaba atado por la narrativa previa: **mientras más fuerte era la postura original frente a una decisión, más difícil era que ese mecanismo generase una nueva narrativa y que ésta fuera aceptada como propia por la persona.**

Sabiendo ahora que había mucha consistencia en esta relación entre cuán extrema era la postura ante una idea y la capacidad de detectar engaños, nos preguntamos cuál era el lugar que ocupaba la confianza en todo esto. **¿Cómo afecta la confianza que tenemos sobre nuestra postura a la hora de ser engañados? ¿Podemos encontrar una forma de medir si podemos evaluar nuestra propia capacidad de ser engañados?** El desafío era tratar de ensayar la precisión del sistema de atribución de confianza (en otras palabras, nuestro reporte introspectivo sobre la seguridad de esa postura). En particular, tratar de entender si los participantes que no habían expresado voluntad de cambiar su respuesta (o sea, que no habían detectado la manipulación) de alguna manera habían 'notado' y hasta reportado diferente sobre esas preguntas. Digamos, **si en algún lugar dentro de ellos habían *sentido* el engaño**.

Decidimos entonces comparar las confianzas en las respuestas manipuladas no detectadas (M) con las no manipuladas (NM). Para eso nos enfocamos en los 816 participantes que no habían detectado ninguno de los dos engaños (ni la primera o la tercera pregunta, que eran las manipuladas y que representaban el 35% de la muestra) y comparamos sus confianzas con las de las respuestas no manipuladas. Al hacerlo, encontramos que **el 60% de ellos (492) expresaban menor confianza en las manipuladas no detectadas respecto de las no manipuladas** ($p < 1e\text{-}171$; sí, dice e-171, en esta casa no se p hackean los resultados, Lisa). Como queríamos estar seguros, para descartar que estas diferencias tuvieran que ver con el

grado de acuerdo (ya que habíamos visto que acuerdo y confianza sí estaban relacionados el uno con el otro), hicimos un segundo análisis. Esta vez **comparamos las confianzas para respuestas manipuladas no detectadas contra las no manipuladas, pero normalizando por el grado de acuerdo.** Como en las respuestas manipuladas el acuerdo era siempre 40 o 60 (elegimos que el sistema, al engañar, les presentara a las personas una respuesta que siempre fuese *contraria a su postura original pero no extremadamente opuesta*), las comparamos con las respuestas no manipuladas que tuvieran acuerdo entre 35 y 45 o entre 55 y 65, respectivamente. Así vimos que mientras que en las respuestas no manipuladas la confianza promedio era de 80 +/- 0.03, en las manipuladas no detectadas (o deberíamos decir no reportadas), la confianza era de 74 +/- 0.01.



Acá podemos ver cómo les atribuímos confianzas distintas a las preguntas manipuladas (M) no detectadas respecto de las preguntas no manipuladas (NM). Sí, no hay barras de error, pero no porque no las hayamos calculado sino porque, una vez más, son tan chicas que no entran en la representación visual. Sí podemos decir que los errores estándar son de 0.01 para M y 0.03 para NM.

En esos 6 puntos de diferencia habíamos encontrado algo extremadamente interesante: **una capacidad de detectar engaño de manera no reportable,** o más bien no reportable a menos que le pidiésemos expresamente al sujeto que indagase en su confianza. Estaba presente algún tipo de **detección inconsciente del auto engaño**, y decimos ***inconsciente*** en términos de que **esa sensación no era reportada directamente, pero sí era capaz de ser reflejada cuando el**

sujeto declaraba una confianza más baja de la esperable: un fenómeno nunca antes observado en la literatura. Había algún tipo de mecanismo de revisión sobre la capacidad de autoengañarnos que estaba operando por debajo de las profundidades de lo reportable (lo *consciente*, si se quiere), y habíamos encontrado una forma para medirlo.

Timothy Ferris describe la conciencia como una monarquía parlamentaria; una donde múltiples agentes pujan por los recursos hasta que emerge una decisión/acción, y es ahí donde un Rey la explica como si fuese una voluntad única. Como si todo cerrara perfectamente. Como si cada pieza del rompecabezas calzara justo, y como si tuviese acceso a todas esas piezas.

Bueno, parecería que el ejemplo es bueno. Parecería que muchas, muchas veces, **llenamos vacíos con narrativas que armamos a medida y que mantienen andando la ilusión de unicidad, volición y orden.**

Gracias a lo que aprendimos en estos experimentos empezamos a desarrollar una idea que, si bien todavía es joven, podría ser una forma interesante de pensar el proceso de generación de relatos sobre nuestras decisiones. Lo novedoso de este enfoque es que enriquece la idea de *El Intérprete* al agregarle a la generación de narrativas un segundo paso, que tiene que ver con el proceso de *evaluación* de esas narrativas. **No sólo nos contamos una historia sino que, de alguna manera, evaluamos qué tan *bueno* es ese relato**. Qué tan despegado está nuestro modelo del mundo del modelo que el mundo nos devuelve al observarlo, qué tan distante, qué tan inconsistente. Así, **a mayor distancia entre lo que declaramos (la postura original) y lo que el sistema nos devuelve (y declara como información que emanó de nosotros mismos), mayor *tensión***. Para las respuestas sobre las que tenemos una postura más extrema, esa puja entre lo que pensamos y lo que nos dicen que pensamos se hace tan grande que nuestro Rey Intérprete ya no puede sostener con narrativas. Ahí, algo se rompe. Así como la historia que nos contamos tiene su origen en un módulo narrativo, ***creemos* que ese Intérprete está sujeto a algún tipo de mecanismo de control**. Una forma de revisar la narrativa y de levantar el banderín cuando tratamos de justificar lo injustificable.

En honor a Gazzaniga, **elegimos referirnos a este mecanismo como *'El Revisor'*:** nuestro guardián ante el autoengaño. Un agente que vigila nuestra propia narrativa. Un pedacito de nosotros que revisa la tensión entre el mundo como creemos que es y el mundo como se nos presenta. Uno que nos cuida de nosotros mismos para que no nos coma el relato.

Los que contribuímos de una u otra manera para hacer este trabajo somos: Alan Frieiro, Andrés Rieznik, Dardo Ferreiro, Facu Alvarez Heduan, Juam a Garrido, Julieta Figini, Lorena Moscovich, Mariano Sigman, Pablo González, Rocco Di Tella, Rodrigo Catalano.
Para leer de Gazzaniga por él mismo, acá.

elgatoylacaja.com/investigacion/el-revisor

Sumate en
eglc.ar/bancar