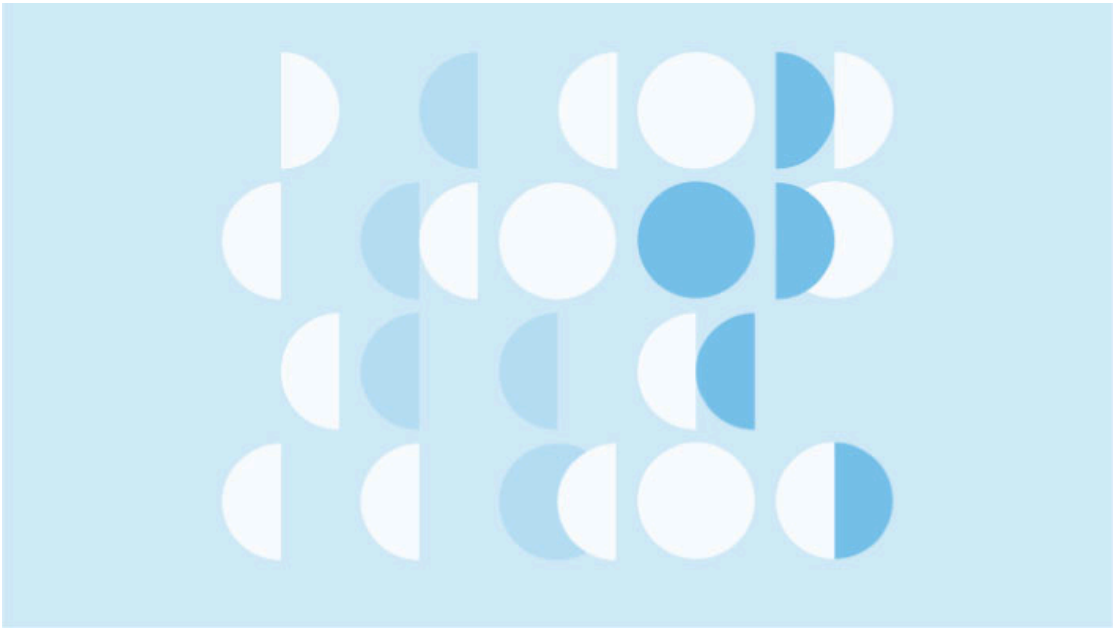


04/01/2018

Ceguera Chile

TXT LABS, ROCCO DI TELLA

¿Cómo hacemos un laboratorio 100% transparente? No sabemos, pero lo estamos intentando. Esto es Diarios de Investigación, capítulo 0: Ceguera Chile.

Diarios de Investigación: Episodio 0

Empezamos Labs con la idea de construir un laboratorio de puertas permanentemente abiertas. Eso empezó diseñando experimentos que pudieran llegar a todos lados, compartiendo los resultados en formatos accesibles, conversando los detalles de diseño y análisis, buscando publicar en revistas de acceso abierto y juntándonos de vez en cuando en algún lugar a charlar los resultados.

Hoy damos otro paso en ese camino de pensar una ciencia más abierta, de imaginar que no puede haber soporte narrativo más efectivo de comunicación pública de ciencia que *hacerla* juntos.

Queremos que tengan acceso completo a la cocina. Al diseño experimental antes de que el experimento pase, a nuestras discusiones incompletas, a los datos crudos, los *scripts* de análisis, a las cosas que dan bien, mal y, especialmente, a las que dan raro.

Hoy empieza Diarios de Investigación, una sección que va a ir creciendo de a poco y en la que pretendemos ir publicando los progresos de cada una de nuestras líneas de investigación a medida que van sucediendo.

Si quieren la posta, realmente no es por acá. Acá van a encontrar cocina, mugre, cosas que dan más o menos y, por sobre todo, conversaciones sobre cómo diseñamos experimentos, interpretamos datos o conectamos nuestros resultados con el conocimiento existente.

Bienvenidos: así hacemos ciencia. No es limpia, eterna ni perfecta, pero es de verdad.

□

Ceguera CL

La semana pasada tuvimos la oportunidad de lanzar el cuarto experimento para nuestra línea de Ceguera a la Elección, donde tratamos de entender sobre procesos mentales haciendo experimentos que tienen que ver con la atención, la introspección y la generación de narrativas. Esta vez le tocó a Chile ser anfitrión del experimento que ya pudimos hacer en Argentina, Brasil y USA aprovechando los contextos políticos polarizados para entender cómo esa polarización afecta la forma en la que elegimos y justificamos nuestras decisiones.

Para contextualizar a los descontextualizados sin repetir y sin soplar, invitamos con una pila de énfasis a los que no sepan de qué hablamos cuando hablamos de 'Ceguera' a empezar por este posteo que sacamos apenas 48 horas después del experimento en Argentina: (y que después se convirtió en este *paper*) y contado

acá por una de las protagonistas en La Nación. ***Esto es clave, porque sin pasar por acá todo lo que sigue adelante va a tener poco sentido.***

TL;DR? Las personas rápidamente desconectan sus decisiones de sus intenciones originales; al punto que pueden hacer propias respuestas manipuladas y usarlas como anclas para construir nuevos argumentos que las justifican, aun cuando no se correspondan con sus primeros deseos. Queríamos saber si reconocemos nuestras decisiones, si podemos ser engañados, si somos capaces de dar razones por decisiones que no tomamos y si todo este procedimiento nos puede llevar a cambiar nuestra decisión original.

Masticar esos primeros resultados nos llevó a desarrollar algunas ideas sobre cómo reaccionamos ante las *narrativas*, tanto internas como externas, y aparecieron una serie de mecanismos que fuimos organizando de a poco en ideas como El Revisor. Ideas que todavía estamos poniendo a prueba, precisamente en experimentos como este.

Recién ahora, con esto masticado, podemos pasar a Chile. Les recordamos en este punto que todos los resultados que van a ver son preliminares, que pueden cambiar a medida que vayamos analizándolos de maneras distintas y que todavía no están publicados, pero en estos años descubrimos la potencia de conversar preliminares. Abrir estas ideas y charlas, tanto entre nosotros como con ustedes, entre todos, fue una experiencia absolutamente positiva que vamos a profundizar. En este caso particular, charlar de Ceguera Argentina nos fue orientando sobre cómo podríamos modificar ligeramente los experimentos para hacer los controles cada vez más sólidos y hasta para hacernos preguntas nuevas.

Cambios metodológicos

Hicimos algunos cambios metodológicos importantes en relación a lo hecho en la Argentina.

Los dos cambios principales fueron implementados con el objetivo de falsear dos explicaciones alternativas a nuestra hipótesis de la existencia de un Revisor, un

módulo mental encargado de monitorear inconscientemente las explicaciones que nos damos a nosotros mismo (explicaciones generadas por otro módulo, el archifamoso, por lo menos entre nosotros, Interprete, propuesto por Michael Gazzaniga).

Cuando en el experimento argentino notamos que la confianza reportada en las respuestas realmente dadas (respuestas NM, no manipuladas) era mayor, en promedio, que aquella confianza reportada en respuetas que el participante no había dado, pero creía haber dado (respuestas MND, manipuladas no detectadas), pensamos que ello se debía a una detección inconsciente del autoengaño. El Revisor *nota* que algo raro está pasando, y aunque no llega a detectar la manipulación (el participante no declara que quiere cambiar su respuesta), ese procesamiento inconsciente emerge a la conciencia como una baja en la confianza reportada.

Sin embargo, una crítica plausible a nuestro argumento es que quienes no detectaron la manipulación en realidad son personas que siempre ponen baja confianza a todo lo que hacen y dicen, mientras que entre las respuestas NM había tanto participantes de alta como de baja confianza. Nuestro propio experimento había mostrado que las personas que van por la vida confiando poco en lo que hacen y dicen son más fácilmente engañables, por lo menos en relación a esos hechos y dichos. Es decir, capaz el ser humano se comporta como un autómata que pone altas o bajas confianzas en general y que la no detección del engaño ocurre más comúnmente en personas que suelen reportar baja confianza. Las bajas confianzas en MND serían independientes de la existencia de un Revisor.

Una forma de saber si esa teoría alternativa puede explicar la diferencia observada en confianzas dadas en respuestas NM y MND, es preguntar la confianza antes y después de la manipulación. En Argentina solo preguntamos la confianza después de la manipulación. En Chile, pudiendo comparar la confianza dada inicialmente con aquella dada al final, podemos ver si, para un mismo nivel de confianza inicial, en las respuestas NM se reportan al final confianzas mayores que en las respuestas MND. Si al principio todos reportan igual confianza, las diferencias al final no se

pueden deber a un cierto sesgo hacia reportar confianzas mayores o menores, sino a los diferentes tratamientos.

Esto requirió un cambio en el relato a los participantes. En Argentina, como no preguntamos la confianza al inicio, la excusa que poníamos para volver a mostrarle sus opiniones a los participantes (y ver si notaban que la habíamos cambiado) era la de que ahora te queríamos preguntar cuánto confiabas en eso que habías -o no- dicho. ¿Ahora qué excusa ibamos a meter? Pusimos una coqueta explicación que decía que queríamos estudiar cómo el cambio de contexto modifica tu acuerdo y tu confianza.

Lo divertido es que eso, que comenzó como un engaño, terminó siendo de alguna manera cierto porque en el grupo de control (en el que no manipulamos nada) tenemos resultados recontra interesantes que nos hablan de... ¡cómo el acuerdo y la confianza se modifican con el contexto (o con las veces que haya pensado en el tema)!

La segunda modificación tuvo el objetivo de falsear una segunda plausible crítica al Revisor. Ahora el ser humano ya no es un autómata que pone confianzas hacia un lado o el otro, sino uno que, dado un nivel de acuerdo, establece una confianza. En Argentina, la manipulación fue así: si, en una escala de 0 a 100 (0 siendo totalmente en desacuerdo y 100 totalmente de acuerdo), habías puesto un valor entre 0 y 50, te decíamos que habías dicho 60; si decías algo entre 50 y 100, te decíamos que habías dicho 40. Te cambio de lado, pero no tanto. El objetivo era que la manipulación fuese sutil, más difícil de detectar.

Ya observamos que el nivel de acuerdo está relacionado linealmente con la confianza (nivel de acuerdo lo definimos como cuán de acuerdo o desacuerdo estás, o sea, qué tan lejos de la posición neutral estás, y lo cuantificamos como |*acuerdo*-50|*2). Entonces, si manipulás de 90 a 40, o sea un nivel de acuerdo mayor a uno menor, claro que habrá menos confianza. En Argentina, desechamos esta crítica controlando por el nivel de acuerdo, es decir, comparamos las confianzas dadas en respuestas MND con aquellas dadas en NM con niveles de acuerdo de 40 o 60. Eso nos quitó mucho poder estadístico, porque muchos sujetos del grupo NM ponen otros acuerdos, no 40 ni 60. Entonces, en Chile, para

ganar poder estadístico, aun a riesgo de ser más obvios en la manipulación, simplemente espejamos el acuerdo reportado, diciendo que la personas había dicho algo del otro lado y con el mismo nivel de acuerdo o desacuerdo, es decir, la misma distancia al centro: si respondias 100 te decíamos que habías dicho 0; si decías 90, 10; 80, 20 y así por delante y viceversa. De yapa, no solo ganamos poder estadístico, si no que ahora la crítica anterior se invalida: estamos manipulando la respuesta, pero manteniendo el nivel de acuerdo.

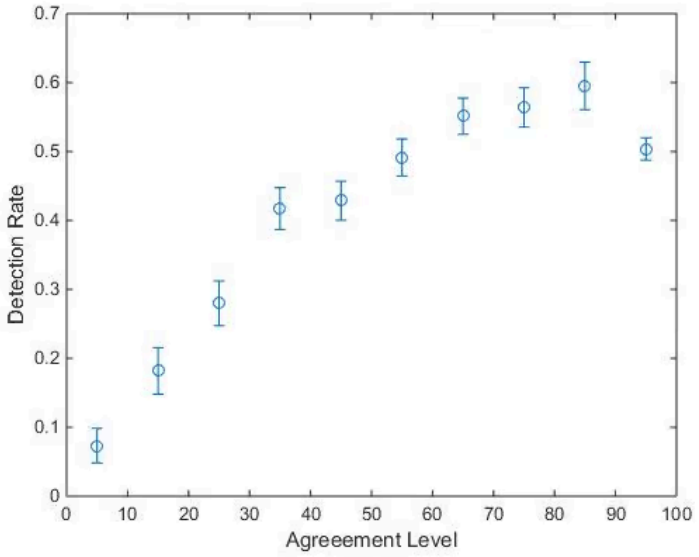
En resumen, hicimos **dos modificaciones importantes: preguntar la confianza antes y después de la manipulación y espejar el acuerdo en las manipuladas en lugar de llevarlo siempre a 40 o 60.**

Resultados (parciales y con pinzas, pero resultados al fin)

1) Tasas de detección

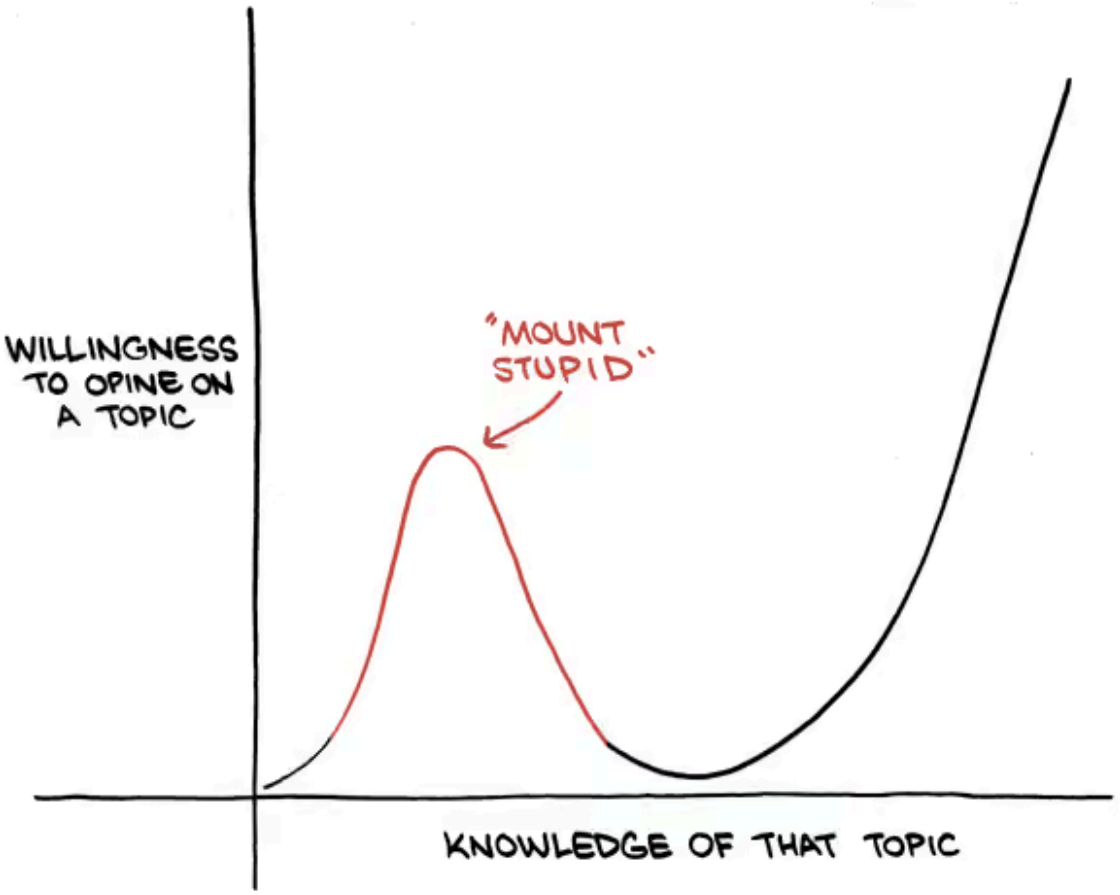
El primer resultado interesante que pudimos ver es la tasa de detección total sobre respuestas manipuladas: 46.4%. Es bajísimo teniendo en cuenta que espejamos la respuesta dada. Esperábamos incluso que, dado que la distancia entre el acuerdo reportado y el mostrado son mayores, el cambio fuese más obvio y en consecuencia las tasas de detección subieran. Sorpresa.

El segundo resultado interesante es que al graficar la tasa de detección de la manipulación como función del nivel de acuerdo, se volvió a repetir **el mismo patrón de siempre.** Otra vez (y van 4: Argentina, Brasil, EEUU y ahora Chile), la capacidad de detección de respuestas manipuladas correlaciona positivamente con qué tan polarizada es la postura de la persona sobre ese tema.



Así salen las figuras de MatLab. Como ya dijimos, estamos experimentando un cachito, pero tenemos muchas ganas de mostrar no solamente las figuras finales lindas sino las que nos escupe directamente el software, así los traemos más cerca de la cocina del análisis.

Algo que apareció en este experimento (y que veníamos viendo en los experimentos anteriores en forma más ruidosa o menos evidente) fue un claro descenso en la tasa de detección para la posición más extrema (la tasa de detección es mayor para respuestas con nivel de acuerdo entre 80 y 90 que entre 90 y 100). Es decir, **la detección de engaño aumenta con el grado de polarización hasta cierto punto, y luego deja de crecer e incluso disminuye en un extremo**. Este comportamiento, quizás algo antiintuitivo, sugiere que un sujeto completamente polarizado es incluso más susceptible a ser engañado que un sujeto un poco menos polarizado, o polarizado pero no taaaanto. ¿Puede eso estar diciendo que los sujetos que reportan confianzas extremadamente altas no son los más introspectivos respecto de ese tema? ¿Dunning-Kruger, sos vos?



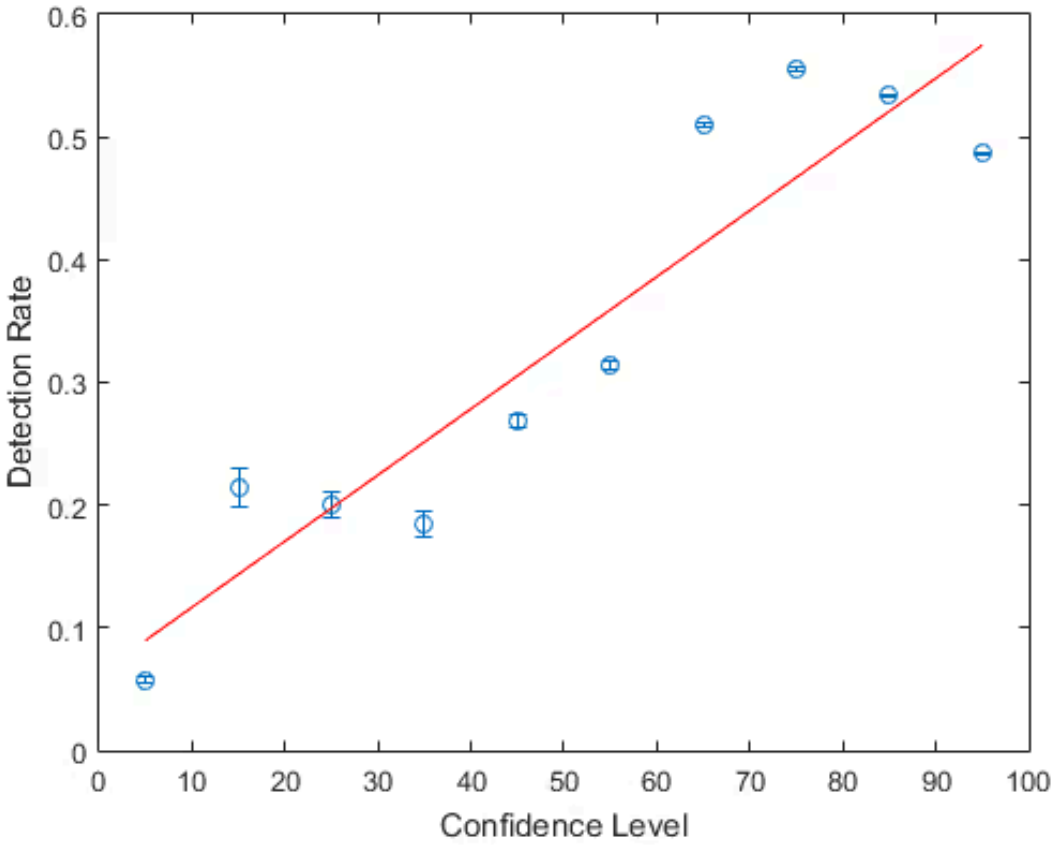
Además del nivel de acuerdo, nos preguntamos qué otras variables influyen sobre la detección del engaño. Analizamos aquellas que en experimentos anteriores se mostraron significativas: la intención de voto (a qué candidato votan), el género y la confianza en la respuesta.

Respecto de la intención de voto, esta vez no vimos una diferencia significativa. Los votantes de Guiller y de Piñera eran igualmente manipulables. Es probable que esto se deba al bajo poder estadístico de nuestra muestra, dado que 1153 manipulados votarían a Guiller y sólo 196 a Piñera. Siempre queremos muestras más balanceadas (en este y en todo aspecto que podamos, sea edad, género, educación), pero esto solamente lo pudimos obtener en Argentina. Tanto en Brasil, USA y Chile nos costó tener acceso a sujetos con intención de voto de *centro derecha* (muchas comillas en esa palabra).

Respecto de la detección por género, Chile volvió a mostrar un patrón que ya vimos en las otras 3 poblaciones. Otra vez, las mujeres fueron menos engañables, con tasas de detección de 43,95 % para hombres y 48,75 % para mujeres (diferencia que es significativa, considerable en tamaño y consistente a lo que vimos en las

otras 3 poblaciones). No tenemos claro qué pensamos que eso significa, pero escuchamos ideas.

Por último, vimos que la confianza reportada también, como en los experimentos anteriores, correlaciona positivamente con la detección del engaño. Aquí, por ejemplo, la confianza después de la manipulación como función de la detección del engaño:



2) El relato se vuelve realidad y nos enseña sobre la Teoría Argumentativa del Razonamiento

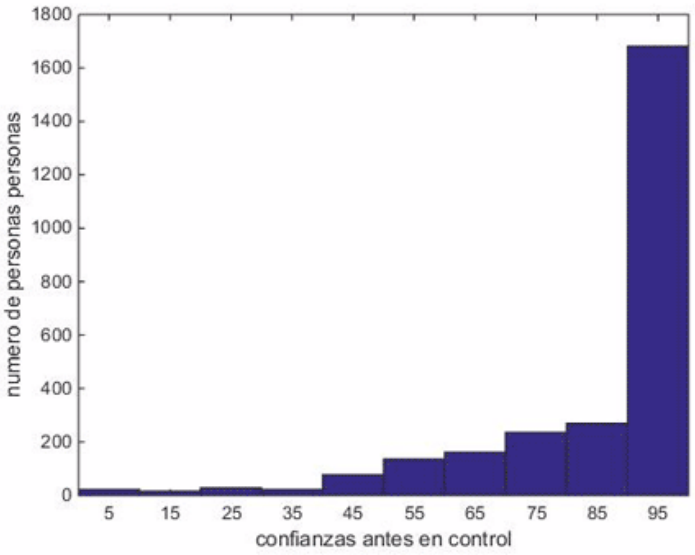
Pensar en el Revisor y en el fenómeno de *Detección inconsciente del autoengaño* (acá el link de nuevo, por las dudas) nos hizo pensar en la idea de cómo tomar una decisión: enunciarla y estar expuestos a esa enunciación (ya sea por decirla en voz alta o por escribirla, o elegir un punto en un *slider*), lograba que esa idea estuviese accesible a módulos mentales que de otra manera no accederían a ella. La pregunta que nos hicimos es si esto pasa todo el tiempo, de manera que cada decisión afecta el estado basal del sistema (algo así como imaginar las posturas no como un punto fijo sino como una distribución de probabilidad, y asumir que el sampleo modifica

esa distribución). Si esto fuese así, pedirle a alguien que declare su confianza debería afectar su próxima declaración de confianza. Y es más. Viendo que en experimentos (ajenos y propiofs), lo grupos de personas en los que varias personas que piensan lo mismo tienden a declarar confianzas mayores en sus respuestas luego de conversar entre ellas, decidimos que no era descabellado pensar a una persona como un grupo.

Pero lo bueno es que no es solamente discursivo, sino que podemos poner esto de manera falsable. Y lo podemos hacer gracias a que tenemos al grupo de control en que no manipulamos ninguna respuesta. Como les adelantamos, el relato de que te preguntábamos de nuevo sobre tu acuerdo y tu confianza para estudiar cómo el contexto las modifica se hizo verdad y tenemos resultados muy interesantes en el grupo control.

De ser esto como lo imaginamos, **esperaríamos que las confianzas posteriores a un primer pedido de declaración de confianza fuesen, en general, mayores que las originales** (estamos hablando del grupo de control en que no se manipulaba nada)**.** Algo así como *'ahora que tuve que expedirme respecto de cuánta confianza tengo sobre este tema y lo veo, más partes de mí que las que son accesibles conscientemente tienen acceso a esta información, y la próxima vez que me pregunten voy a estar un poquito más convencido que esta'*.

Y no podrán creer lo que pasó a continuación:



Noten que las dos columnas de la derecha suben cuando pasamos de azul a verde y todas las demás bajan. O sea, en el

grupo de control, en que no hay manipulaciones, las confianzas después de la repregunta aumentan, en promedio, en relación a la confianza inicialmente dada en la pregunta, hay más personas que reportan confianzas altas.

Las confianzas viran hacia posiciones más convencidas (un aumento de 1,5 puntos de confianza, que es chico, pero al mismo tiempo significativo, con p=0,0249 usando *Wilcoxon ranksum test*).

¿Es este resultado preliminar? Sí. ¿Estamos entusiasmados? Muy.

Esta idea es completamente consistente con la Teoría Argumentativa del Razonamiento, que pone en el centro de la discusión la idea de que el razonamiento aparece en la evolución a partir de la interacción social. La idea fundamental de esta teoría es que razonamos no tanto para llegar a la verdad sino más bien para argumentar y convencer, consciente o inconscientemente, *a priori*, a otros, pero creemos que también a nosotros mismos. Con esa idea en mente, tratamos de ver si un fenómeno grupal puede darnos una pista sobre lo que pasa en un individuo. Esta idea está todavía un poco cruda, pero nos entusiasma mucho conversarla, profundizarla y pulirla, así que hagámolos en los comentarios. Que la revisión por pares venga temprano y esta idea tenga que sostenerse por sus propios medios.

3) El Revisor sale victorioso (por ahora, como siempre en la ciencia)

Cuando comparamos cuánto cambia la confianza (la confianza al final menos la confianza al principio) en respuestas MND y NM, vemos que, como predicho por nuestra teoría del Revisor, la confianza en NM en promedio aumenta y en las MND disminuye, incluso controlando por la confianza inicial. No solo aumenta la proporción de personas que baja su confianza en relación al grupo de control (30% versus 22%, p = 1e-8), sino que la confianza baja mucho más: baja un promedio de 30 puntos sobre 100 en las respuestas MND y de 14 puntos en las NM, p = 1.2e-12. Así, los cambios metodológicos nos permitieron verificar las predicciones de nuestra teoría que propone la existencia de un Revisor y falsear dos explicaciones alternativas.

Final de algo que recién empieza

Como vimos en Resultados 2, la modificación experimental en la que empezamos a pedir confianza antes y después nos permitió estudiar si el sólo hecho de pedirle a alguien que declare su confianza es en sí mismo un tratamiento. Pero ¿cómo puede ser esto? ¿qué haría que este control en realidad fuese una especie de experimento cero?

Hasta ahora nos había interesado el fenómeno que describimos como *El Revisor*, en el que vemos cómo reacciona una persona al exponerla a información inconsistente con sus propias decisiones. Lo que nunca nos habíamos preguntado es cómo reacciona una persona ante información *verdadera* sobre sus decisiones. Esta es la puerta más importante que vemos abrirse y una en las que vamos a estar tratando de profundizar.

Mucho de lo que encontramos tiene dos elementos principales: por un lado, gran certeza estadística (p valores muy pequeños), por el otro, la capacidad de detectar efectos de tamaño muy chicos (una diferencia de pocos puntos para Acuerdo o Confianza). Esto hubiese sido absolutamente imposible sin la enorme cantidad de datos, lo que nos lleva a la última parte de este posteo: agradecer.

Ceguera CL no hubiese sido posible sin un montón de apoyo de personas e instituciones interesadas en hacernos preguntas y compartir las respuestas.

Los que pensamos, diseñamos, corrimos y analizamos Ceguera somos:

Rocco Di Tella: Líder de proyecto Ceguera CL | Diseño experimental | Análisis | Vínculos institucionales

Andrés Rieznik: Diseño experimental | Análisis

Lorena Moscovich: Diseño experimental | Comunicación | Vínculos institucionales

Rodrigo Catalano: Backend

Javi Goldschmidt: Frontend

Dardo Ferreiro: Diseño experimental | Líder de proyecto Ceguera BR

Juan Garrido: Diseño experimental

Facu Álvarez: Análisis

Pablo González: Líder de proyecto Ceguera AR/US | Diseño experimental | Análisis | Comunicación

No queremos dejar de mencionar personas e instituciones que bancaron la difusión y agradecerles: como siempre, a la Comunidad Gato ❤ que está siempre, firme para etiquetar, compartir y escribirle a alguien en el territorio que sea donde querramos hacer un experimento, así como a los vínculos que fuimos construyendo a medida que maduraba la idea de correr Ceguera en Chile. Hacer este experimento nos permitió colaborar con grandes equipos como Fundación Ciencia Ciudadana, Ciudadano Inteligente y El Mostrador: buscamos su apoyo tanto por difusión como por consejo sobre desarrollo experimental y en ambas áreas mostraron ser fundamentales. A Rossana Castiglioni, Olivia Sohr, Silvana Lauzan, Suzana Foxley, David Altman, Pierre Ostiguy, Patricio Navia por numerosas conversaciones que enriquecieron significativamente el proyecto. Y por su ayuda con la difusión agradecemos a Marcos Ortiz, Manuel Aris y Marco Enriquez Ominami.

elgatoylacaja.com/ceguera-diario-de-investigacion
