

Teoría de sesgos en el sistema educativo de la democracia del siglo XXI

Nuevas garantías para la libertad de pensamiento, el “Derecho a no ser engañados”¹

Theory of biases in the educational system of democracy in the 21st century

New guarantees for freedom of thought, the "Right not to be deceived"

Luis Miguel González de la Garza

Profesor de Derecho Constitucional y Teoría del Estado
Universidad Nacional de Educación a Distancia
E-mail: lmdegarza@der.uned.es

Resumen: La forma en la que la mente humana procesa la información es un aspecto fundamental de nuestra manera de aproximarnos a la realidad, íntimamente conectado con lo anterior se encuentra el problema de cuando la mente humana procesa esa información de forma errónea o defectuosa como consecuencia del procesamiento de esta a través de los sesgos cognitivos que nos caracterizan. Los sesgos cognitivos juegan un papel fundamental en nuestra visión de la realidad, del mundo que nos rodea y, por ello, su identificación es fundamental para prevenir los peores efectos de tales sesgos sobre el procesamiento de la información y los efectos que esas distorsiones de procesamiento pueden tener sobre

¹ Este Trabajo es resultado de proyecto de investigación financiado por el Ministerio de Ciencia, Innovación y Universidades RTI2018-096103-B-100 “Enseñar la Constitución. Educar en Democracia”, aprobado en la Convocatoria I+D+i “Retos de investigación” correspondiente a 2018.

nosotros. El conocimiento de los sesgos o de las heurísticas debería formar parte de la educación de las personas en los currículos formativos, si bien, es importante determinar el momento en el que esa educación debe incorporarse a la formación de los jóvenes. La formación de la ciudadanía en nuestras modernas democráticas de información debe conocer las fuentes de distorsión autónoma de procesamiento de la información que depende de nuestra propia naturaleza humana y evolutiva para poder así desarrollar estrategias que prevengan, en la medida de lo posible, tales errores, en muchos casos catastróficos, y que pueden conducir a sociedades mejor adaptadas para comprender y razonar en entornos cada vez más complejos y exigentes.

Palabras clave: Sesgo cognitivo; emociones; racionalidad; antivacunas; irracionalidad; educación; derecho a no ser engañados.

Abstract: The way in which the human mind processes information is a fundamental aspect of our approach to reality, intimately connected with the above is the problem of when the human mind processes that information erroneously or defectively as a result of the processing of it through the cognitive biases that characterize us. Cognitive biases play a fundamental role in our vision of reality, of the world around us and, therefore, its identification is essential to prevent the worst effects of such biases on information processing and the effects of these processing distortions. They can have on us. The knowledge of biases or heuristics should be part of the education of people in training curricula, although it is important to determine the time at which this education should be incorporated into the training of young people. The formation of citizenship in our modern democratic information must know the sources of autonomous distortion of information processing that depends on our own human and evolutionary nature in order to develop strategies that prevent, as far as possible, such errors, in many cases catastrophic, and that can lead to societies better adapted to understand and reason in increasingly complex and demanding environments.

Keywords: Cognitive bias; emotions; rationality; anti-vaccines; irrationality; education; right not to be deceived.

1. *Introducción*. 2. *Los Sesgos cognitivos*. 3. *El ser humano en su dimensión emocional*. 4. *Movimientos antivacunas*. 5. *Sesgos en el sistema BERT (Representaciones de Codificador Bidireccional de Transformadores) de Google*. 6. *Educación en la Universidad y sesgos cognitivos*. 7. *Tres ejemplos de arquitecturas de toma de decisiones colectivas diseñadas para neutralizar poderosos sesgos cognitivos en la toma de decisiones: el pre morten; la formación de los pilotos de aviación civil y los filtros colaborativos*. 8. *Conclusiones y discusión*. 9. *Bibliografía*.

1. Introducción

La *posverdad* como argumenta McIntyre (2008: 126) una derivación de la *posmodernidad* que es un movimiento filosófico-cultural y por lo tanto un *constructo humano cambiante y erróneo* que ha generado y genera daños incalculables a nuestras modernas sociedades democráticas (Castellanos: 2019: 353). Es muy diferente de los *sesgos cognitivos* que, en cambio, son una realidad evolutiva estudiada por la psicología científica moderna *que nos determinan permanente* en nuestra forma de *procesar la información*, por ello, deben enseñarse en los sistemas educativos.

Si podemos convenir en que cada persona puede tener su propia opinión sobre cualquier tema, nadie está en cambio autorizado a tener sus “*propios hechos*” sobre tales temas. Esto, que en el paradigma analógico era cierto puede dejar de serlo en el paradigma de la *realidad virtual*, singularmente cuando estos hechos pueden ser falsificados de forma tan precisa, eficaz y poco costosa que la realidad puede dejar de ser lo que hasta ahora ha sido para convertirse en un campo de excentricidades, dudas, omisiones de diseño intencional y certezas construidas *ad hoc* con una erosión severa y de alcance incalculable en la sociedad.

Frankfurt (2006: 83) señala que las civilizaciones *nunca* han podido prosperar, *ni podrán hacerlo*, sin cantidades ingentes de información *fiable sobre los hechos*. Tampoco pueden florecer si están acosadas por las problemáticas infecciones de creencias *erróneas*. Para crear y mantener una cultura avanzada es preciso que no nos dejemos debilitar por el error y la ignorancia. Necesitamos saber un gran número de verdades, y también, desde luego cómo hacer un uso productivo de ellas.

La protección *de la verdad* en la democracia del siglo XXI y en la vida en general es la esencia que hay que preservar frente a los ataques que van a ir in crescendo: primero fueron las *Fake News*, ahora se suman las *Deepfakes* y veremos nuevas dimensiones de estas patologías informacionales, pero ambas tienen un doble denominador común, *su ataque a la verdad mediante la tergiversación o falsificación de los hechos y la manipulación simétrica y explotación deliberada y consciente de los “sesgos cognitivos humanos”* que estudia la *psicología de la toma de decisiones*. Siendo *el objetivo de estas distorsiones* de muy diversa naturaleza los objetivos de manipulación, por ejemplo, de la opinión pública son los que en este análisis nos van a ocupar, ya que afectan a un elemento medular de la democracia. Una opinión pública nutrida de afluentes *no tóxicos* que permitan a los ciudadanos formarse juicios veraces basados en hechos ciertos y que puedan ser procesados cognitivamente con el menor número de distorsiones derivadas de nuestra propia naturaleza humana.

Los cambios que en la identidad humana se puede producir en virtud de las modernas tecnologías de la sociedad de la información pueden ser estudiados desde diversas perspectivas de análisis. Una de ellas es la que estudia esos posibles cambios *una vez que* las tecnologías empiezan a producir sus efectos, es decir *ex post*. En líneas generales estos análisis son reactivos y buscan soluciones *ad hoc* para los casos o *constelaciones de efectos* que se presentan en base a conjuntos de valores sociales y jurídicos establecidos en cada sociedad en un momento histórico dado. La capacidad “*disruptiva*” de muchas tecnologías que hoy se encuentran en la “*frontera del conocimiento*” toma por sorpresa a esas estructuras de valor y generalmente las desborda, porque no ha existido una planificación de posibles efectos, lo que es lógico porque la evolución de los efectos se genera en una *nebulosa de incertidumbre* de, seguramente, imposible control.

Pero, a nuestro juicio, la otra forma de estudiar y prevenir los efectos disruptivos, o al menos aquellos que podamos considerar más distópicos en lo que respecta a las modificaciones de la identidad humana tendría o podría tener su fundamento en la educación. Pero, no en cualquier tipo de educación, sino en aquella que *ex ante* fije su interés en las *distorsiones autónomas de la forma de procesamiento de la información por los seres humanos*. El cambio de la identidad humana por las tecnologías es una consecuencia, en parte, de que

somos víctimas de nuestros propios sesgos cognitivos como especie, en lo que afecta a: *la comprensión, a las conductas y a la acción*. La libertad de pensamiento es un presupuesto (Ferrajoli: 2011: 304) de la libertad de elección y de la participación crítica y consciente en la vida política. Nosotros diríamos más, de la vida *responsable* en la actividad general en todas sus esferas o dimensiones intelectuales fundamentalmente internas si bien con dimensiones externas cuando esa libertad se exterioriza. Aquí seguiríamos la definición de Hayek (2008: 37) de *libertad interior* que no se debe confundir en ningún sentido con el libre albedrío.

El desarrollo de la teoría de los sesgos cognitivos supone un hallazgo muy relevante en la forma en la que los seres humanos elaboramos nuestros juicios de forma inconsciente. Precisamente por ello y porque la función crítica y consciente se puede desarrollar del modo más amplio y profundo, libre de sesgos cognitivos que generan sistémicamente los errores de pensamiento, disponer en la actualidad de conocimientos explícitos sobre su forma de actuar en el procesamiento de la información por la mente humana hace posible llenar de contenido una libertad que carecía, hasta ahora, de un contenido sumamente especial, en vez de evanescente como lo podía ser otrora, que no es otro que disponer de una identificación precisa de heurísticas que al ser conocidas puedan ser neutralizadas o limitadas en su alcance y efectos perniciosos, saber cómo cometemos errores de procesamiento de la información distorsionada nos proporciona estrategias y herramientas para que nuestra libertad de pensamiento sea tal libertad ya que no se es libre con un pensamiento multidimensionalmente defectuoso producto de sesgos cognitivos. Nozick (1995: 112) tenía razón cuando sostenía que los filósofos se habían vuelto tecnológicamente obsoletos en el estudio de los procesos fiables de adquisición de creencias y que serían reemplazados por científicos cognitivos, así empieza a ser. Russell (2017: 191) señalaba que la educación debería tener dos objetivos: primero dar un conocimiento definido, lectura y escritura, lengua y matemáticas, y además; segundo, crear los hábitos mentales que permitirán que la gente adquiera conocimiento y se forme juicios sensatos por sí misma. Es a esta formación de juicios sensatos a la que se dirige no de forma única pero sí con otros instrumentos de razonamiento complementarios como la lógica, entre otros, la teoría de sesgos cognitivos ya que parece apta para la identificación de las heurísticas que no nos facilitan la elaboración de juicios sensatos.

2. Los sesgos cognitivos

La noción de sesgo cognitivo fue introducida por Daniel Kahneman y Amos Tversky (2012: 21) en 1972. Entendemos por sesgo cognitivo un fenómeno psicológico de naturaleza involuntaria ya que se produce de forma automática y que describe una alteración en la mente humana que la conduce a una *distorsión de la percepción*, a una distorsión cognitiva y o a juicios imprecisos o a interpretaciones ilógicas o irracionales (Sutherland: 2015: 27) de los hechos. Los sesgos cognitivos tienen siempre una razón de ser, tuvieron un origen evolutivo en la historia de la humanidad resultando de utilidad en un pasado remoto, el problema es que al cambiar las condiciones en las que los mismos tenían utilidad, las ventajas que podían tener esas formas de procesar la información de forma rápida y sin una carga cognitiva de atención focalizada, en un entorno moderno y con exigencias por completo distintas de en las que se desarrollaron esas *heurísticas*, se convierten en trampas que puede tener consecuencias letales para los individuos y de forma agregada para la sociedad.

Los hay de diversa naturaleza como la *Pareidolia*. *Si miras cualquier pared manchada con varias manchas o con una mezcla de diferentes tipos de piedras, si estás a punto de inventar alguna escena, podrás ver en ella un parecido con varios paisajes diferentes adornados con montañas, ríos, rocas, árboles, llanuras, amplios valles y varios grupos de colinas. También podrás ver diversos combates y figuras en movimientos rápidos, expresiones extrañas de rostros y disfraces extravagantes, y un número infinito de cosas que luego puedes reducir a formas separadas y bien concebidas.* Estas palabras las escribió Leonardo da Vinci en relación con un fenómeno denominado *Pareidolia*² un fenómeno de base neurobiológica, una ilusión como la describió Karl Jaspers (1997: 1056) que nos afecta a todos los seres humanos y que consiste en la tendencia hacia la percepción incorrecta de

² Se supone que esta capacidad es el resultado de miles de años de selección natural que favorecen a las personas más capaces de identificar rápidamente el estado mental a través de los sentimientos expresados en el rostro, por ejemplo, de amenaza a las personas de un grupo humano o una tribu, lo que brinda al individuo la oportunidad de huir o atacar preventivamente. En otras palabras, procesar esta información de manera subcortical, por lo tanto, inconscientemente, antes de pasarla al resto del cerebro para un procesamiento detallado, acelera el juicio y la toma de decisiones cuando se necesita una reacción rápida. Esta capacidad, aunque altamente especializada para el procesamiento y el reconocimiento de las emociones humanas, también funciona para determinar el comportamiento en los entornos evolutivos de la vida salvaje en la que se identifican animales que podrían ser potencialmente peligrosos y sus estados de actividad: caza; descanso; indiferencia.

un estímulo como un objeto, patrón o significado conocido por el observador, tales como ver formas en las nubes, ver caras en objetos inanimados o patrones abstractos, o escuchar mensajes ocultos en la música, entre otros muchos ejemplos. Desde luego no existen en la realidad esos patrones, son interpretaciones erróneas, ilusiones, distorsiones en suma de nuestro sistema de procesamiento de la información sensorial. Pero lo cierto es que nuestra mente se equivoca³.

Algunos investigadores como Funder y Krueger (2004: 318) han sugerido la posibilidad de ver los sesgos cognitivos no como errores, sino como *atajos mentales* empleados por la mente humana a la hora de *predecir* y tomar decisiones, sobre todo cuando *no hay mucha información disponible*, es decir, *bajo incertidumbre* que es -por otro lado- el modo habitual en el que solemos tomar decisiones los seres humanos, sin perjuicio de que el resultado de esos razonamientos será con alta probabilidad erróneo.

Lo anterior hay que ponerlo necesariamente en conexión con la forma en la que el cerebro humano toma decisiones, en este sentido, es esencial considerar la teoría revolucionaria de la “*mente predictiva*” (Ayan: 2019: 34) que confiere al *automatismo* de la mente un papel protagonista. La mente *nos permite predecir los acontecimientos venideros de forma rápida y segura cuando hay baja incertidumbre*. El aprendizaje, la experiencia y la consciencia tienen por finalidad *la mejora de las predicciones implícitas* que se generan de manera continua. Es más, para autores como Solms (2019:45) el “*consciente*” surge sólo cuando fallan las predicciones del cerebro. En este modelo precisa Seth (2019: 56), el cerebro realiza *inferencias bayesianas*⁴ generando sin cesar predicciones de las señales sensoriales y comparando estas predicciones con las señales que recibimos del mundo externo. La diferencia entre las señales predichas y las reales dan lugar a los llamados errores de predicción, dejando al cerebro dispuesto a nuevas tandas de predicción. Desde este punto de vista (Searle: 2018: 118) nuestras percepciones provienen de “*adentro hacia fuera*” en

³ Se equivoca de una forma análoga a lo que sucede con el ilusionismo. La magia no existe, el ilusionismo sí y podemos experimentar sus efectos de forma perfectamente perceptible e inevitable, sobre cómo la visión genera ilusiones falsas pero efectivas, puede verse en: Macknik, S.L, S. Martínez-Conde y Sandra Blakeslee, “*Los engaños de la mente. Cómo los trucos de magia desvelan el funcionamiento del cerebro*”, Destino, Barcelona, 2012.

⁴ Todos los modelos de inferencia bayesianos tienen en común la asignación de la probabilidad como medida de creencia de una hipótesis, por lo tanto, la inferencia es un proceso de reajuste de medidas de creencia al conocerse nuevos datos que reajustan la certidumbre de la inferencia inicial.

la misma medida, si no más, que de “*afuera hacia dentro*”. Más que un registro pasivo de una realidad objetiva externa, la *predicción* se erige como un proceso de construcción activa, *una alucinación controlada*, como se la ha acabado llamando. En línea con lo señalado, Friston (2020: 49) precisa que el cerebro elabora pronósticos sobre el mundo permanentemente y para ello emplea las reglas de la estadística bayesiana señaladas, de esta forma es posible sintetizar toda la actividad cerebral hacia una única meta: evitar las sorpresas de nuestro entorno en la medida de lo posible.

El primer paso pues para saber cómo falla nuestro razonamiento y protegernos frente a nosotros mismos y protegernos de quienes exploten tales “*defectos de procesamiento distorsionado de la información*” es conocer y comprender cómo funcionan tales sesgos, siendo la Educación, por lo tanto, basada en un método que permita disponer a las personas de las herramientas que faciliten la *identificación, diagnóstico y modificación* de lo que son “*pautas de pensamiento distorsionado*”: como: *dogmatismos, pensamiento de grupo, el antiintelectualismo* o el pensamiento basado exclusivamente en las “*emociones*”, la que debe jugar un papel relevante en el futuro de los menores y jóvenes. Son errores mentales como argumenta Matute (2019: 57) que cometemos sistémicamente todos y que se pueden predecir, dado que en todos nosotros ocurren en las mismas situaciones y funcionan de la misma manera y en la misma dirección. Una descripción general (Phol: 2004: 40) de estos sesgos son los que vemos seguidamente y que no agotan el catálogo de los hasta ahora identificados:

- *Anclaje (anchoring)*: es la tendencia a juzgar una situación con base en información recibida recientemente sobre ella. Cuando conocemos muy poco sobre un asunto, tendemos a confiar en la información que tenemos actualmente o que nos es proporcionada. Un ejemplo de este sesgo son las rebajas en los supermercados: cuando un producto es ofertado como un 30% más barato, asumimos que su precio es conveniente, sin tener en cuenta su valor absoluto o su relación con otros productos.
- *Aversión a la desposesión o aversión a la pérdida (efecto dotación)*: es la tendencia a preferir evitar las pérdidas en lugar de la posibilidad de adquirir ganancias (teoría de las perspectivas). Así, se demandará más dinero para renunciar a un objeto de lo

que se estaría dispuesto a pagar para adquirirlo. Vender bienes que uno normalmente usa activa regiones del cerebro asociadas al disgusto o la pena. Comprar también activa esas áreas, pero solo cuando los precios se consideran demasiado elevados, cuando se siente que un vendedor está recibiendo un dinero que excede el valor de cambio. Las imágenes del cerebro también indican que comprar a precios particularmente bajos es algo placentero. Hoy en día sabemos que el efecto dotación nos afecta incluso con sólo haber tenido el objeto en la mano durante unos instantes. Por eso mismo, muchas empresas invitan a probar productos y dejan que se prueben y posean todo lo que sea necesario. El vínculo sentimental que se establece con el objeto pone en marcha el efecto dotación, un poseedor se siente como propietario y se activa inmediatamente la aversión a perderlo, a no tenerlo. Deshacerse de algo, aunque se haga de forma voluntaria, se puede interpretar como una pérdida, y nadie desea perder.

- *Descuento hiperbólico*: Dadas dos recompensas similares, las personas muestran mayor preferencia por aquella que llegue más pronto que por una posterior. La tendencia se incrementará conforme ambos beneficios estén más cerca al presente. En este sentido, se descuenta el valor del beneficio posterior por un factor que aumenta con la duración del tiempo de espera.
- *Efecto Bandwagon o efecto de arrastre*: Es la tendencia a hacer (o creer en) algo porque muchas personas lo hacen (o lo creen). Está relacionado con el pensamiento de grupo o el comportamiento gregario. El efecto bandwagon (que tomó su nombre de la carreta que lleva la banda musical de un tren de circo) dicta que la probabilidad de que una persona adopte una creencia o conducta es *directamente proporcional a cuántos otros ya la tengan*, lo que significa que existe una tendencia psicológica a seguir o imitar las acciones y pensamientos de los demás, porque preferimos ajustarnos a lo preexistente, ya que es imposible no derivar nueva información de lo que otros piensan y hacen.
- *Efecto espectador (pasividad)*: Es un fenómeno psicológico por el cual es menos probable que alguien intervenga en una situación de emergencia cuando hay más personas que cuando se está solo.

- *Efecto denominación*: Es la tendencia a gastar más dinero cuando está denominado en pequeñas cantidades (por ejemplo, en monedas) más que en grandes (por ejemplo, billetes).
- *Efecto de encuadre*: Es la tendencia a alterar las decisiones según sea presentada una misma opción. Como resultado, se deducirán conclusiones diferentes a partir de la misma información dependiendo de la manera como se presenta la información. En concreto, las personas tienden a seleccionar opciones, dependiendo si la cuestión está enmarcada para concentrarse en las pérdidas o en las ganancias.
- *Ilusión del control*: Es la tendencia a sobreestimar el grado de influencia sobre otros eventos externos. De esta manera, los seres humanos tienden a creer que pueden controlar o al menos influir en las consecuencias o resultados que claramente no pueden controlar ni influir.
- *Sesgo de atención*: La tendencia implícita a que los estímulos emocionalmente relevantes en el propio entorno mantengan o limiten la atención de manera preferencial.
- *Sesgo de confirmación*: Es la tendencia de las personas a favorecer la información que confirme sus propios presupuestos o hipótesis, sin importar si la información es verdadera.
- *Sesgo de congruencia o sesgo de compatibilidad*: La tendencia a comprobar las hipótesis exclusivamente por medio de pruebas directas, en contraste con las pruebas de posibles hipótesis alternativas.
- *Sesgo de distinción*: Es la tendencia a ver dos opciones como más diferentes cuando se las evalúa simultáneamente que cuando son evaluadas de manera separada.
- *Sesgo de impacto*: Es la tendencia a sobrevalorar la duración e intensidad de los futuros estados emocionales, basándose en experiencias previas.
- *Sesgo de información*: Es la tendencia a buscar información, incluso cuando esta no puede afectar a la decisión a tomar. Puede crear la falsa impresión de que, por tener más información, el razonamiento o la conclusión será más veraz.
- *Sesgo del experimentador*: Cuando los experimentadores tienden a creer, certificar y publicar datos que concuerdan con sus expectativas con respecto al resultado de

- un experimento y desechar, desacreditar o infravalorar las ponderaciones correspondientes a los datos que parezcan estar en conflicto con sus expectativas.
- *Sesgo del status quo*: Las personas actúan de la misma forma, se quedan con lo que tienen a menos que exista una poderosa razón para cambiar, o incluso aunque exista una buena razón para cambiar no cambian. La aversión a las pérdidas y el sesgo de status quo suelen producirse de forma asociada creando una poderosa fuerza que inhibe la propensión al cambio.
 - *Sesgo de punto ciego*: Es la tendencia a no darse cuenta de los propios prejuicios cognitivos o a verse a sí mismo como menos sesgado que los demás.
 - *Sesgo retrospectivo o prejuicio de la elección comprensiva*: Es la tendencia a recordar las decisiones propias como mejores de lo que realmente fueron.
 - *Prejuicio de desconfirmación o sesgo de disconformidad*: Es la tendencia a realizar un crítico escrutinio de la información cuando contradice sus principales creencias y aceptar sin criterio aquella información que es congruente con sus principales creencias.
 - *Percepción selectiva*: Tendencia en la cual las ansias, esperanzas o ilusiones afectan a la percepción.
 - *Efecto del falso consenso*: Se refiere al hecho experimentalmente comprobado de que la mayoría de las personas juzgan que sus propios hábitos, valores y creencias están más extendidos entre otras personas de lo que realmente están, ya sean por motivación, sesgo de confirmación o percepción selectiva. Es la tendencia a sobrestimar el grado en el cual los demás coinciden o están de acuerdo con ellos. Las personas cambian rápidamente sus propias opiniones, creencias y predilecciones para parecer más acordes o en sintonía con el público general. Como los miembros de un grupo alcanzan un consenso y raramente encuentran personas que disientan, tienden a creer que todo el mundo piensa del mismo modo.

El campo de la economía se ha adaptado paulatinamente en los últimos veinte años a incluir los estudios conductuales y la teoría de sesgos en su desarrollo más profundo hasta el punto de impulsar con extraordinaria pujanza la denominada “economía conductual” en la que se

han concedido diversos Premios Nóbel.⁵ La economía conductual centrada principalmente en la microeconomía experimenta desarrollos basados en la superación de modelos erróneos contruidos sobre premisas falsamente aceptadas durante décadas, como la racionalidad perfecta de los agentes económicos (Thaler: 2017: 36) como dijo Mark Twain “*lo que te mete en problemas no es aquello que no sabes; es aquello que estás seguro de que sabes, y que resulta que no es verdad*”. Las personas tienden a tener exceso de confianza porque no se molestan en documentar su registro de predicciones pasadas equivocadas, y posteriormente empeoran las cosas cayendo en el error de pernicioso sesgo de confirmación: tan sólo buscan la evidencia que confirme sus hipótesis preconcebidas. La única protección posible contra el exceso de confianza es la búsqueda sistemática de datos y, en particular, de datos que demuestren que nos equivocamos.

Pensamos que por análogas razones a por las que la psicología conductual ha corregido y forma ya parte del cuerpo de doctrina económica más avanzada, ésta debe también ser introducida en la educación como corrección de modelos anticuados o incompletos que producen en la formación múltiples y sistémicos errores de pensamiento derivados de nuestra propia naturaleza. Conocer los errores de procesamiento mental de la información es esencial para poder mitigarlos y desarrollar estrategias educativas que tiendan a su minimización o evitación en la sociedad del futuro.

3. El ser humano en su dimensión emocio-racional.

Hay que recordar, como señala Damasio (2010: 140) que cuando preguntamos por qué desarrollamos las artes, la ciencia, la tecnología, los gobiernos... una razón importante es que desarrollamos esas instituciones por la motivación que recibimos a partir de nuestro sistema de creencias. Habiendo pasado desapercibida la idea de que los sentimientos de dolor o sufrimiento o de *intentar estar bien*, motivan toda la creación de esas prácticas e instrumentos de la cultura, es decir, que los *sentimientos* son los cimientos de nuestra mente. Hay que señalar que, el concepto de “*dignidad humana*” pocas veces ha sido debidamente evaluado desde esta perspectiva. Por ello no es una cuestión menor, pero es una cuestión que debe ser modulada como veremos más adelante. Debemos precisar con López Rosetti

⁵ Daniel Kahneman; Vernon Smith; Richard Thaler.

(2018: 16) que los seres humanos no somos seres prioritaria o fundamentalmente racionales –si así fuera seríamos máquinas de procesamiento de información análogas a los ordenadores- sino, *seres emocionales que razonan*, que obviamente no es equivalente. No sólo somos irracionales, sino previsiblemente irracionales; (Ariely: 2019: 17) es decir que nuestra irracionalidad se produce siempre del mismo modo una y otra vez. Durante los cientos de miles de años de evolución de nuestra especie no existió la razón ni el pensamiento racional, solamente existían las emociones primero y los sentimientos después y hace muy poco tiempo, evolutivamente hablando, llego a través de la evolución el pensamiento racional con el lenguaje. Así lo señalan los brillantes estudios de Haidt (2019: 57). En efecto, *el procesamiento de la información emocional* no es ni Jeffersoniano, ni Platónico, es *Humeano*⁶. Los seres humanos hacen juicios morales de manera rápida y emocional y el razonamiento moral es, principalmente, una búsqueda *post hoc* de razones para justificar los juicios que ya se han hecho, unos de forma consciente y otros de forma inconsciente. Lo anterior es consistente con los hallazgos de Khaneman (2012: 45) en relación con el Sistema 1 y el Sistema 2 de procesamiento de información cerebral. Debemos precisar que el Sistema 1 es el responsable habitual de los grandes errores en el procesamiento de la información, singularmente en aquella que no evalúa correctamente las informaciones falsas, razón por la cual Zimbardo (2019: 551) advierte sobre la necesidad imperativa de mantener la máxima atención contra las *influencias situacionales*.

Esa educación en la *psicología de la toma de decisiones* pensamos que es fundamental en una sociedad cada vez más atomizada y que precisamente por esa atomización en más sensible al uso de los *sesgos cognitivos con finalidades disruptivas*. Es decir, si las personas saben cómo piensan y qué defectos tiene el pensamiento, estamos previniendo su abuso por parte de quienes empleen esos sesgos, bien conocidos para el dominio de las personas y por lo tanto para su explotación intencional.

Siempre ha sido cierta aquella expresión que señalaba que no hay nada más peligroso que la ignorancia en acción, hoy habría que añadir que no hay nada más peligroso que operar con *pensamientos distorsionados*. Y vivimos en la era de la “*distorsión del pensamiento*”.

⁶ Platón y la razón (Timeo) como elemento central del hombre; Jefferson postula una idea intermedia y doble: emoción – juicio -razón. Y Hume: La razón no es más que el sirviente de las pasiones.

Señalemos que no estamos haciendo referencia a pensamientos clínicos, sino a la prevención de formas de pensamientos que, en la actualidad, tenemos ya fundamentalmente bien identificados y que representan factores que, debidamente conocidos, pueden suponer formas adecuadas de *corregir, limitar y por ello resolver los efectos que esos sesgos producen en el procesamiento de la información* con la esperanza de que cuanto menor sea el pensamiento o razonamiento distorsionado la conducta de los ciudadanos podrá ser, sobre todo más *resistente al engaño*, a las falsedades, lo que puede contribuir a que la sociedad pueda disponer de instrumentos más adecuados para razonar en entornos complejos donde precisamente los “*sesgos cognitivos producen los peores efectos*”. Es importante insistir en que los sesgos cognitivos una vez identificados pueden *contrarrestarse* y de ahí se desprende la importancia de su conocimiento, veamos un ejemplo. En el *sesgo de negatividad* (Rozin y Royzman: 2001: 297) una sencilla expresión lingüística como decir “*bien*” no es suficiente si se trata de transmitir entusiasmo a otra persona o a un auditorio, siendo necesario que esa expresión se rodee de un mayor énfasis si se pretende que se enfrente a otra de carácter *negativo* como “*mal*”. En cambio, una valoración *negativa* de carácter lingüístico no necesita de soporte adicional ya que su efecto es dominante (Liebrecht y colaboradores: 2019: 170), es decir, *las palabras negativas causan mayor impresión que las positivas*, pero cuando se intensifican los adjetivos positivos (con términos como: *extraordinario, maravilloso, sin igual*) el sesgo de negatividad disminuye.

Podemos identificar concretas formas de pensamiento distorsionado, como por ejemplo aquellas que tienen que ver con los “*movimientos antivacunas*” grupos extensos de personas que confían en la “*mentira*”. Pensamiento distorsionado que se ha visto extraordinariamente potenciado en una magnitud exponencial gracias a las redes sociales, a sus efectos polarizadores de grupos de retroalimentación positiva y a su capacidad basada en la “*homofilia*” de concentrar a aquellas personas que comparten semejantes sesgos cognitivos. Uno de los principios fundamentales para comprender las redes sociales, señala Kadushin (2013: 315) es el de la *homofilia*, expresión que proviene del griego y cuyo significado es “*amor a lo mismo o los iguales*” concepto introducido en la teoría social por Lazarsfeld y Merton en 1978. Este es el principio que se encuentra tras aquellas palabras de Karl Liebknecht: “*Los pájaros del mismo plumaje siempre vuelan juntos*” y conduce, directamente, a los sistemas sociales ya que es de aplicación tanto a personas como a grupos

o colectivos. La *homofilia* se produce a través de la mera proximidad, parentesco o afinidad, estatus comunes u otros atributos socialmente definidos incluido atributos como el color de la piel o, simplemente, por coincidir en el mismo lugar en el mismo momento. La *homofilia* es un proceso y también un resultado, las personas con los mismos valores y actitudes tienden a asociarse con otras semejantes. La interacción conduce a su vez a una mayor probabilidad de tener valores comunes. Las personas con la misma posición estructural o la misma clase, ocupación y formación tienden a compartir los mismos valores y a participar en actividades comunes y de ese modo tienden a asociarse unas con otras. Las redes sociales basadas en la homofilia son radicalmente injustas; la *homofilia* dificulta en gran medida a los planificadores sociales la superación de la segregación y de los prejuicios de cualquier clase. Existe también un proceso de retroalimentación implícita: con el tiempo, las relaciones tienden a evolucionar de forma que se hacen más *homófilas*. Existen dos tipos de causas para que se produzca la *homofilia*. Las normas o valores comunes pueden unir a *nodos* con atributos comunes, o a la inversa, los atributos y contactos comunes pueden conducir a normas comunes, siendo esto así en el caso tanto de individuos como en el de colectividades y surgiendo en redes pequeñas como grandes.

El circuito de retroalimentación entre la estructura de la red y la preferencia individual se va haciendo especialmente acusado con el paso del tiempo estando implicados cuatro procesos:

- a) *Los mismos tipos de personas se juntan.*
- b) *Las personas influyen unas en otras y durante ese proceso van pareciéndose cada vez más.*
- c) *Las personas pueden acabar en el mismo lugar.*
- d) *Una vez están en el mismo lugar, ese mismo lugar influye en ellas para que se parezcan más.*

En el proceso que acabamos de describir, se superpone el “*Sesgo de confirmación*” es decir, la tendencia de las personas a favorecer la información que confirme sus propios presupuestos o hipótesis, sin importar la información verdadera.

En ausencia de redes sociales las comunicaciones estaban mediadas por concentradores, es decir por *intermediarios selectivos*, estos podían ser los *consejos de redacción de la prensa, de la televisión o de la radio*. Pero en las redes, cualquier persona puede ser un concentrador y lo que se ha comprobado *es que las redes lo que unen es a las personas parecidas*. Los medios, por ello, tenían una capacidad disolvente de los peores efectos de la concentración y de la ignorancia. Tenían capacidad selectiva y eso orientaba al ciudadano que *no se las tenía que ver consigo mismo* como selector de información incompetente, es decir, como alguien que diseña el modelo informativo que recibe y que ofrece. Una vez que esas personas empiezan a interactuar se activa el *“tribalismo” que forma parte de nuestra herencia evolutiva* y con él un conjunto de actitudes esenciales del mismo, como aferrarnos estrechamente a nuestro grupo, asumir y defender la matriz moral del mismo y dejar de pensar por nosotros mismos, lo que es una consecuencia de un principio básico de la psicología moral y es que *“la moralidad une y ciega”*, ceguera moral que como señala Goleman (1997: 265) se distingue por consistir en la creencia implícita en la justicia y moral de las decisiones del grupo mediante una falsa suposición que permite a sus miembros ignorar las consecuencias morales de sus decisiones.

4. Movimientos antivacunas

Las personas necesitan abrigar un sólido sentimiento de identidad y de su lugar en el mundo (Hog: 2019: 75) y para muchas personas el ritmo y la magnitud de los cambios pueden resultar alienantes, la incertidumbre amenaza seriamente nuestra identidad y los grupos sociales constituyen una poderosa fuente de identidad que atenúa la incertidumbre personal, en particular, si los grupos son distintivos y sus miembros comparten un sentimiento de interdependencia. Interdependencia que se reafirma con el sesgo *Bandwagon* o efecto de arrastre, que es la tendencia a creer o hacer algo porque muchas personas lo creen o lo hacen. Es un efecto del pensamiento de grupo o tribal.

Las Redes sociales son en la actualidad un lugar ideal para mitigar el desasosiego que provoca la incertidumbre personal porque proporciona acceso sin interrupción a información ilimitada, que con mucha frecuencia seleccionan los propios individuos y de manera inadvertida pero altamente eficaz, también los algoritmos que aprenden de las interacciones de los individuos para mostrarles lo que habitualmente más buscan o desean.

De este modo la gente sólo accede a la información que confirma su identidad. El *sesgo de confirmación* actúa y se acentúa en tiempos de incertidumbre, separando los universos de la información y de la identidad que fragmentan y polarizan la sociedad. *La conexión en línea permite dar con grupos que no se encontrarían con facilidad en el entorno físico*. Por ello Internet fortalece aún más el sesgo de confirmación porque las personas necesitan rodearse de quienes piensan igual que ellos para que se confirmen continuamente sus identidades y visiones del mundo. Los contornos de la “*verdad*” entonces se ajustan a estos universos de identidad social autosuficientes y autorreferenciales que generan una clausura operativa de los grupos en los que se genera confianza (Luhman: 2005: 39) ya que ésta juega un papel relevante en la reducción de la complejidad del universo de opciones informativas rivales y complejas (Hardin: 2010: 103) haciendo posible y facilitando lo que Laski (2017: 32) denominara la *seducción de la mera inercia* que es siempre poderosa y que nos permite, en efecto, trazar la pequeña senda de nuestra existencia sin que nos sintamos amenazados por ningún desastre inminente. Nos ahorra la molesta tarea de someter a examen hábitos que siempre resulta peligroso analizar y, a veces, pernicioso destruir. Con todo -advertía Laski- no hay duda de que el proceso de la *inercia* es, a la larga, la pérdida del sentido cívico entre la multitud. A lo anterior hay que añadir un grave problema aún no resuelto (Sunstein: 2019: 84) que es cómo hacer la *disensión* menos costosa o incluso recompensar esta disensión especialmente cuando quienes disienten no se benefician ellos mismos sino a otros. Es decir, cómo se deberían diseñar estructuras de jerarquización de información en los grupos que posibiliten la disensión en la que la información privada de los miembros del grupo no quede sepultada por el temor de las reacciones del grupo. Aquí el diseño institucional puede jugar un papel importante. El problema central de la difusión de la *conformidad* es que oculta a los miembros del grupo información valiosa que debe conocer.

En estos escenarios no hay “*verdades absolutas*” u objetivas, ni motivaciones para explorar e incorporar otros puntos de vista, ya que estos debilitarían la capacidad de identidad social para evitar la incertidumbre personal. Lo que se denomina *sesgo de desconfirmación* que es la tendencia a realizar un escrutinio absolutamente crítico de la información cuando contradice las principales creencias del sujeto y aceptar sin criterio aquella información que es congruente con sus creencias principales bien sean personales o del grupo al que

pertenece. Este sesgo intensifica a su vez el *sesgo de inercia* que es la tendencia a evitar, siempre que sea posible, tomar decisiones que supongan cambios, deseando que las cosas sigan como están generando un circuito cerrado y circular de polarización grupal *en equilibrio*. La polarización social conduce a la reducción de la necesaria diversidad social que es, tan importante, como la biodiversidad. No sólo alimenta la inteligencia colectiva y la innovación (Helbing y otros: 2018: 74) sino que mejora la *resiliencia social*⁷ es decir, la capacidad de nuestra sociedad de gestionar las conmociones inesperadas. Con frecuencia la reducción de la *sociodiversidad* también merma o disminuye simétricamente el rendimiento de la economía y de la sociedad en su conjunto.

Y es frecuente que, a su vez, en el grupo o en miembros del grupo se manifieste el *sesgo de reactancia*, que es la tendencia a que se tomen decisiones por el mero hecho de contradecir a los demás miembros de otros grupos como *reafirmación de la propia autonomía grupal*. Será frecuente que en estos grupos se halle presente el pensamiento *conspirativo* (Quattrociocchi: 2018: 26) que se caracteriza por la incapacidad de asignar a los hechos adversos un determinante causal (fuerzas de mercado, presión evolutiva, complejidad), lo que, según Martin Bauer, implica un modo casi religioso de pensar en los procesos. Aunque Internet garantiza el acceso a cantidad ingente de información (Pluviano y Della Sala: 2019: 16) también ha difundido la idea de que basta utilizar un buscador y confiar en la veracidad de los resultados para hacerse un experto en un tema. Este *sesgo cognitivo* se conoce como efecto “*Dunning-Kruger*”. En 1999 los psicólogos Justin Kruger y David Dunning, describieron la *incapacidad metacognitiva* de las personas *inexpertas* de reconocer las propias limitaciones y su tendencia a sobrevalorarse. “Todos somos idiotas seguros de nosotros mismos”. Con suma frecuencia “*el ignorante no sabe que lo es*”. En cambio, tiene una confianza exagerada e ilusoria en sí mismo y en creerse en posesión de la verdad. Se trata de autosugestión. Por esa razón, es necesario recurrir a expertos que ayuden a

⁷ El concepto de resiliencia tiene su origen en el campo de la física y de la ingeniería. En tal sentido, se entiende por resiliencia la magnitud que cuantifica la cantidad de energía que un concreto material puede absorber o almacenar al deformarse elásticamente, pudiendo romperse o recuperarse dicha deformación producida por efecto de un impacto, por unidad de superficie de rotura. La resiliencia se distingue de la tenacidad en que ésta cuantifica o mide la cantidad de energía absorbida, por unidad de superficie de rotura, bajo la acción de un esfuerzo progresivo y no por impacto, como en el caso de la resiliencia.

diferenciar hechos e informaciones falsas que difunden personas incompetentes más o menos conscientes de ello.

Nuestra mente no es reproductiva sino reconstructiva como señalan Pluviano y de Della Sala (2019: 22). Y en esa reconstrucción de la realidad comete inevitablemente errores ya que se basa en los propios prejuicios y *marcos de referencia*⁸. Nuestra mente asimila y recuerda los hechos al vincularlos en un marco de referencia (Lakoff: 2016: 393) creando una narrativa que tratamos de mantener coherente. Por ello, *desmontar una información* incluso si se revela totalmente incorrecta o falsa *deja un vacío* en el modelo mental, por lo que preferimos un modelo de pensamiento falso pero completo, es decir coherente, a uno parcial e incoherente y que no encaja con nuestros conocimientos. El autoengaño (Mele: 2016: 111) juega un papel relevante en el mantenimiento de ese estado mental cómodo pero falso. Esto es de especial trascendencia en la psicología del testimonio y en la *falibilidad* de éste como argumenta Mazzoni (2019: 78) La clásica expresión “*Testis unus testis nullus*” “Un testigo solo, testigo nulo”. Paulo: Digesto 48, 18, 20, *unius testimonio non esse credendum* (no se debe dar crédito a un único testimonio) y Ulpiano: Digesto 22, 5, 12, *ubi numerus testium non adicitur, etiam duo sufficient* (donde no se exprese el número de testigos, bastarán dos) atesora una intuición neurobiológicamente correcta a lo largo de la historia que coincide con lo que nos muestra la ciencia de la conducta. En efecto y como precisa Gazzaniga (2006: 129-138) de todas las cosas que recordamos, lo más sorprendente es que algunas sean ciertas. Los errores de distorsión pueden corromper la memoria en cualquier fase: desde la codificación de un recuerdo hasta cada una de las ocasiones en que se recupera o reactiva dicho recuerdo. Algunos errores añaden información falsa a los recuerdos (errores de obra) y otros se deben a que olvidamos u omitimos información (errores de omisión). Además, parece que la mera repetición de la información errónea incrementa la probabilidad de que la información se recuerde posteriormente como cierta con mayor seguridad. Es lo que se denomina «efecto de verdad ilusoria». Por último, varios estudios sobre la inserción de recuerdos falsos emplean un paradigma, llamado «*paradigma de Deese-Roediger-McDermott*», o DRM, por los investigadores que lo desarrollaron, para

⁸ Los marcos crean la realidad social enfocando nuestra atención en el asunto inmediato y desviándola al mismo tiempo de todo lo que no tenga relación directa con él. Todo aquello que queda fuera del marco no existe. Op Cit., “*El punto ciego. Psicología del autoengaño*”, Plaza y Janés, Barcelona, 1997.

estudiar los recuerdos falsos. Como ejemplo del paradigma DRM, se puede hacer la siguiente prueba. Se puede pedir a una persona que lea la siguiente lista de palabras y que intente recordar tantas como pueda: *cama, descanso, despierto, cansado, dormir, cabezada, soñar, despertar, manta, sueño, modorra, roncar, siesta, tranquilidad, bostezo, somnoliento, ensoñación*. Ahora puede esperarse cinco minutos y pedir a esa persona que escriba o repita todas las palabras que recuerde.

¿Ha escrito la palabra *dormir*? Esta palabra, denominada «*señuelo crítico*», no figuraba en la lista, y sin embargo la mayoría de las personas recuerda haberla visto, a menudo con gran seguridad. Dado que todas estas palabras están relacionadas con *dormir*, es probable que este vocablo se active en la mente del sujeto, provocando la confusión entre el pensamiento interno (consciente o inconsciente) de la palabra *dormir* y la aparición real de la palabra. Es un buen ejemplo de cómo interfiere el cerebro en la información que recibe, distorsionándola antes de que se almacene. Como estudio Schacter (2003: 23) diversas disfunciones afectan de forma muy importante a la memoria y por lo tanto a los recuerdos, nos afectan a todos tales como: el *transcurso*, la memoria se desvanece; la *distractibilidad*, atención y recuerdo; el *bloqueo*, en la punta de la lengua; la *atribución errónea*, atribuciones erróneas de testigos oculares; la *sugestionabilidad*, la influencia sobre los testigos: la *propensión*, cómo éramos depende de cómo somos; y la *persistencia*, los recuerdos con etiquetas emotivas.

A los sesgos señalados tenemos que unir que vivimos en un mundo cada vez más polarizado en el que la verdad queda a menudo sepultada por desinformación que procede verosímelmente de dos fuentes, aquella que se originan en propagadores profesionales e intencionales de falsedades destinadas a sumir en el error a las personas y a los grupos con la finalidad de generar *inestabilidad aguda o crónica* en función de sus pretensiones de poder: político; de mercado o social, y aquella otra que como se establece en el principio de la navaja de Hanlon: “*Nunca le atribuyas a la maldad lo que puede ser explicado por la estupidez*” que como advertiera Goethe: “*Los malentendidos y la negligencia crean más confusión en el mundo que el engaño y la maldad. En todo caso, estos dos últimos son mucho menos frecuentes*”. Las dos fuentes se refuerzan y actúan sinérgicamente con efectos devastadores. La suma de las dos fuentes origina necesariamente *incertidumbre* y según la

expresión que se le atribuye a Immanuel Kant: “*La inteligencia de un individuo se mide por la cantidad de incertidumbres que es capaz de soportar*”, esa incertidumbre es - precisamente- la que mitigan los sesgos cognitivos. En nuestro mundo de conectividad informativa extrema la incertidumbre que proviene de las redes sociales hace posible que tales personas bajo esa incertidumbre adopten y den por ciertas y validas, construcciones irreales, falsas y ficticias que generan graves daños para sí mismas y para la sociedad, organizadas entorno a grupos mesiánicos como los anti vacunas como hemos considerado u otros movimientos como los “terraplanistas” aquellos grupos que piensan que la tierra es plana contra toda evidencia científica y racional.

5. Sesgos en el sistema BERT (Representaciones de Codificador Bidireccional de Transformadores) de Google

El otoño pasado, Google presentó una tecnología de inteligencia artificial innovadora llamada BERT que cambió la forma en que los científicos en lingüística construyen sistemas que aprenden cómo las personas escriben y hablan. BERT⁹ es una red neuronal de código abierto que ha sido entrenada para procesar el lenguaje natural. BERT procesa las palabras en el contexto de una oración, en lugar de palabra por palabra.

Pero BERT, que ahora se está implementando en servicios como el motor de búsqueda de Internet de Google, tiene un problema: podría estar detectando y asimilando sesgos en la forma en que un niño imita el mal comportamiento de sus padres. BERT es uno, de una serie de sistemas de inteligencia artificial que aprenden de mucha información digitalizada, tan variada como libros antiguos, entradas de Wikipedia y artículos de noticias. Las décadas e incluso siglos de sesgos, junto con algunos nuevos, probablemente se encuentran en todo

⁹ Técnicamente, BERT es un modelo de Redes Neuronales Artificiales (RNA) aplicado al campo del Natural Language Processing (NLP), específicamente al subcampo del Natural Language Understanding (NLU). En otras palabras, BERT es una técnica de Inteligencia Artificial (IA) que le permite a un sistema informático comprender mejor el lenguaje humano. BERT se convierte en la segunda implementación de IA más importante en el algoritmo, después de RankBrain. Su objetivo es ayudar al motor de búsqueda de Google a tener una comprensión mucho más profunda de las intenciones e intereses de búsqueda de los usuarios. Esto al analizar no solo las palabras clave o keywords, sino también el resto de las palabras a su alrededor, las cuales le “dicen” al algoritmo cuál es el contexto y la temática específica de cada intención de búsqueda.

ese material, es decir, dado que el sistema BERT opera sobre textos que contienen cientos de miles de datos con sesgos, estos necesariamente se recogerán y perdurarán.

BERT y sus sistemas equivalentes tienen más probabilidades de asociar a los hombres con la programación de computadoras, por ejemplo, y generalmente no otorgan suficiente crédito a las mujeres como ha señalado Kurita y otros (2019: 31). Un programa informático de procesamiento de lenguaje natural decidió que casi todo lo escrito sobre el presidente Trump era negativo, incluso si el contenido real era positivo, es decir, cuando el algoritmo de procesamiento de la información se retroalimenta de información sesgada se genera un proceso en el que la información verdadera es omitida. Pero lo más significativo del sistema BERT u otros sistemas equivalentes es su capacidad para replicar o reproducir sesgos allí donde los encuentre y reproducirlos en formulaciones actualizadas de resultados que los recogen. Informaciones sesgadas almacenadas en repositorios de la más variada naturaleza, una vez digitalizada y procesada reproducirá los sesgos que en ella se encuentren. Además, BERT tiene la capacidad de realizar *evaluaciones emocionales*¹⁰ haciendo posibles clasificaciones de esta naturaleza, lo que en el ámbito de los sesgos cognitivos emocionales que venimos considerando facilita el uso de estas clasificaciones con las más diversas finalidades siendo singularmente peligrosas en el ámbito de la propaganda cognitiva virtual, la transformación social basada en emociones y el uso de datos emocionales con finalidades comerciales.

6. Educación en la Universidad y sesgos

En los Estados Unidos nos recuerdan Haidt y Lukianoff (2019: 39) en su importante monografía, se está sufriendo lo que podríamos denominar una infección cultural que se basa en tres postulados falaces pero que provienen, en su mayor parte de las *redes sociales y de la polarización*, estas son tres ideas disruptivas que operan de forma conjunta.

La falsedad de la *fragilidad* de las personas y su postulado: “*Lo que no te mata te hace más débil*”. La falsedad *del razonamiento emocional* y su corolario: “*Confía siempre en tus*

¹⁰ En promedio, existen 200 factores que Google tiene en cuenta para clasificar las páginas web en las SERP, pero no todos ellos están relacionados con el análisis del contenido textual. Pueden verse los factores, también los emocionales en: <https://backlinko.com/google-ranking-factors>

sentimientos” y la falsedad del pensamiento dicotómico y de su consecuencia natural: “*nosotros contra ellos*” en el que la vida es una batalla entre las buenas y las malas personas. En buena medida estas ideas podrían ser identificadas como una sobre simplificación e infantilización de determinados estratos sociales, pero son el efecto de errores derivados en amplia medida de sesgos cognitivos. Haidt y Lukianoff señalan tres argumentos básicos que deben cumplirse simultáneamente para identificar la falsedad de una idea, estos son:

En primer lugar, la idea distorsionada contradice el saber antiguo (ideas que se encuentran muy a menudo en la literatura sobre la sabiduría de muchas culturas). En segundo lugar, contradice la investigación psicológica moderna sobre el bienestar y por último perjudica al individuo y a las comunidades que la adoptan. Estas tres grandes falsedades propias de la época de la *posverdad* son concreciones que se han articulado en medidas y movimientos políticos que las respaldan y están causando problemas graves a los jóvenes, a las Universidades es decir a su función educadora y a las democracias liberales.

La primera falacia identificada afirma que los estudiantes son frágiles, los estudiantes están empezando a demandar protección *ante el lenguaje* porque han aprendido involuntariamente a emplear las mismas distorsiones cognitivas que la Terapia Cognitivo Conductual intenta corregir. Muchos padres, maestros de primaria y secundaria, así como profesores y administradores de universidades habían enseñado inconscientemente a una generación de estudiantes a desarrollar hábitos mentales comúnmente presentes en personas que padecen ansiedad y depresión.

Los estudiantes han aprendido a reaccionar con temor e ira a las palabras, los libros y los oradores invitados como ponentes a las universidades porque se les ha enseñado a *exagerar el peligro*, a emplear el pensamiento dicotómico (o binario), y a magnificar sus primeras *reacciones emocionales* y a desarrollar una peligrosa sensibilidad a las denominadas *microagresiones* en los campus universitarios que suponen un cambio moral significativo, ya que representa el giro de la valoración de los actos, pasando de la *intención* de los actos al *efecto* de los mismos, siendo *el efecto* el detonante de las reacciones y despreciando o no valorando la intención. En el pensamiento emotivo se encuentra la “*seguridad emocional*”. La *seguridad emocional* hace creer que los alumnos deben estar a salvo no sólo de los accidentes físicos, sino también de las personas y *de las ideas* que están en desacuerdo con

las que se tienen. Es decir, el *emotivismo* alimenta movimientos identitarios transversales productores de intolerancia, favorece una cultura de la *acusación pública* y finalmente de la *autocensura individual* y de la *censura colectiva* que se puede articular jurídicamente a través de las “*leyes mordaza*” (Coleman: 2018: 109) o contra el discurso del odio¹¹ en sociedades cada vez más cerradas, y a la vez fragmentadas en bloques polarizados irreductibles y tendentes no a la libertad de expresión sino a la homogeneidad en las *expresiones políticamente correctas*. El cambio de paradigma es fundamental en la universidad norteamericana, pasando de la idea de que la intención de la educación no debería hacer sentirse cómodas a las personas: su propósito es hacerlas pensar, al paradigma de que la intención de la educación no debería hacer pensar a las personas, sino que su sentido es hacerlas sentir cómodas lo que supone claramente un error categorial de inmensa magnitud.

Lo anteriormente señalado tiene un fuerte impacto negativo sobre los postulados centrales de lo que hemos denominado las *modernas democracias deliberativas* (Elster: 2000: 129) en las que se representa la permanente tensión entre la *política del dogma* y la *política del descubrimiento* como señalara Deutsch (1976: 230) a través del diálogo y del debate informado. El aislamiento de los grupos su radicalización y enclaustramiento influenciados hoy más que nunca por los sesgos cognitivos actuando y potenciándose a través de la redes sociales levantan muros de aislamiento que hacen cada vez más difícil el debate y la discusión honesta, es decir, aquella orientada *por la política del descubrimiento*, en vez de por la política que podríamos denominar de los “*micro-dogmatismos*”, es decir, aquella que no se residencia en un sólido bloque ideológico dogmático sino en un cosmos o miríada de dogmatismos cruzados o transversales que se difunden por toda la sociedad. El aislamiento de los grupos es un elemento esencial para debilitar la confianza en un espacio público tendencialmente abierto y dispuesto a la confrontación pública de las ideas. La ruptura del espacio común de la opinión pública con grupos cada vez más cerrados y autorreferenciales deteriora gravemente la democracia deliberativa precisamente porque afecta a la capacidad de deliberación abierta en las sociedades abiertas (Popper: 1957: 392) cuando los *micro-dogmatismos* se aferran a sus postulados centrales y no están dispuestos a modificar

¹¹ La libertad de expresión, es decir lo que la gente no quiere oír. George Orwell.

mediante el aprendizaje los puntos en los que se sostiene la argumentación falsa pero consistente de las ideas que animan a cada uno de estos. La ecología social dominante que parece consistente con los argumentos señalados es aquella en la que la difusión de ideas falsas, tales como las *Fake News* o las *Deepfakes* encuentra un terreno apropiado no para su rechazo sino para su difusión, debiendo recordar que sin libertad de pensamiento no puede existir cosa alguna como la sabiduría, y tampoco algo como la libertad pública sin libertad de expresión (Gordon: 1721: 43). De alguna forma todo aquello que se dirija a la manipulación de la libertad de pensamiento ya sea *endógeno* o *exógeno* falseando los hechos afecta directamente a la libertad de expresión, ya que expresar libremente un pensamiento manipulado o deliberadamente erróneo no es a la postre sino efecto de una causa a la que los poderes públicos deben prestar especial atención desde el sistema educativo.

7. Tres ejemplos de arquitecturas de toma de decisiones colectivas diseñadas para neutralizar poderosos sesgos cognitivos en la toma de decisiones: el pre mortem; la formación de los pilotos de aviación civil y los filtros colaborativos

El conocimiento de los sesgos cognitivos que como seres humanos forman parte de nuestra naturaleza puede tener importantes consecuencias positivas si nos permite desarrollar instrumentos o herramientas procesales que puedan corregir esas deformaciones en nuestras formas individuales y colectivas de razonar o procesar la información distorsionada en la búsqueda de las respuestas más racionales a los problemas que tenemos que resolver, ello tiene especial relevancia en el campo de la educación en lo que Sunstein y Thaler (2009: 103) denominan la “*arquitectura de la decisión*”, es decir, en el desarrollo de arquitecturas de pensamiento correctas capaces de neutralizar o mitigar los peores efectos los sesgos en el procesamiento de la información.

¿Puede el optimismo que resulta del exceso de confianza ser superado mediante desarrollos procedimentales adecuadamente diseñados? Khaneman (2012: 346) señala que ha habido numerosos intentos de enseñar a suspender la confianza en intervalos en los que, con solo unos pocos informes de fuentes modestas, quedase reflejada la imprecisión de los juicios basados en el exceso de confianza. En otras situaciones, el exceso de confianza fue mitigado (pero no eliminado) animando a quienes juzgaban sobre algo a considerar hipótesis

alternativas. Sin embargo, el exceso de confianza es una consecuencia directa de aspectos de procesamiento de la información del Sistema 1, el cual puede ser domado, pero no derrotado. El principal obstáculo es que la confianza subjetiva viene determinada por la coherencia de la historia que uno ha construido, no por la calidad y la cantidad de la información en que se sustenta. Las organizaciones son más capaces que los individuos de domar el optimismo. La mejor manera de hacerlo es la que ideó Gary Klein, que generalmente defiende la decisión basada en la *intuición* frente al papel de los sesgos y es típicamente hostil a los algoritmos. Sobre la importancia de la intuición en la toma de decisiones y el *sesgo de reconocimiento* puede verse el trabajo de Gigerenzer (2008: 49). Klein denominó su propuesta como el *pre mortem*. El procedimiento es sencillo en su planteamiento: cuando una organización ha tomado una decisión importante, pensemos en un proyecto o una inversión, pero no la ha ejecutado formalmente, Klein propone que un grupo de individuos entendidos e informados sobre esa decisión se reúna en una breve sesión. La premisa de la sesión es una breve alocución que se basa en la siguiente idea: “Imaginemos que ha transcurrido un año” desde que se adoptó ese proyecto. Habíamos puesto en práctica el plan tal como ahora lo conocemos. El resultado fue un completo desastre. Tómense, por favor, 5 o 10 minutos para escribir una breve historia de tal desastre. La idea de Gary Klein del *pre mortem* suele suscitar un entusiasmo inmediato ya que es una fórmula para desarrollar una *autopsia* a un proyecto fracasado antes de que fracase en el mundo real. El *pre mortem* tiene dos ventajas principales: frena el *sesgo de pensamiento grupal* que afecta a tantos equipos cuando parece que se va a tomar una decisión y libera la imaginación de los individuos entendidos en un sentido muy necesario.

Cuando un equipo converge en una decisión —y especialmente si surge un líder con autoridad que la apoya—, las dudas manifestadas sobre el acierto de la acción planeada se desvanecen y eventualmente son tratadas como pruebas de escasa lealtad al equipo y a sus líderes o jefes en vez de una crítica impersonal al proyecto. La desaparición de las dudas contribuye al *exceso de confianza* en un grupo donde sólo los que apoyan la decisión tienen voz. La principal virtud del *pre mortem* es que legitima las dudas. Además, anima a quienes apoyan la decisión a buscar posibles riesgos que no se habían considerado antes. El *pre mortem* no es una solución extraordinaria y no proporciona una protección completa contra las sorpresas desagradables, pero enseña una “arquitectura de la decisión” destinada a

reducir el perjuicio que puede ocasionar un plan sometido a los sesgos y al optimismo acrítico. De forma análoga podemos considerar la estrategia o arquitectura de la decisión que se emplea en la mayor parte de las líneas aéreas en la formación de sus pilotos y en relación con la toma de decisiones de riesgo entre el comandante de la aeronave y el segundo comandante y que se desarrolló fundamentalmente a partir de la trágica experiencia del accidente aéreo de Tenerife del 27 de marzo de 1977 entre dos Boeing 747 con resultado de 584 víctimas mortales (Brafman: 2009: 19). Dos sesgos, *la aversión a la pérdida* y el *sesgo diagnóstico* fueron -al margen de otros aspectos que se dieron cita en aquél accidente- los que estuvieron presentes en la mente del comandante de la compañía aérea KLM Jacob Van Zanten. El *sesgo de aversión a la pérdida* es la tendencia a hacer cuanto sea posible para evitar pérdidas potenciales monetarias antes que obtener ganancias equivalentes, es decir, la pérdida tiene un peso emocional superior a la ganancia en nuestra mente. Las pérdidas (Tversky: 1992: 316) son valoradas psicológicamente entre 1,5 y 2,5 veces más intensamente que las ganancias. El *sesgo diagnóstico*, por su parte, constituye una forma de ceguera psicológica ante todas las pruebas que contradicen nuestra valoración inicial de una persona o situación, tras *la atribución de valor* derivada de una percepción inicial a esa persona o situación en el caso que nos ocupa, ya que la situación fue una percepción de seguridad dramáticamente errónea.

El comandante del Boeing 747 de la KLM, Jacob Van Zanten era uno de los pilotos más experimentados y consumados del mundo, no sólo tenía una habilidad singular para pilotar, sino que era conocido por el planteamiento metódico en todos los detalles de los vuelos que comandaba, de hecho, era el director del programa de seguridad de KLM. Los hechos, muy sintéticamente relatados fueron los siguientes. El vuelo de la KLM originariamente se dirigía desde Ámsterdam a las Palmas en las Islas Canarias, sin embargo el avión tuvo que variar su destino debido a que un artefacto explosivo había explotado en el aeropuerto de Las Palmas, teniendo que dirigirse al aeropuerto de Los Rodeos en Santa Cruz de Tenerife, esa circunstancia afectó al tráfico aéreo de ese día por lo que otros aviones, y entre ellos otro Boeing 747 de Pan Am también tuvo que aterrizar en un aeropuerto que no estaba diseñado para gestionar el tráfico de tantas aeronaves. El comandante Van Zanten no permitió que desembarcara el pasaje, era consciente de que en unas pocas horas desde el aterrizaje en el aeropuerto de Los Rodeos correspondía el turno de descanso de la

tripulación, tampoco sabía cuándo se podría despegar y si habría plazas de hotel para todo el pasaje en las inmediaciones del aeropuerto. La situación implicaría para la compañía aérea demoras y pérdidas económicas si se quedaban en ese aeropuerto, por ello ordenó llenar de combustible el avión en lo que consideraba que podría determinarse si se podría despegar en algún momento de aquel día.

El aeropuerto empezó a ser cubierto por la niebla y desde la torre de control no había buena visibilidad, la torre permitió al 747 de Pan Am que empezase a situarse en la parte superior de despegue de la pista, pero el Boeing de Pan Am no siguió correctamente -por la niebla y el desconocimiento de las pistas- las instrucciones que se le dieron. Van Zanten por su parte, en un momento dado adopta la decisión de despegar al recibir la noticia de que el aeropuerto de Las Palmas quedaba nuevamente abierto al tráfico; una vez que terminó de repostar el comandante tomó una decisión incomprensible para no verse forzado a pasar la noche en Los Rodeos. Aceleró los motores y el avión empezó a rodar por la pista. El copiloto -como señalaron las transcripciones de cabina- dijo a Van Zanten que esperara, que no tenían autorización de despegue. Ya lo sé, replicó Van Zanten y frenando el avión instó al copiloto a que la pidiese. El piloto comunicó con la torre y obtuvo la autorización de vuelo – es decir la aprobación del plan de vuelo- pero la torre no autorizó la orden de “despegue”. Van Zanten estaba resuelto a despegar y a plena potencia rodó por la pista a toda velocidad cubierta de niebla. Entre tanto el 747 de Pan Am estaba cruzando la pista mientras que el 747 de KLM avanzaba a velocidad de despegue, ya en vuelo la cola del 747 de Van Zanten chocó contra el fuselaje del Boeing de Pan Am, estallando el avión de KLM en el aire lleno de combustible. Van Zanten, toda la tripulación y los pasajeros del vuelo de la KLM murieron, así como muchos pasajeros del 747 de Pan Am, los fallecidos fueron en total 584 personas.

El desastre tuvo una influencia duradera en la industria aeronáutica, destacando en particular la importancia vital del uso de las expresiones estandarizada en las comunicaciones de radio. Los procedimientos de la cabina también se revisaron en profundidad, lo que contribuyó al establecimiento de la gestión de los recursos de la tripulación como parte fundamental de la capacitación de los pilotos de las aerolíneas. Van Zanten había sentido un estado de frustración que le llevó a ignorar precisamente lo que él

enseñaba como director del programa de seguridad que se había ensayado en multitud de ocasiones. La aversión a la pérdida, y el sesgo diagnóstico condujeron al comandante a adoptar una decisión irracional que produjo el accidente más grave de la historia aeronáutica.

Desde entonces a los pilotos se les entrena para cuestionar al comandante de la aeronave cuando detectan una desviación de los protocolos de seguridad. Lo hacen siguiendo un procedimiento de tres fases. En la primera fase se declaran los hechos, por ejemplo: “Nuestra velocidad de aproximación es muy baja”. Si esa declaración no es atendida, el siguiente paso es “cuestionar” empleando el nombre del comandante y añadir un cuantificador al hecho que se declara. Esta estrategia intenta sacar al comandante la visión de túnel que tal vez esté experimentando. Si los dos pasos anteriores no dan resultado se pasa a acciones de naturaleza más ejecutiva como comunicar la situación a los controladores de vuelo o de torre para que adopten estas las decisiones que el comandante no podrá ignorar. Como podemos observar se trata de cambiar una arquitectura de la decisión en la que el comandante era una autoridad incuestionada en todo momento, por una nueva estrategia en la que el comandante tiene que ajustar sus criterios a los de la tripulación cualificada, es decir, al segundo comandante que debe supervisar la conducta del primer comandante si esta pone en peligro la seguridad de la aeronave.

Los *filtros colaborativos* son herramientas que permiten obtener información relevante de bienes de experiencia¹², en los que para obtener esa experiencia es necesario haber disfrutado del bien previamente. Frente a publicidad optimista del prestador del servicio o fabricante del bien que puede disponer de la mejor información sobre el bien o servicio los filtros colaborativos constituyen un esfuerzo por resolver un problema de *arquitectura de las decisiones*. Si se sabe qué suelen preferir personas como nosotros, podremos animarnos a seleccionar productos o servicios que no conocemos a partir de esa información. A muchas personas el filtro de colaboración les facilita muchas elecciones difíciles, hoy estos

¹² Los bienes de experiencia son aquellos bienes cuyas características más importantes tales como: calidad, resistencia, durabilidad o adecuación no pueden ser determinados por el consumidor hasta que han sido adquiridos o consumidos. En este tipo de bienes se pueden producir incentivos a que se generen fraudes respecto a los elementos del bien como puede ser la calidad de este. Este problema se acentúa aún más en los denominados bienes de confianza, en los que el resultado del bien solo puede ser juzgado por un experto. La categoría de bienes de experiencia fue estudiada inicialmente por Philip Nelson, “Information and Consumer Behavior”, 78(2) *Journal of Political Economy* 311-329, 1970.

filtros están ampliamente difundidos, desde opiniones en Amazon de miles de usuarios sobre los más variados productos, pasando por redes de alojamientos como Booking, Trivago, etc. He incluso alcanzando los filtros de valoración colaborativa a la prestación de servicios profesionales: médicos; abogados; profesores etc., lo que ayuda a identificar la *ineficiencia* con beneficio para los usuarios.

La descripción temprana del efecto Halo se puede ver ya en la obra de Maquiavelo (1999:91) cuando señalase: *Pues los hombres, en general, juzgan más con los ojos que con las manos porque todos pueden ver, pero pocos tocar. Todos ven lo que parece ser, más pocos saben lo que eres; y estos pocos no se atreven a oponerse a la opinión de la mayoría, que se escuda detrás de la majestad del Estado.* Los sesgos cognitivos “Halo” o la también denominada *coherencia emocional exagerada* en el caso de los profesionales, pueden ser corregidos o mitigados en base a información derivada de la experiencia lo que puede suponer un indicador útil a la hora de adoptar una decisión como por ejemplo contratar un servicio. Este sesgo fue identificado por Edward L.Thorndike (1920: 28). Si nos gusta la política del presidente de nuestro país, por ejemplo, o de un determinado líder político de la oposición, es muy probable que nos guste su voz y su apariencia también. La tendencia a gustarnos o disgustarnos todo de una persona —incluyendo cosas que no hemos observado directa o indirectamente— de forma unificada, se conoce como efecto Halo. Consiste en un error asociado a la existencia de datos muy limitados sobre una persona y elevada incertidumbre en la emisión de un juicio sobre circunstancias o personas, que conlleva afirmaciones exageradas o irreales sobre habilidades, capacidades o atributos de una persona o de una cierta circunstancia. Es habitual en el ámbito de los juicios de tipo social que se realizan a partir de las opiniones de expertos reconocidos en un área concreta del conocimiento, cuando se refieren a asuntos fuera de su ámbito profesional, donde se “*presupone erróneamente*” que el conocimiento de la persona abarca cuestiones de la más diversa índole por causa de alguna suerte de capacidad especial que abarca aquellos aspectos en los que el experto no lo es pero que forman ese “*halo*” entorno a sus opiniones, valoraciones, juicios o capacidades. Solomon Asch (1946: 283) observó que el atractivo físico es la variable que más evoca el efecto halo. El atractivo físico, siendo habitualmente la primera información que tenemos de otra persona, da a las personas que juzgan sesgadamente la impresión del efecto una información medible sobre el mismo, y son

algunas características del atractivo físico las que mejor evocan este efecto. Son estos rasgos específicos los que nos hacen juzgar la personalidad de un individuo por la apariencia. Son muchos estos rasgos específicos, por ejemplo: la altura, la complexión, el color de los ojos, color del cabello, el peso, la voz, etc.

8. Conclusiones y discusión.

La idea fundamental que sostenemos en este trabajo y en base a lo señalado anteriormente es que *todos estos sesgos* deberían ser enseñados y formar parte de la educación de los alumnos de forma que se pueda controlar lo que es perfectamente previsible y de esa forma poder avanzar hacia sociedades mejor preparadas para identificar los *patrones incorrectos o distorsionados de nuestro razonamiento evolutivo*. No sabemos todavía cómo desarrollar algoritmos capaces de neutralizar todos los efectos de los sesgos, como vimos en el caso del sistema BERT porque los contextos son excesivamente variados, pero es claro que quizá una de las misiones importantes que podrá desarrollar la inteligencia artificial es proveer de instrumentos tanto para *identificarlos* como para *sugerir* vías para su neutralización que se ofrezcan a los individuos. Mientras eso llega una *educación avanzada* no puede desconocer estos sesgos o heurísticas *si queremos fomentar un pensamiento crítico elaborado sobre sólidas bases científicas*. Desafortunadamente nuestros sistemas educativos ignoran en los planes de estudio actuales una materia de la relevancia como la que hemos tratado de esbozar en este estudio, pero parece razonable que debería plantearse formar parte de estos por su relevancia social, toda vez que pueden ser abordados desde la perspectiva de la *protección de la libertad del pensamiento* en el sentido de que se garantiza esa libertad cuando identificamos formas de procesamiento distorsionado de la realidad y proponemos estrategias para su corrección favoreciendo un pensamiento no distorsionado o erróneo, lo que nos aproxima más a la libertad y nos aleja del condicionamiento y del error sistémico. Libertad entendida pues como correctora de errores o sesgos cognitivos.

Tampoco se debe perder de vista que en la sociedad de la información en la que los “*mercaderes de la atención*” que podemos identificar en los grandes operadores de plataformas: Google, Facebook, Amazon, Twitter etc., que comercian con información procedente de los datos personales extraídos de los ciudadanos y que explotan o pueden explotar los sesgos cognitivos y emocionales de los mismos, éstos adquieren un poder de

influir en la conducta de las personas como nunca nadie antes había podido soñar. En el caso de Twitter, por ejemplo, esta red social realiza análisis de sentimientos que se refiere al uso del procesamiento de lenguaje natural, análisis de textos, lingüística computacional y biometría para identificar, extraer, cuantificar y estudiar sistemáticamente estados afectivos e información subjetiva de los usuarios de la red, con la cobertura del control del discurso del odio pero del que es fácil extrapolar su traslación al mercado de la influencia política de los ciudadanos, debemos recordar que de las 200 categorías de datos que evalúa Google, se encuentran también los datos de los sentimientos o emocionales de sus usuarios. Sensibilizar a la sociedad con los sesgos en los que incurrimos y su identificación puede contribuir a que las regulaciones normativas de *lege ferenda* incorporen tales conocimientos específicos para, por ejemplo, restringir o limitar que los algoritmos puedan hacer uso de esa información emocional que es perfectamente capaz de convertirse en una poderosa herramienta de control ideológico susceptible de dar forma a las actitudes de los ciudadanos de manera individualizada o agregada.

Recordemos que, en opinión de Jefferson, las ideas no constituían una propiedad privada de nadie. En cuanto un individuo expone públicamente sus opiniones, éstas pertenecen a todo el mundo y todos se pueden aprovechar de ellas. El hecho de que las ideas se extiendan libremente de uno a otro por todo el mundo para instrucción moral y recíproca de los hombres parece un designio especial y benévolo de la naturaleza, que las ha hecho como el fuego extensible por todo el espacio sin perder en ningún punto su densidad y como el aire donde nos movemos, respiramos y emplazamos nuestro ser material, no susceptible de confinamiento ni de apropiación exclusiva. Este designio especial que determinara Jefferson: la libre extensión de las ideas por todo el mundo para instrucción moral; que constituiría la vertiente activa de la libertad de comunicación pública, se conjuga y adecúa necesariamente con la vertiente pasiva o el derecho a recibir ideas que positivamente configuraría el derecho de buscar, investigar y recibir informaciones y opiniones, que Fichte señalara magistralmente en su discurso (reivindicación de la libertad de pensamiento, 1793) en respuesta al Edicto de Censura promulgado, el 19 de diciembre de 1788, por el ministro conservador Wöllner, bajo el reinado de Federico Guillermo II de Prusia con las siguientes palabras: “*El derecho a recibir libremente todo lo que nos es útil forma parte constitutiva de nuestra personalidad; pertenece a nuestro destino servirnos libremente de todo aquello*

que está a nuestra disposición para nuestra formación espiritual y moral. Sin esta condición, la libertad y la moralidad serían un regalo inútil para nosotros. Una de las fuentes más ricas de nuestra instrucción es la comunicación entre un espíritu y otro. No podemos renunciar al derecho a abastecernos de esta fuente sin renunciar a nuestra espiritualidad, a nuestra libertad y personalidad [...]. La libre investigación de todo objeto posible de la reflexión, llevada en cualquier dirección posible y hasta el infinito, es, sin duda alguna, un derecho del hombre. Nadie, salvo él mismo, puede determinar su elección, su dirección y sus límites”. Hoy tras más de doscientos años de aquellas palabras, Jefferson y Fichte no podrían adivinar lo que representaría la aparición de Internet como medio sin igual de comunicación con vocación universal para la difusión de las ideas correctas, incorrectas o meramente falaces, tampoco podrían imaginar las intensas externalidades negativas que suponen los usos abusivos o ilícitos de la información y la capacidad de control social que bajo regulaciones normativas sumamente perfectibles puede llegar a explotar esta tecnología. Pero no debemos olvidar los elementos extraordinariamente positivos de Internet, tan sólo debemos procurar desarrollar normas significativamente más eficaces basadas en los nuevos paradigmas que venimos considerando que protejan de forma adecuada la intimidad, la protección de datos, así como la identificación y regulación de todas aquellas actividades que afecten a la dignidad de las personas en las redes de comunicaciones electrónicas. Balleisen (2017: 143), señala con acierto que cada vez que se traspasan fronteras tecnológicas —sea la invención del servicio postal, el teléfono, la radio, el ordenador o los títulos respaldados por hipotecas—, esos avances permiten a las personas aproximarse a las fronteras tanto de la tecnología como de la deshonestidad a través del fraude. Sólo más adelante, en cuanto se hayan establecido las capacidades, los efectos y las limitaciones de una tecnología, podemos determinar los medios tanto deseables como abusivos para utilizar esas nuevas herramientas. Por ejemplo, Balleisen revela que una de las primeras funciones del servicio de correos de los EE. UU., era vender productos *que no existían*. Se tardó un tiempo en descubrir el problema del fraude postal, que a la larga fue el prelude de un contundente conjunto de regulaciones que ahora garantizan la gran calidad, eficiencia y confianza en este importante servicio. Si pensamos en los avances tecnológicos desde esta perspectiva, llegamos a la conclusión de que debemos dar las

gracias a algunos estafadores creativos por ciertos progresos e innovaciones que hemos llevado a cabo.

No es una tarea imposible desde luego como vemos, pero lo que se exige es que la sociedad sea consciente de los límites que no se deben franquear y legisladores sensibles con de su deber y dispuestos a fijar esos límites en normas jurídicas desarrollando apropiadamente los derechos fundamentales que amparan las facultades de uso de las tecnologías por los ciudadanos, es una cuestión de tiempo y de prudencia. Es singularmente importante la admonición de Fichte cuando expresaba: *“Nadie, salvo él mismo, puede determinar su elección, su dirección y sus límites”*. Un problema que se cierne sobre todos nosotros acontece cuando la elección de los individuos es orientada por poderosos sistemas de procesamiento de datos emocionales capaces de guiar subrepticamente la conducta de forma silenciosa y sibilina y cuando la dirección y límites de la curiosidad de las personas, sus gustos o deseos son conducidos clandestinamente por poderes privados o públicos que le dirigen, que conforman sus opiniones, que tratan de gobernar nuestra vida emocional siendo capaces de moldear nuestra libertad y vida a los códigos de conducta diseñados en laboratorios de experimentación psicológica de gobiernos o multinacionales, eso ya ha sucedido con Cambridge Analítica en el campo electoral en los Estados Unidos como hemos estudiado en trabajos anteriores y con resultados positivos para quienes han explotado la orientación conductual de masas inermes de ciudadanos desprotegidos frente a tal amenaza. Es claro que se hace necesario la articulación dogmática de nuevos derechos, como el Derecho a no ser engañado que deben situarse en la esfera de los derechos universales fundamentales desde donde su irradiación puede ser incorporada a las Cartas de Derechos de los Estados. El Derecho a no ser engañado puede ser un útil instrumento para que no se empleen los sesgos cognitivos contra los titulares del Derecho, de él puede nacer un haz de facultades que reordene diversas manifestaciones que integrarían su ámbito normativo en la sociedad de la información global. Su expresión podría ser: *Todas las personas tienen el Derecho a no ser engañadas en la sociedad de la información, a que no se empleen o exploten contra sí mismas los sesgos cognitivos que nos afectan como individuos o como grupo, a que la dignidad inherente al ser humano exija que las autoridades públicas vigilen activamente y limiten con efectividad que los algoritmos*

públicos o privados no sean empleados para el dominio social, político o comercial de las personas en defensa de la libertad del pensamiento como ausencia de coacción.

9. Bibliografía

Asch, Solomon. E. "Forming impressions of personality", *Journal of Abnormal and Social Psychology*, 41, 258-290, 1946.

Ariely, Dan, "*Las trampas del deseo*", Ariel, Barcelona, 2019.

Ayan, Steve, "El inconsciente, el piloto automático de la mente", *Mente y Cerebro*, N° 99, 34-50, 2019.

Balleisen, Edward J, "*Fraud, and American history from Barnum to Madoff*", Princeton University Press, 2017, New Jersey, USA.

Brafman, Ori y Rom Brafman, "*El impulso irracional*", Espasa, Madrid, 2009.

Castellanos Claramunt, Jorge, "Participación ciudadana y posverdad: la amenaza de la posverdad participativa", *Persona y Derecho*, Vol 81, 349-384, 2019.

Coleman, Paul, "*La censura maquillada. Cómo las leyes contra el discurso del odio amenazan la libertad de expresión*", Dykinson, Madrid, 2018.

Damasio, Antonio, "*Y el cerebro creó al hombre ¿cómo pudo el cerebro generar emociones, sentimientos, ideas y el yo?*", Destino, Barcelona, 2010.

Da Vinci, Leonardo. John, R; Don Read, J (eds.). "Cuadernos organizados y prestados en inglés". Empire State Book Co. 1923.

D'Ancona Matthew, "*Posverdad. La nueva guerra contra la verdad y cómo combatirla*", Alianza Editorial, Madrid, 2019.

Deutsch, Karl W, "*Política y gobierno*", Fondo de cultura económica, Méjico, 1976.

Elster, Jon, comp. "*La democracia deliberativa*", Gedisa, Barcelona, 2000.

Ferrajoli, Luigi, "*Principia Iuris. Teoría del Derecho y de la Democracia*", Trotta, Madrid, 2011.

Frankfurt, Harry G., “*Sobre la charlatanería y sobre la verdad*”, Paidós, Barcelona, 2018.

Friston, Karl., “Friston y su teoría de la energía libre”, *Mente & Cerebro*, Investigación y Ciencia, Marzo/Abril, Nº 101, 2020.

Funder, David C., y Joachim I. Krueger, “Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition”. *Behavioral and Brain Sciences*, vol 27(3):313–76, 2004.

Gazzaniga, Michael S., “*El cerebro ético*”, Paidós, Barcelona, 2006.

Gigerenzer, Gerd, “*Decisiones instintivas*”, Ariel, Barcelona, 2007.

Goleman, Daniel, “*El punto ciego. Psicología del autoengaño*”, Plaza y Janés, Barcelona, 1997.

Gordon, Thomas, Carta 15. Sobre la libertad de expresión: que resulta inseparable de la libertad pública, en: Cueva Fernández, Ricardo, “*Cartas de Catón*”, BOE, Madrid, 2018, p. 43.

Haidt, Jonathan, “*La mente de los justos*”, Deusto, Barcelona, 2019.

Haidt, Jonathan y Greg Lukianoff, “*La transformación de la mente moderna*”, Deusto, Barcelona, 2019.

Hardin, Russell, “*Confianza y confiabilidad*”, Fondo de Cultura Económica, México, 2010.

Hayek, Friedrich A, “*Los fundamentos de la libertad*”, Unión Editorial, Madrid, 2008.

Helbing, Dirk y otros, “Democracia digital o control del comportamiento”, *Temas* (Investigación y Ciencia), 1 trimestre, nº 91, 2018.

Hog, Michael A., “Identidad y populismo”, *Investigación y Ciencia*, nº 518, noviembre 2019.

Jaspers, Karl, “*General Psychopathology*”, Practika, Moscow, 1997.

Kahneman, Daniel, “*Pensar rápido, pensar despacio*”, Debate, Barcelona, 2012.

Kadushin, Charles, “*Comprender las redes sociales. Teorías, conceptos y hallazgos*”, CIS, Madrid, 2013.

Kaushik, Basu, “*Una república fundada en creencias*”, Grano de sal, Méjico, 2018.

Klein Gary, puede verse en Harvard Business Review: <https://hbr.org/2007/09/performing-a-project-premortem> (última visualización 05/2020).

Keita Kurita y otros, “Measuring Bias in Contextualized Representations”, *1st ACL Workshop on Gender Bias for Natural Language Processing 2019*. Puede verse en: Cornell University <https://arxiv.org/abs/1906.07337v1> (visualizado 02/01/2020).

Lakoff, George, “*Cómo piensan progresistas y conservadores*”, Capitán Swing, Madrid, 2016.

Laski, Harold, “*Los peligros de la obediencia*”, Sequitur, Madrid, 2017.

Liebrecht, Christine, Letticia Hustinx y Margot van Mulken, “The Relative Power of Negativity: The Influence of Language Intensity on Perceived Strength”, *Journal of Language and Social Psychology*, vol 38(2) 170-193, 2019.

Luhmann, Niklas, “*Confianza*”, Anthropos, Barcelona, 2005.

López Rosetti, “*Emociones y sentimientos*”, Ariel, Barcelona, 2018.

Macknik, S.L, S. Martínez-Conde y Sandra Blakeslee, “*Los engaños de la mente. Cómo los trucos de magia desvelan el funcionamiento del cerebro*”, Destino, Barcelona, 2012.

Maquiavelo, Nicolás, “*El Príncipe*”, Aleph, Argentina, 1999.

Matute, Helena, “Ilusiones y sesgos cognitivos”, *Investigación y Ciencia*, noviembre, 2019.

Mazzoni Giuliana, “*Psicología del testimonio*”, Trotta, Madrid, 2019.

McIntyre, Lee, “*Posverdad*”, Cátedra, Madrid, 2008.

Mele, Alfred R, “*El autoengaño desenmascarado*”, Cátedra, Madrid, 2016.

Nelson Philip, “Information and Consumer Behavior”, 78(2) *Journal of Political Economy* 311-329, 1970.

- Nozick, Robert, “*La naturaleza de la racionalidad*”, Paidós. Barcelona, 1995.
- Pluviano, Sara, Sergio Della Sala, “El autoengaño de los antivacunas”, *Mente y Cerebro*, nº, 95, 2019.
- Pohl, Rüdiger F, “*Cognitive Illusions. A Handbook of Fallacies and Biases in Thinking, Judgment and Memory*”, Psychology Press, Taylor & Francis Group, New York, 2004.
- Popper, Karl, “*La Sociedad abierta y sus enemigos*”, Paidós, Barcelona, 1957.
- Quattrociocchi, Walter, “La era de la (des) información”, *Temas (Investigación y Ciencia)*, 1 trimestre, nº 91, 2018.
- Rozin Paul, y Edward B. Royzman, “Negativity Bias, Negativity Dominance, and Contagion”, *Personality and Social Psychology Review*, Vol. 5, nº 4, 296-320, 2001.
- Russell, Bertrand, “*Viaje a la revolución*”, Ariel, Madrid, 2017.
- Schacter, Daniel L., “*Los siete pecados de la memoria*”, Ariel, Barcelona, 2003.
- Searle, John R, “*Ver las cosas tal como son. Una teoría de la percepción*”, Cátedra, Madrid, 2018.
- Seth, Amilk, “La construcción cerebral de la realidad”, *Investigación y Ciencia*, 55-60, noviembre, 2019.
- Solms, Mark, en: Ayan, Steve, “El inconsciente, el piloto automático de la mente”, *Mente y Cerebro*, Nº 99, 34-50, 2019.
- Sunstein, Cass, R, “*Conformity. The Power of Social Influences*”, New York University Press, NY, 2019.
- Sunstein, Cass, R. Richard H. Thaler, “*Un pequeño empujón*”, Taurus, Barcelona, 2009.
- Sutherland, Stuart, “*Irracionalidad: El enemigo interior*”, Alianza Editorial, Madrid, 2015.
- Thaler, Richard H, “*La Psicología económica*”, Deusto, Barcelona, 2017.
- Thorndike, E. L. “A constant error on psychological rating” *Journal of Applied Psychology*, IV, 25-29, 1920.

Tversky, Amos; Kahneman, Daniel, “Advances in prospect theory: Cumulative representation of uncertainty”. *Journal of Risk and Uncertainty* 5: 297-323, 1992.

Zimbardo, Philip, “*El efecto Lucifer. El porqué de la maldad*”, Paidós, Barcelona, 2019.