

La inversión de un paradigma filosófico-jurídico de la justicia: de Rawls a la era digital.

Annalisa Verza

1. La Teoría de la Justicia: la última de las grandes construcciones teóricas

Creo que puede decirse que en el mundo occidental, en el ámbito de la filosofía jurídico-política-social, el paradigma teórico de la justicia dominante del último siglo ha sido el elaborado por John Rawls en su obra maestra *A Theory of Justice*, publicada en 1971¹, y luego revisada en su *Political Liberalism* de 1993². La teoría rawlsiana de la justicia ha representado quizás el esfuerzo más completo, elaborado y exitoso jamás realizado para integrar, en una única visión coherente, una teoría fundacional del contrato social que está en la base de nuestro sistema político y jurídico, con una visión normativa, racional e imparcialmente justificada, de los principios básicos que dicho sistema debe aplicar concretamente.

Como es sabido, a nivel sustantivo, esta compleja y rica construcción ha llegado a señalar como predominantes dos principios de justicia dirigidos a las instituciones básicas de una sociedad bien ordenada: aquel por el que debe garantizarse la máxima libertad para cada individuo, compatible con la de los demás, y aquel por el que cualquier diferencia en la distribución de los bienes debe considerarse aceptable sólo si es capaz de producir, al final, una mejora incluso en las condiciones de los más desfavorecidos (*maximin principle*).

Aún más representativo de este enfoque es el modo en que, en la tesis de Rawls, se definen estos principios: mediante la construcción de un dispositivo de elección (una nueva forma de contrato social) que, eliminando procedimentalmente (mediante el descenso, en la "*posición originaria*", de un "*velo de ignorancia*") para cada parte contratante toda posible tentación de parcialidad a favor de las concepciones del bien que le son personalmente vinculadas, obliga a cada uno a deliberar de manera puramente racional e imparcial (según un modelo, es decir, de "*justicia procedimental pura*").

Sólo el ejercicio independiente, por parte de cada miembro de la sociedad, de esa deliberación racional podría conducir, de hecho, a la definición de un modelo de justicia verdaderamente neutral, capaz de conciliar las diferencias entre las concepciones del bien elegidas por los

¹ John Rawls, *A Theory of Justice*, Cambridge (MA), Harvard University Press, 1971.

² John Rawls, *Political Liberalism*, New York, Columbia University Press, 1993.

individuos, todas las cuales deben ser valoradas precisamente por ser elegidas autónomamente (siempre que sean *razonables*, es decir, respetuosas a su vez con la autonomía de los demás).

Como se desprende de este brevísimo resumen, entre los elementos centrales de esta gran construcción teórica destacan, sin duda, al menos dos principios que, además, están interconectados:

(a) una visión del hombre científicamente entendido como un ser racional interesado en maximizar su propio beneficio, según la visión "clásica" del *homo oeconomicus*. Es decir, la capacidad humana de *racionalidad* es vista como una facultad que ciertamente puede ejercerse en condiciones que permitan desprenderse de la atracción ejercida por el interés propio, y es concebida y exaltada como el único medio que, corrigiendo nuestras tendencias a la parcialidad, puede conducir a la justicia;

b) la centralidad otorgada al valor de la *autonomía* de cada ciudadano, que, incorporado en el primer y más importante principio, viene a constituir, en el plano moral, político y por tanto también jurídico, el principal valor entre los dignos de protección.

La prioridad de estos pilares conceptuales no surgió ciertamente de la nada: la centralidad otorgada a la racionalidad humana de cada individuo, de hecho, reflejaba una matriz racionalista kantiana que en esas mismas décadas, desbordaba de la filosofía moral a la esfera de la psicología y más allá (el enfoque psicológico de la moral de Lawrence Kohlberg – racionalista y centrado, precisamente, sobre la preeminencia del valor de la justicia – marcó el conjunto de los últimos cuarenta años del siglo pasado), también encontró aplicación y desarrollo, en paralelo, en las teorizaciones contemporáneas de la elección racional, en sus diversas aplicaciones – desde la lógica, hasta la economía, pasando por las ciencias (piénsese, por ejemplo, en el famoso "dilema del prisionero", y en los trabajos de John Nash).

Del mismo modo, el valor (también "kantiano") de la autonomía individual, con su corolario de la consideración (primero neutra, y luego activamente protectora) hacia los múltiples matices de la identidad (sea o no objeto de elección), encontró más tarde su exaltación, en el ámbito filosófico-político, en la teorización del concepto de interseccionalidad (a nivel personal), y en las tesis de apoyo al multiculturalismo (a nivel colectivo).

No pretendemos aquí resumir en unas pocas líneas el enorme debate que siguió a la publicación de la obra de Rawls³; a los efectos de este razonamiento, basta con recordar que los esfuerzos

³ Para una reconstrucción del debate, me remito a Annalisa Verza, *La neutralità impossibile*, Milano, Giuffrè, 2000.

analíticos de los numerosos estudiosos que durante décadas se concentraron en la exégesis del paradigma de Rawls pronto identificaron, entre otras cosas, un importante defecto en su construcción. Era un defecto que estaba tan bien escondido e incrustado en su propia estructura como que era capaz de invalidar en última instancia, debido a su centralidad, la propia pretensión de neutralidad de los resultados del contrato social.

Como se ha señalado, de hecho⁴, el valor de la autonomía, aparentemente derivado como resultado final del ejercicio del proceso de elección racional en condiciones de imparcialidad, ya estaba implícitamente sugerido a las partes contratantes cuando el esquema rawlsiano les exigía que ejercieran una elección ya libre (y, por tanto, autónoma) entre los diversos principios que fundamentan las concepciones del bien.

Fue este elemento del "procedimiento" (el ejercicio de una elección libre), en definitiva, el que, al excluir implícitamente la participación en la deliberación de los partidarios de concepciones del bien incompatibles con la elección autónoma, condujo necesariamente al redescubrimiento, al final del procedimiento, de un ingrediente en realidad ya inserto en el circuito desde el principio: el valor pivotante de la autonomía. El valor de la autonomía, transmitido a través del marco, no era más que el presupuesto de un razonamiento racional que confirmaba sus propios puntos de partida. En otras palabras, si la elección de los contratantes en la deliberación rawlsiana bajo el "*velo de ignorancia*" podía efectivamente decirse que era racional, imparcial y neutral, no podía decirse lo mismo del *framing* de la misma, es decir, de su estructura, o de su arquitectura, pues ello ya ocultaba el valor que luego se elegiría por unanimidad.

Cincuenta años después, el paisaje teórico aparece literalmente al revés.

Por un lado, en efecto, la fuerza teórica de los conceptos de racionalidad y autonomía – piedras angulares dominantes de la concepción de Rawls – se ha visto completamente desbordada por los nuevos estudios impulsados por el enorme impacto de la "revolución digital"⁵, y la por la neurociencia y la psicología cognitiva, hasta el punto de que no sólo ha caído la confianza en tales principios, sino que nuevas técnicas de control social, basadas en la intersección entre tales descubrimientos y el conocimiento capilar de nuestros datos proporcionado por las redes digitales,

⁴ Véase, por ejemplo, Brian Barry, *How Not to Defend Liberals*, in *Liberalism and the Good*, a cura di R. B. Douglass, G. M. Mara, H. S. Richardson, New York: Routledge, 1990, p. 44; Stephen Macedo, *Liberal Virtues*, Oxford: Clarendon Press, p. 262.

⁵ Luciano Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano, Raffaello Cortina, 2014.

se están injertando en la misma convicción de la *inconsistencia* de la racionalidad humana, produciendo formas de control oblicuamente *heterónomas*.

Por otro lado, y en sentido contrario, lo que se había identificado como una de las principales críticas al marco de Rawls – es decir, la direccionalidad subliminal dada al proceso deliberativo por la incorporación del valor de la autonomía en el mismo *frame* de la toma de decisiones – ha pasado de ser una limitación a ser una verdadera técnica de regulación, apuntalando toda una serie de nuevas formas de control y dirección de la sociedad, en su mayoría extralegales y completamente alejadas de la discusión política y del escrutinio de alguna forma de contrato social. Precisamente "*la piedra desechada por los constructores*", en definitiva, se ha convertido rotundamente en "*piedra angular*".

2. La crisis de los conceptos de racionalidad y autonomía

Incluso antes del desarrollo actual de las ciencias cognitivas, la confianza en la racionalidad humana siempre ha ido acompañada de una prudente conciencia de sus lados oscuros. Ya H.L.A. Hart, en su *The Concept of Law*, había atribuido a la limitación (pero no a la ausencia) de la racionalidad humana un valor explicativo central en relación con el fenómeno jurídico. Pero incluso en tiempos más cercanos a la discusión que nos ocupa, recuerdo vivamente la atención que se prestó al tema de la *akrasia*, la incapacidad de actuar según los propios principios razonables, en los cursos de postgrado oxonianos celebrados a finales de los años noventa por un académico que, como Joseph Raz⁶, participaba como protagonista en el debate sobre la justicia desde el "frente" liberal, si bien mantenía una visión perfeccionista de la misma.

Pero esta primera piedra angular del constructo de Rawls, la de la fiabilidad de la *racionalidad* humana, ha sido objeto de estudios en el nuevo siglo que, por el contrario, y desde diversas perspectivas, se han dedicado a desentrañar y comprender en profundidad los mecanismos que tutelan nuestras tendencias *irracionales*. No sólo los estudios autorizados e influyentes de la psicología cognitiva, sino también las observaciones basadas en las nuevas técnicas digitalizadas de *neuroimagen*, y los experimentos a muy gran escala realizados en las redes sociales, han revelado la preponderancia, en la determinación de nuestro comportamiento, del peso de los automatismos reales, que sólo son corregidos por nuestra racionalidad en condiciones muy particulares.

⁶ Joseph Raz, *The Morality of Freedom*, Oxford: Oxford University Press, 1988.

En primer lugar, como ha demostrado el premio Nobel Daniel Kahneman⁷, nuestro comportamiento, pero también la afirmación de nuestras ideas, están constantemente dominados por una base instintiva, que nos predispone a elecciones muy rápidas, regidas por prejuicios que anticipan el juicio (los *biases* son muchos: desde el *bias* de confirmación, que nos inclina hacia lo que es coherente con ideas y elecciones previamente validadas, hasta el *bias* de inercia, pasando por el *bias* de influencia social, y muchos otros).

Como afirman Kahneman y Tversky, son casi siempre estas tendencias automáticas, ya arraigadas en nuestra constitución mental, y seleccionadas en el curso de la evolución humana en función de las ventajas evolutivas que han dado a nuestra especie, las que mueven, con su poder animal, nuestro pensamiento y nuestra actuación, y no la racionalidad, que, en su mayor parte, se limita a elaborar a posteriori formas lógicas y plausibles de justificación de conductas ya definidas. En definitiva, las condiciones para un ejercicio de pensamiento verdaderamente deliberativo y exploratorio, y no meramente confirmatorio, rara vez se dan, y requieren, como ha escrito Philip Tetlock⁸, la presencia simultánea de tres presupuestos

- que los responsables de la toma de decisiones, antes de formarse una opinión, ya saben que tendrán que justificarla ante los demás;
- que las opiniones de estos otros no se conocen;
- que los responsables de la toma de decisiones consideren que la audiencia está bien informada y se interesa por la exactitud de las decisiones tomadas.

Estas condiciones – de hecho, las mismas que la posición original de Rawls – que, sin embargo, sólo se dan en la realidad en muy raras ocasiones, a favor, en cambio, del uso de *heurísticos* de decisión rápida (atajos mentales automáticos).

Hoy en día, pues, el desarrollo de las técnicas de *neuroimagen* digital (como la IRMf y la TSE) que permiten interrogar directamente al cerebro, y mostrar comparativamente sus activaciones ante diferentes estímulos y en distintas condiciones, ha dado una confirmación concreta a estos estudios, mostrando también con claridad el modo en que factores como el estrés y la prisa, o el desencadenamiento de dinámicas de grupo (tribalistas y polarizantes), consiguen inhibir al

⁷ Central es el *Biases and Heuristics Program* de Kahnemann e Tversky.

⁸ J.S. Lerner e Philip Tetlock, *Bridging Individual, Interpersonal, and Institutional Approaches to Judgement and Decision Making: The Impact of Accountability on Cognitive Bias*, in S.L. Schneider e J. Shanteau, *Emerging Perspectives on Judgement and Decision Research*, New York: Cambridge University Press, 2003.

máximo el potencial de reflexión racional, acelerando y potenciando, en cambio, nuestras respuestas automáticas.

Pero la consecuencia más importante es, quizás, otra. Como sostenía Francis Bacon, el conocimiento es poder: una vez conocidas, estas tendencias son también, y sobre todo, *predecibles* y, por tanto, prácticamente explotables.

Pero si el comportamiento y el pensamiento irreflexivo, con los estudios de Kahnemann y Tversy, han salido así por primera vez del ámbito de la racionalidad, otros estudios, como los de Jonathan Haidt⁹ y Joshua Greene¹⁰, han mostrado cómo incluso nuestro propio *pensamiento moral*, con los principios y valores que lo constituyen, lejos de ser, *a' la Kohlberg*, una isla residual del ejercicio de la reflexión racional, en realidad reproduce patrones de comportamiento fijados evolutivamente en nuestra constitución biológica, y que se pueden rastrear a "emociones morales" irreflexivas, y no a juicios considerados. Así, principios y valores hasta ahora idealizados y sacralizados, como los de la protección de los vulnerables, de la equidad, de la lealtad al grupo, y sentimientos "activadores" como el de la empatía, el de la vergüenza, etc., se revelan también como conectados a pulsiones emocionales automáticas, con las mismas consecuencias prácticas que las anteriores.

Finalmente, otros estudios más recientes¹¹, retomando de hecho los argumentos más antiguos y conocidos de varios autores del realismo jurídico americano (piénsese en Jerome N. Frank y Karl N. Llewellyn), han destacado cómo incluso el último elemento que queda para sostener el estandarte de la racionalidad humana, es decir, el juicio deliberadamente racional, aunque ejercido conscientemente con un desprendimiento deliberado de heurísticos, emociones y automatismos, es profundamente sensible a la presencia incontrolable de elementos de contorno que, si bien irrelevantes para los fines del juicio, terminan por influir en él confiriéndole vaguedad, mutabilidad e imprevisibilidad.

Se trata del "ruido" (*noise*), constituido por elementos que pueden ser ambientales (piénsese en la influencia de la temperatura, o en cualquier otro posible elemento límite presente en el contexto de elección – el *frame*), idiosincrásicos (las tendencias de carácter de la persona que emite el juicio – por ejemplo, su nivel de severidad) o incluso ocasionales (como la buena predisposición tras una comida satisfactoria).

⁹ Jonathan Haidt, *Menti tribali*, Torino: Codice, 2013 (orig 2012).

¹⁰ Joshua Greene, *Moral Tribes*, London: Atlantic Books, 2013.

¹¹ Daniel Kahnemann, Olivier Sibony e Cass Sunstein, *Noise. A Flaw in Human Judgement*, London, William Collins, 2021.

Mientras que la clara direccionalidad de los prejuicios y de las emociones morales nos empuja hacia elecciones predecibles – y por tanto, como se ha señalado, manipulables –, el ruido, al hacer que cada juicio sea dirigible hacia una gama demasiado amplia e incierta de posibles resultados, en función de su composición, lo convierte en indefinido y, en consecuencia, no se presta a proporcionar elementos para guiarlo. Sin embargo, por otro lado, y precisamente por esa vaguedad de la que está revestido el juicio humano, la constatación del poder de influencia del ruido se presta a *justificar* el control externo heterónimo del comportamiento y a sustituir el juicio humano "defectuoso" por fórmulas más susceptibles de garantizar la justicia y la precisión¹². Tanto más cuanto que un juicio descrito como a merced de los *biases*, de los impulsos emocionales y de la influencia del ruido ya no puede decirse que sea verdaderamente "autónomo" en el sentido rawlsiano, y ya no gozará de la inviolabilidad que se le atribuye en ese paradigma.

El desmantelamiento del mito de la racionalidad humana, en definitiva, sólo puede haber desbordado necesariamente incluso el valor cardinal rawlsiano del máximo respeto a la *autonomía* de juicio y elección de cada persona, que está vinculado a él. Si las elecciones, los comportamientos adoptados e incluso los juicios resultantes de la reflexión son irracionales y están condicionados en gran medida por otra cosa, de hecho, también el presupuesto de la deferencia que se les otorga llega a perder su fundamento principal.

Entretanto, además, esta caída de la confianza y de la valorización de la autonomía había madurado también como resultado de otra vía científica paralela: la puesta en marcha por el conductismo de Burrhus F. Skinner, que fue a partir de los años 30, y durante décadas, un influyente profesor de psicología en Harvard y otras universidades.

Según los estudios de Skinner, corroborados además por los resultados de sus estudios experimentales con animales, el comportamiento humano consistiría esencialmente en esquemas de respuesta condicionados por estímulos externos. También según este paradigma, por tanto, la acción humana estaría totalmente determinada y, en consecuencia, también sería deliberadamente determinable. De ahí que Skinner comenzara su largo estudio de los mecanismos del "condicionamiento operante", las causas que producen las elecciones y las acciones, con el trasfondo de la asunción de una negación total y completa de la capacidad humana de autonomía. Sin embargo, si todos estamos condicionados de la misma manera, porque estamos programados para responder de forma determinista a los estímulos y refuerzos, entonces esta direccionalidad

¹² Esta es, de hecho, la recomendación de los autores de *Noise*.

puede hacer que no sólo los individuos, sino también sociedades enteras, masas, "enjambres", sean programables para seguir esquemas de comportamiento preestablecidos, más aún teniendo en cuenta la disponibilidad actual de redes digitales capaces de llegar a todos y conocer a todos. El legado de Skinner, y sus tesis sobre el aprendizaje programado y las "máquinas de aprender", ha desembocado, de hecho, en el conocimiento de las técnicas de programación algorítmica que hoy, cruzadas con los *big data*, estructuran nuestra relacionalidad y nuestras acciones online y, en cierta medida, offline. Desde este punto de vista, como ha afirmado más recientemente Alex Pentland, director del *Human Dynamics Lab* del *MIT Media Lab*, siguiendo la estela de Skinner, el ser humano se disuelve en la computabilidad predecible de lo colectivo y de los grandes números determinables algorítmicamente¹³.

El individuo autónomo, en definitiva, no parece ser más que un "percance estadístico"¹⁴.

3. El *framing*, de limitación crítica a recurso operativo: nuevas técnicas de regulación social

Mientras que, por un lado, los puntos fuertes del marco rawlsiano (racionalidad y autonomía) sufrían una reinterpretación paralela, una transformación inversa y paralela afectaba a la comprensión del sutil mecanismo del *framing*, es decir, de la capacidad directiva y reguladora de la (presupuesta y no discutida) arquitectura de la elección. El mecanismo en cuestión, antes censurado como un grave límite crítico capaz de socavar la resiliencia de una construcción como la de Rawls, pasó a ser elevado, ante la necesidad de remediar el déficit de racionalidad humana "externamente", al estatus de recurso operativo. Lo que antes era un problema, en otras palabras, se ha convertido ahora en una fuerza y una herramienta, no sólo aceptada, sino explícitamente invocada como un método no invasivo (de hecho, incluso inadvertido) de regulación social capaz de "empujar" suavemente hacia conductas y objetivos sin tropezar con deliberaciones humanas inevitablemente defectuosas, sino más bien invirtiendo a su favor el propio conocimiento y predicción de los automatismos que limitan nuestra racionalidad.

Estas limitaciones cognitivas se recontextualizan ahora como elementos que hay que consentir y explotar con vistas a objetivos heterónomamente establecidos como "buenos", en lugar de como vicios que hay que combatir (inútilmente).

¹³ Alex Pentland, *Social Physics. How Social Networks Can Make Us Smarter*, London: Penguin, 2014.

¹⁴ Shoshana Zuboff, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma: Luiss University Press, 2019, p. 457.

Se trata de una inversión del discurso, un movimiento de *aikido* teórico que se asemeja mucho a la "inversión" estratégica similar recomendada en el *Esprit des Lois* cuando, frente a la desilusión total por la capacidad de los poderosos de no abusar de su poder, Montesquieu nos instó a no luchar frontalmente contra esta limitación, sino a utilizarla en nuestro beneficio: *¡controlemos el poder* (por desgracia, abusivo) *a través del poder* (por desgracia, abusivo)! También aquí es la condena del vicio a corregir – en este caso la irracionalidad – la que sirve de justificación para el uso, con efectos e intenciones de ingeniería social, de técnicas destinadas a gestionarlo y pilotarlo explotando la propia irracionalidad¹⁵: ¡que controlemos la irracionalidad a través de la irracionalidad!

Técnicas, sin embargo, lejos de pasar por la discusión política previa o la contractualización: incluso en este aspecto, el neocontractualismo rawlsiano ha dado paso a su exacto opuesto.

3.1. *Tuning: del nudge al neuromarketing.*

Como escribe Shoshana Zuboff en su perturbadora obra *The Age of Surveillance Capitalism*, entre estas técnicas hay algunas que explotan la previsibilidad de nuestros impulsos de comportamiento aprovechando los mecanismos de *tuning*, es decir, de ajuste fino. Al igual que el *tuning*, en música, consiste en producir una afinación común de diferentes instrumentos "atrayéndolos" para que armonicen hacia un estándar acústico, también aquí se trata de técnicas destinadas (de forma muy diferente a la clásica regulación jurídica expresada en normas públicas con sanciones) a explotar los diversos aspectos de la arquitectura de la elección para sintonizar y unificar imperceptiblemente (e inadvertidamente) las conductas hacia los resultados deseados (considerados, de forma un tanto apodíctica, coherentes con los intereses del *nudgee* "tal y como los juzgan las personas mismas¹⁶"). La expresión más conocida de estas técnicas es la que se ha impuesto en el ámbito socio-político, el *nudge*, el "empujón suave", producido por pequeñas intervenciones de "ortopedia" conductual destinadas a orientar a los ciudadanos, de forma no coercitiva pero sabia, hacia un comportamiento racional aprovechando nuestra propia irracionalidad.

Los ejemplos propuestos por sus más famosos defensores, Richard Thaler y Cass Sunstein, son bien conocidos: un aula en la que, para estimular la atención, todas las sillas están orientadas hacia

¹⁵ Alex Pentland, *Social Physics*, cit.; Dominique Cardon, *Che cosa sognano gli algoritmi. Le nostre vite al tempo dei big data*. Milano: Mondadori, 2016; Shoshana Zuboff, *Il capitalismo della sorveglianza*, cit.

¹⁶ Cass Sunstein e Richard Thaler, *Nudge. Improving Decisions About Health, Wealth and Happiness*, Milano: Feltrinelli, 2014, p. 5 e p. 13.

un pupitre elevado; una cafetería que, por razones de salud, propone la ensalada de frutas a la altura de los ojos y antes que el pudín; un contrato de seguro médico que prevé su renovación automática para proteger a los individuos del olvido.

En todos estos casos, el mecanismo irracional del *bias* de inercia nos empujará a dirigir nuestra mirada hacia donde el contexto ya nos predispone a dirigirla, o a aceptar una estructura ya propuesta por *default* (la del contrato ya predispuesto para la renovación automática) adaptándonos a ella, en lugar de cuestionar la estructura. Otros *biases* que pueden utilizarse con este fin son, por ejemplo, el *bias* de disponibilidad, que lleva a considerar como comparativamente relevante o característico aquello que en el contexto dado se señala como especialmente relevante o característico, o el *bias* de influencia social, que empuja a adaptarse y conformarse con las elecciones que son percibidas como propias por la mayoría: también en este caso, el resorte activador reside en operar sobre la arquitectura de la elección, presentando un comportamiento como elección por *default*.

Otro *bias* especialmente activador es el de la excesiva consistencia, que implica la resistencia y preponderancia de las impresiones iniciales y, por tanto, la mayor dificultad para cambiar una opinión una vez formada. Es satisfaciendo esta forma de irracionalidad que se ha comprendido, por ejemplo, que en los países en los que es obligatorio el etiquetado de calorías en los envoltorios de los alimentos envasados, es más eficaz colocarlo en el lado izquierdo del envase, para que, al ser leído en primer lugar, produzca inmediatamente una impresión que condicionará el proceso mental posterior (por las mismas razones, en los países en los que se escribe hacia la izquierda, como Israel, la escritura se colocará en cambio en el lado derecho¹⁷).

El *tuning* realizado a través del instrumento del *nudge* es adoptado abiertamente por muchos gobiernos importantes, y ahora parece estar en camino hacia una institucionalización cada vez más embrizada y amplia¹⁸, a pesar de carecer de una politización básica: aunque los creadores afirman la coherencia del *nudging* con los objetivos de los *nudges*, de hecho, éstos no son en absoluto objeto de discusión y debate político-social previo. Como escribe Silvia Vida¹⁹, no hay nada que

¹⁷ Daniel Kahneman, Olivier Sibony, Cass R. Sunstein, *Noise. A Flaw in Human Judgement*, cit., p. 172.

¹⁸ Véase el informe de la OCDE de 2017 "*Behavioural Insights and Public Policy. Lessons from Around the World*", o la creación en 2010, en Gran Bretaña, del *Behavioural Insights Team* (BIT), una institución gubernamental dedicada a desarrollar políticas públicas basadas en el uso de las ciencias del comportamiento, y, en Estados Unidos, del *Social and Behavioral Sciences Team* (un subcomité del *National Science and Technology Council*) en 2014.

¹⁹ Matteo Galletti, Silvia Vida, *Libertà vigilata. Una critica al paternalismo libertario*. Firenze: IF Press, 2018, p. 364.

garantice que estos estándares ideales coincidan realmente con los objetivos reales que tienen los individuos sometidos a *nudging*. Pero sobre todo, y precisamente por ello, la decisión racional que implementan explotando la irracionalidad no está en absoluto vinculada a la *decisión autónoma* de los sujetos a los que se aplican: no puede serlo por definición.

Pero, además del ámbito político, el *tuning* también ha demostrado ser extremadamente productivo y eficaz en el ámbito económico, donde ha permitido desarrollar técnicas de *neuromarketing* cada vez más sofisticadas. También aquí fue la explotación de nuestros *biases* y tendencias irracionales, sondeados y medidos a través de las citadas técnicas de *neuroimagen* por empresas como EmSense, Neurosence, MindLab, International y Nielsen, la que revolucionó la forma de incitarnos a consumir, alejándola de la intuición de los anunciantes y basándola en fundamentos y técnicas científicas capaces de revelar computacionalmente lo que realmente consigue modular nuestras preferencias de consumo (deseos, emociones, afán de pertenencia).

Uno de los automatismos más explotados en este campo es el uso de los llamados "marcadores somáticos", elementos simbólicos destinados a asociar el producto con el recuerdo de emociones positivas o negativas anteriores para dirigirnos rápidamente en la elección de un producto. Un champú "no más lágrimas", por ejemplo, que evoca, precisamente, lágrimas, hará que uno corra para protegerse comprando el producto. Pero también encontramos técnicas capaces de inducir preferencias, como la publicidad subliminal, ya identificada por James Vicary en los años 50: estímulos inferiores a 30 milisegundos, y por tanto no interceptados por la conciencia, capaces de evocar inadvertidamente estados de ánimo, actitudes y deseos. O como el uso de símbolos eficaces para evocar un sentimiento de pertenencia, y atraernos hacia un producto capaz de provocar el placer inducido por este sentimiento (se ha visto que las marcas socialmente evocadoras, como Apple, estimulan las mismas zonas del cerebro vinculadas a la afiliación religiosa²⁰). O como el uso de los poderosos mecanismos de imitación, ligados al funcionamiento de las neuronas espejo, que nos impulsan a identificarnos con los modelos bellos asociados a determinados productos o prendas de vestir. También encontramos una hábil confianza en el poder de los mecanismos neurofisiológicos ligados a la consecución de objetivos (por ejemplo, terminar una colección de puntos) para liberar un neurotransmisor "dopante" como la dopamina, que luego nos empujará a querer repetir una y otra vez la conducta asociada a ella (como en el fenómeno de la compra

²⁰ Martin Lindstrom, *Neuromarketing. Attività cerebrali e comportamenti d'acquisto*, Santarcangelo di Romagna: Maggioli editore, 2013 (2008), pp. 111-131.

compulsiva, que da felicidad a corto plazo, pero luego desencadena la ansiedad por reproducir la misma sensación lo antes posible).

Pero este conocimiento requiere una gran cantidad de datos para ser cada vez más preciso. Por ello, el *neuromarketing* se apoya hoy en día en Internet. Y la sinergia de intereses dada por la intersección de ambos ha llevado, para obtener un beneficio, a investigar y codificar cada vez más las fuerzas que impulsan nuestro comportamiento.

3.2. Herding

El *tuning* actúa, como se ha visto, sobre el individuo, aprovechando sus prejuicios, sus mecanismos afectivos, su estrés, sus influencias inadvertidas. El *herding*, en cambio, se refiere a las masas: literalmente, "*herding*" significa "arrear". Uno de sus significados, según el Cambridge Dictionary online, es "to make people move somewhere as a group, often against their wishes or with difficulty". Es, en otras palabras, la estimulación de un comportamiento social convergente, producido por la alineación de los pensamientos o comportamientos de diferentes individuos en ausencia de una dirección central²¹.

Es una dinámica que reproduce, en el ámbito humano, las fascinantes geometrías que se hacen visibles cuando los bancos de peces, las bandadas de pájaros o las manadas de animales salvajes se mueven de forma misteriosamente coordinada, como una unidad, a pesar de la ausencia de una jerarquía y un liderazgo evidentes. En el ámbito humano, puede observarse en los movimientos y la dinámica de la multitud, una poderosa fuerza natural irreductible a la mente de los individuos que la componen, analizada con reveladora agudeza hace más de 100 años por un brillante testigo de aquellas explosiones de *herding* salvaje que se produjeron en el caos de las décadas posteriores a la Revolución Francesa, Gustave Le Bon. Fue él quien identificó claramente el "contagio social" como el resorte esencial del fenómeno: según Le Bon, la imitación, en su mayoría inconsciente²², se produce como efecto del contagio, y es fundamental para entender los fenómenos sociales. El ser humano, de hecho, es un imitador por naturaleza y es de la necesidad de imitar de donde surgen fenómenos direccionales comunes como la moda y otras formas de influencia colectiva, ya sean

²¹ Cfr. Ramsey R. Rifaat, Nick Chater, Chris Frith, *Herding in Humans*, in *Trends in Cognitive Sciences*, 13/10, 2009, pp. 420-428.

²² Gustave Le Bon, *Psicologia delle folle. Un'analisi del comportamento delle masse*, Milano: TEA, 2004 (orig. 1895), p. 174.

opiniones, ideas, manifestaciones literarias, costumbres: las multitudes se guían por modelos, no por argumentos.

Como escribió, las principales características del individuo en una multitud son la anulación de la personalidad consciente, el predominio de la personalidad inconsciente, la orientación determinada por la sugestión y el contagio de sentimientos e ideas en una dirección, la tendencia a transformar inmediatamente las ideas sugeridas en actos. El individuo ya no es él mismo, sino un autómeta, incapaz de guiarse por su propia voluntad²³.

El autor captó claramente el poder de la pulsión instintiva y "dopante" que gobierna al ser humano cuando se encuentra envuelto en estas formaciones colectivas: la sugestión, que por contagio se transmite a todos los cerebros indicando la dirección a tomar, tiende a convertirse en acción, y la multitud no admite ningún obstáculo entre un deseo y su realización²⁴.

Inmersos y disueltos en la multitud, en la masa, en la dinámica tribal, somos llevados, como observó Le Bon, a dejar de lado nuestra individualidad y a someterla completamente a la voluntad colectiva.

El contagio tampoco requiere la presencia simultánea de los infectados, como lo demuestra el hecho de que la explosión revolucionaria de 1848 arraigue en toda Europa.

En comparación con los tiempos del polifacético pensador francés, hoy en día se ha desentrañado y sacado a la luz lo que era la misteriosa "alma" de la multitud, junto con los mecanismos de "contagio social" que la supervisan, gracias, de nuevo, a las enormes cantidades de datos puestos a disposición por la omnipresencia computacional de la dimensión digital de nuestras vidas.

Si bien el mencionado mecanismo de las neuronas espejo, descubierto por Giacomo Rizzolatti²⁵, representa ciertamente, a nivel neuronal, una de las claves para explicar el fenómeno, hoy en día es sobre todo el uso de modelos matemáticos y algorítmicos lo que ha abierto nuevos horizontes para comprender los fenómenos de contagio social. La ciencia de las redes²⁶ también se está ocupando de ello, partiendo inicialmente de "modelos" abstractos capaces de representar los fenómenos colectivos, a partir de metáforas físicas, en las cuales los individuos se representan

²³ Ibidem, p. 55.

²⁴ Ibidem, p. 61.

²⁵ Giacomo Rizzolatti e Corrado Sinigaglia, *So quel che fai. Il cervello che agisce e i neuroni specchio*, Milano: Raffaello Cortina, 2006.

²⁶ Albert Laszló Barabasi, *Link: la scienza delle reti*, Torino: Einaudi, 2004.

como átomos, sus vínculos como moléculas, la orientación con respecto a un campo magnético como metáfora de la emergencia de un consenso político.

Pero si hablamos de contagio – y la metáfora remite naturalmente al orden de la biología- el nivel de complejidad que hay que entender y representar transportándolo al plano social se ve magnificado por el hecho de que, mientras que en el plano biológico la exposición a una, dos o más fuentes de contagio no produce diferencias en el resultado, en el plano social el efecto final es sensible tanto al aumento de las fuentes de contagio como a las cualidades del sujeto infectante (según sea o no una persona conocida) al que se está expuesto. Por tanto, la calidad de los datos sobre los que se calculan las simulaciones es relevante.

Precisamente por ello, esta perspectiva está traspasando la frontera más allá de los primeros "modelos de juguete", avanzando hacia sistemas predictivos reales²⁷ gracias a las enormes cantidades de datos que por fin están disponibles, recogidos en Internet a través del análisis de las redes sociales, de los motores de búsqueda, de los centros de compra online, de los datos transmitidos por las tecnologías wearables, del IOT, etc., relativos a millones de individuos "reales", todos ellos con preferencias, redes sociales e inclinaciones que ahora se conocen. Todos estos datos hacen posible simulaciones de fenómenos de contagio, resultado de las interacciones diarias en la red social, relacionadas con la difusión de ideas, conocimientos, creencias políticas, modas, etc, en la que los elementos "contagiosos" son las ideas, o incluso las emociones, y los resultados son agregaciones instintivas y tribalizaciones ocasionales en función de los estímulos (aunque sean efímeros y "sin alma"²⁸, como diría Byong-Chul Han), operativas y energéticas, pero en las que el razonamiento autónomo se ve frenado por la fuerza de arrastre del "coloide humano" así ligado. En efecto, como señala Jonathan Haidt, la triangulación "pertenecer-creer-hacer" que supervisa la dinámica del grupo dificulta el ejercicio, en ese contexto, del razonamiento racional²⁹.

3.3. *Condicionamiento y programación del comportamiento mediante algoritmos*

Como ya se ha dicho, el conocimiento es poder: poseer las claves de esta dinámica, es decir, comprender las fuerzas que determinan los fenómenos de contagio, es la base para poder producirlos y gobernarlos también – una notable fuente de poder regulador, tanto más eficaz en la

²⁷ Alessandro Vespignani, *L'algoritmo e l'oracolo. Come la scienza predice il futuro e ci aiuta a cambiarlo*. Milano: Il saggiatore, 2019, p. 50.

²⁸ Byung-Chul Han, *Nello sciame. Visioni del digitale*, Roma: nottetempo, 2015 (2013).

²⁹ Jonathan Haidt, *Menti tribali*, cit., p. 319.

medida en que es aplicable a individuos colocados en una condición, precisamente, de no racionalidad y no autonomía.

La última dimensión, la del condicionamiento proactivo operado gracias a los algoritmos, reúne la estrategia activadora del *tuning* con el conocimiento de la dinámica del *herding*, y lo hace, como hemos visto, desde una perspectiva verdaderamente a gran escala. Y es internet, el mundo digital ahora totalmente fusionado con nuestras vidas offline, la dimensión que hace posible tal condicionamiento.

Shoshana Zuboff, en este sentido, describe dos importantes experimentos de contagio social realizados por Facebook, a través de sus investigadores, sobre sus usuarios.

El primero de ellos, realizado con 61 millones de usuarios como conejillos de indias, se describe en un ensayo publicado en *Nature* en 2012³⁰. El objetivo era comprobar la eficacia de la influencia social en la movilización política: el contexto real y el pretexto del estudio fueron las elecciones de mitad de mandato de 2010 para el Congreso de Estados Unidos. Tras definir un grupo de control para las comparaciones necesarias, cuyo *newsfeed* no fue manipulado, se envió al *newsfeed* del conjunto de usuarios estudiados un mensaje en el que se les instaba a votar, acompañado de un enlace informativo, un botón que decía "he votado" y un contador que, idealmente, indicaba el número de usuarios que ya habían votado, incluidos algunos de los amigos del usuario (hasta seis) cuyas fotos se mostraban.

Los resultados mostraron, como era de esperar, que la sugerencia (*tuning* proactivo) de la consistencia de una colectividad, mostrada como afín y dirigida a un determinado tipo de comportamiento (*herding*), creando motivación y presión para ser como los amigos, aumentó perceptiblemente la alineación conductual del grupo estudiado, en el que la participación en la votación fue significativamente mayor que en el grupo de control³¹.

Sin embargo, no solo se puede influir en el comportamiento, sino también en los estados de ánimo, es decir, en los estados afectivos, como se observó en otro experimento sobre las pruebas experimentales de contagio emocional masivo a través de las redes sociales, realizado también con

³⁰ Robert M. Bond et al., *A 61-Million-Person Experiment in Social Influence and Political Mobilization*, in *Nature* 489, n. 7415, 2012, pp. 295-298.

³¹ El mecanismo del *herding*, por el que la atracción gravitatoria de un cuerpo social orientado hacia una idea o comportamiento lo lleva a crecer progresivamente a medida que reúne, por imitación y miedo de la exclusión, a otros individuos, motivados aún más por la existencia misma del grupo que por la causa, se reproduce también en los últimos tiempos en la práctica, a veces abusiva, de la recogida de firmas entre colegas universitarios.

la participación de investigadores de Facebook y publicado después en los *Proceedings of The National Academy of Sciences* en 2014³².

El objetivo del estudio era comprobar si una exposición, aunque sólo fuera subliminal, a un contenido emocional sugerido (*tuning*) podría haber influido consecuentemente en los posts, ramificados en sus comunicaciones con amigos, de las personas observadas (difundiendo así, se podría decir, un *herding* afectivo). De nuevo, el estudio confirmó la hipótesis, identificando en el cambio de carácter afectivo-emocional de las comunicaciones así condicionadas una dinámica de "influencia social subyacente" que demostró que los estados emocionales también pueden transferirse en una red social por contagio, llevando a las personas a compartir las mismas emociones sin ser conscientes de ello. Como escribe Shoshana Zuboff, la difusión de mensajes emocionales subliminales positivos o negativos se combinó con el estado motivacional activado por las comparaciones sociales. Esto, lo sabemos, es el resultado de la empatía humana, es decir, la capacidad de compartir estados emocionales y asumir los puntos de vista de otras personas, la facultad que hace que la risa y el bostezo sean contagiosos y que lleva a reproducir inconscientemente las expresiones faciales y el lenguaje corporal de otra persona³³.

Pero incluso la difusión del juego de realidad aumentada *Pokémon Go*, cuya moda estalló en 2016, constituyó en sí misma un experimento que demostró cómo era posible, a través de estímulos como la gamificación (que Skinner leería como "condicionamiento operante"), llevar también en la realidad offline a un enorme número de personas hasta donde el juego decidiera por ellos, en busca de puros estímulos virtuales, en una mezcla de *tuning*, *herding* y condicionamiento con un poder evidente, y unas implicaciones económicas inmediatas (*Niantic lab*, la empresa creadora del juego, había llegado a acuerdos con establecimientos comerciales a los que, por una cuota, desviaba a los jugadores: De este modo, los lugares patrocinados recibieron 500 millones de visitantes³⁴).

Finalmente, ni siquiera el ámbito de la política ha renunciado a aprovechar esta facultad de heterodirección y subversión total del respeto a la autonomía individual, con evidentes

³² Adam D.I. Kramer, Jamie E. Guillory, Jeffrey T. Hancock, *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*, in *Proceedings of The National Academy of Sciences* 111, n. 24, 2014, pp. 8788-8790.

³³ Shoshana Zuboff, *Il capitalismo della sorveglianza*, cit., cap. 10.

³⁴ https://techcrunch.com/2017/05/31/pokemon-go-sponsorship-price/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2x1LmNvbS8&guce_referrer_sig=AQAAAGqejSqfx3sp80Nj15k2uF9lpcIsDgkdvOK3Fe6aa6l0RcPvVaNvEx12xAeUDqSF-dEKWkeZpknKV0_n_uYb7RzWC2f0NICj9Ya1gt_C0NY14TR87EqGMN8Cgq3dwrPUYm2Lz0rSwmm9MIJqgOW3E7Af9kO7BITypSAPfPqfOKL

repercusiones no sólo en las esperanzas de neutralidad y justicia que encendió el gran paradigma de Rawls, sino en la propia resiliencia de la democracia. En las elecciones de 2008, Obama ya había recopilado datos sobre cientos de millones de votantes; en 2012, la campaña hizo uso de métricas, perfiles y bases de datos avanzadas basadas en informaciones de los votantes extrapoladas de las redes sociales³⁵, para entender cómo hablar a los votantes y a qué emociones apelar, implementando un *micro-targeting* conductual basado no en intereses (racionales), sino en los miedos, sospechas y enfados (emocionales) de los ciudadanos³⁶.

Una técnica, ésta, que ha marcado la pauta, y que ha sido puntualmente reproducida en contextos electorales posteriores (incluido el referéndum británico sobre el *Brexit*), suscitando escándalos como el de *Cambridge Analytica*, cuyo tardío renacimiento constituye en sí mismo, sin embargo, la demostración más llamativa del poder de la tendencia que tanto ha asombrado. De hecho, hace ya muchos años que se emprendió el largo camino que ha conducido al progresivo desmantelamiento de los supuestos de la racionalidad humana, debilitando progresivamente los valores colocados como garantía de nuestra autonomía, y allanando evidentemente el camino para el florecimiento de los estudios sobre las respuestas automáticas que inadvertidamente nos vemos inducidos a dar a los estímulos de heterodirección, y para la afirmación de las técnicas de condicionamiento heteronómico que encuentran uso en la economía y la política³⁷.

Y el hecho de que las implicaciones manipuladoras de todo esto sólo hayan salido a la luz recientemente, en un despertar súbito, escandalizado y atónito, es desgraciadamente una confirmación de su tremenda capacidad obturante, y de la vulnerabilidad que, en caída libre desde Rawls hasta nuestros días, amenaza cada vez más nuestra autonomía y nuestra capacidad de representarnos como subjetividades libres, partes reales del contrato social que nos engloba.

³⁵ Marco Cacciotto, *Il nuovo marketing politico*, Bologna: Il Mulino 2019; Giovanni Diamanti, *I professionisti della comunicazione politica, dalle origini al digitale*, in Claudio Riva, *Social media e politica. Esperienze, analisi e scenari della nuova comunicazione politica*, Torino: UTET Università, 2021, pp. 27-43, a p. 31.

³⁶ Isabella Borrelli, *Cosa influenza le nostre idee politiche online?* in Claudio Riva, cit., pp. 129-146, a p. 138

³⁷ Alessandro Vespignani, *L'algoritmo e l'oracolo*, cit., pp. 169-170.