

<http://doi.org/10.21555/top.v680.2374>

## Towards a Historically Bounded Rationality: Herbert A. Simon's Critique of the Neoclassical Notion of Rational Agent

### Hacia una racionalidad históricamente acotada: la crítica de Herbert A. Simon a la noción neoclásica de "agente racional"

Josafat I. Hernández Cervantes

División de Estudios Multidisciplinarios del Centro de  
Investigación y Docencia Económica  
México

[josafat.hernandez@cide.edu](mailto:josafat.hernandez@cide.edu)

<https://orcid.org/0000-0002-8077-5900>

Recibido: 27 - 04 - 2022.

Aceptado: 30 - 06 - 2022.

Publicado en línea: 06 - 12 - 2023.

Cómo citar este artículo: Hernández Cervantes, J. I. (2023). Hacia una racionalidad históricamente acotada: la crítica de Herbert A. Simon a la noción neoclásica de "agente racional". *Tópicos, Revista de Filosofía*, 68, 35–68. DOI: <https://doi.org/10.21555/top.v680.2374>.



This work is licensed under a Creative Commons Attribution  
-NonCommercial-ShareAlike 4.0 International License.

### Abstract

This article presents Herbert A. Simon's critique of the two basic assumptions of the neoclassical notion of agent: perfect rationality and self-interest. It shows how Simon criticized both assumptions by recognizing that the cognitive capacities of agents are limited, showing that there are specific problems that do not have optimal solutions and that there are other motivations irreducible to self-interest, such as loyalty and group identification, which are fundamental for understanding the behavior of organizations. It is shown how Simon pointed towards a notion of rationality that is historically bounded by the interaction of the agent with its environment, which presents a complex and dynamic vision of human behavior. However, the article also shows the difficulties Simon had in achieving the full realization of this project by assuming an assumption of individualistic cognition that prevented him from giving a more substantive role to the environment in structuring agents' behaviors. It ends by showing how the notion of habit can give a more substantive role to the environment in order to develop a historically bounded conception of rationality.

*Keywords:* bounded rationality; satisficing; heuristics; organizations; cognitive individualism; habits.

### Resumen

En el artículo se expone la crítica de Herbert A. Simon a los dos supuestos básicos de la noción neoclásica de "agente": la racionalidad perfecta y el interés propio. Se muestra cómo Simon criticó ambos supuestos al reconocer que las capacidades cognitivas de los agentes son limitadas, mostrando que existen problemas específicos que no tienen soluciones óptimas y que hay otras motivaciones irreducibles al interés propio, como la lealtad y la identificación grupal, que son fundamentales para entender el comportamiento de las organizaciones. Se muestra cómo Simon apuntó hacia una noción de "racionalidad" históricamente acotada por la interacción del agente con su entorno, lo cual presenta una visión compleja y dinámica del comportamiento humano. Sin embargo, el artículo también muestra las dificultades que tuvo Simon para lograr la realización

plena de este proyecto al asumir un supuesto de cognición individualista que le impidió dar un papel más sustantivo al entorno en la estructuración del comportamiento de los agentes. Se termina mostrando cómo la noción de “hábito” puede dar un papel más sustantivo al entorno para desarrollar así una concepción históricamente acotada de la racionalidad.

*Palabras clave:* racionalidad acotada; satisfaccionismo; heurísticas; organizaciones; individualismo cognitivo; hábitos.

## 1. Introducción

Herbert A. Simon (1916-2001) acuñó la expresión de “racionalidad acotada” (*bounded rationality*) desde mediados del siglo XX y con ello se le ha reconocido como fundador de la llamada vieja economía del comportamiento (*Old Behavioral Economics*) (Heidl, 2016; Heukelom, 2014; Angner y Loewenstein, 2012). Con su trabajo, Simon (1955) buscó realizar una drástica revisión del *homo economicus* que lo llevó a plantearse una sustitución completa de esa noción en la economía. Sin embargo, es controversial explicar exactamente qué fue lo que Simon entendió por “racionalidad acotada” y, en particular, su relación con la noción neoclásica de “agente racional”.

En la literatura de economía del comportamiento y de psicología cognitiva hay muchas interpretaciones, a menudo opuestas, sobre cómo entender el carácter acotado de la racionalidad. Algunos lo entienden como decisiones no óptimas o *second best* (Laibson y List, 2015; Sargent, 1993). Otros autores (Kahneman, 2011; Thaler, 2015) entienden la racionalidad acotada como decisiones rápidas basadas en razonamiento heurístico que pueden ser falibles y que pueden ser tendientes al error, o a la resolución de problemas de una forma suficientemente buena (*good enough*) (Gigerenzer, 2008; Berg y Gigerenzer, 2010). En estas interpretaciones, sin embargo, no se destaca un elemento que Simon enfatizó en sus últimos trabajos: la importancia de la relación agente-entorno dentro de una visión histórica (Simon, 1990 y 1998).

En este contexto, la tesis principal de este escrito es que la noción de “racionalidad acotada” de Simon se puede entender como *históricamente acotada*. La razón es la siguiente: luego de criticar y rechazar la noción neoclásica de “agente racional”, la cual expondremos en la sección

2, Simon fue elaborando conceptos que lo llevaron a subrayar cada vez más la importancia de la interacción entre el agente y el entorno social dentro una visión evolucionista. Primero, como veremos en las secciones 3 y 4, Simon criticó con su modelo satisfaccionista y su trabajo en inteligencia artificial uno de los supuestos estructurales de la noción neoclásica de “agente”: la racionalidad perfecta optimizadora. Luego, como veremos en la sección 5, criticó otro supuesto estructural de la noción neoclásica de “racionalidad”: el interés propio ilimitado. Las nociones de “satisfaccionismo” y “razonamiento heurístico”, los problemas no formalizables y su teoría de las organizaciones llevaron a una noción de “entorno” que, como veremos en la sección 6, Simon (1990 y 1998) siguió desarrollando, incorporando más variables que permiten esbozar una noción de “racionalidad históricamente acotada”. Esta noción permite mostrar cómo los comportamientos de los agentes pueden variar en diferentes épocas históricas y en diferentes grupos sociales. Sin embargo, como veremos en la sección 7, Simon compartió con la economía neoclásica un individualismo metodológico porque la unidad básica de análisis sigue siendo el individuo, solo que en este caso aparece como una variante un poco diferente que llamaremos “supuesto de cognición individualista”. Este supuesto consiste en asumir que el razonamiento heurístico y otros procesos cognitivos relevantes para la toma de decisiones de los agentes ocurren en el interior de la cabeza de los individuos (Zerubavel y Smith, 2010; Martínez, 2016). Este supuesto de cognición individualista, como veremos, le impidió a Simon tener una noción más fuerte de “interacción agente-entorno” que le permitiera entender mejor cómo se moldea y transforma la racionalidad de los agentes a lo largo del tiempo. En ese sentido, sus avances hacia una noción de “racionalidad históricamente acotada” fueron limitados, pero, como veremos en la sección 8, estos avances pueden desarrollarse más si se abandona el supuesto de cognición individualista y se avanza a una noción más social de la cognición retomando una noción relacional, como la de “hábito” del filósofo pragmatista norteamericano John Dewey. La sección 9 presenta comentarios finales.

## **2. La noción neoclásica de “agente racional”**

La noción neoclásica de “agente” tiene sus orígenes en la revolución marginalista del último cuarto del siglo XIX, donde se integró al análisis económico el uso intensivo del cálculo diferencial e integral, así como la filosofía moral utilitarista (Morgan, 2012; Ross, 2014). Con los

trabajos pioneros de William Stanley Jevons se inició la generalización en la economía del uso de funciones y curvas de utilidad. Los economistas neoclásicos postularon estas funciones para cada individuo atribuyéndoles, además, una motivación única para tomar decisiones: la maximización de la utilidad. Esto para que el agente logre la felicidad, la cual se entiende como un estado mental donde solo hay placer y ausencia de dolor (Mill, 2010). Sin embargo, con Jevons se vuelven equivalentes los términos “placer” y “felicidad con ganancia”, así como “dolor” y “sufrimiento con costos” (Morgan, 2012; Ross, 2014). El agente racional debe buscar la felicidad, entendiendo por esto la obtención de la máxima cantidad de beneficios económicos, y reducir al máximo los costos o pérdidas económicas.

Las funciones y curvas de utilidad de Jevons se axiomatizaron a mediados del siglo XX por medio del modelo de utilidad esperada. Este modelo fue formulado por primera vez por los teóricos de juegos John von Neumann y Oskar Morgenstern (1953). La ecuación básica de ese modelo es la siguiente:

$$UE = U(x_1, x_2, \dots, x_n) * P(x_1, x_2, \dots, x_n)$$

Ahí se define una función de utilidad a partir de los parámetros  $x$ , que constituyen las diferentes opciones que el agente puede elegir, y se multiplica por la probabilidad de la ocurrencia de dichas opciones. Esta multiplicación nos dará como resultado la utilidad esperada. La idea central del modelo es que el agente racional, por ser maximizador de utilidad, tomará la decisión que le genere mayor utilidad esperada. Es decir, la opción que tenga mayores beneficios, menores pérdidas, y más probablemente vaya a ocurrir.

El modelo de utilidad esperada representa la toma de decisiones de agentes maximizadores de utilidad situados en contextos de riesgo e incertidumbre, donde cada opción tiene un grado de probabilidad de ocurrencia. En situaciones de riesgo, donde la probabilidad es medible debido a la regularidad de los eventos, se calculan probabilidades usando frecuencias (probabilidad objetiva). Cuando se trata de situaciones de incertidumbre, donde las regularidades son poco frecuentes, se usan métodos bayesianos (probabilidad subjetiva que mide grados de creencia), donde la probabilidad de ocurrencia de un evento se va actualizando conforme la información disponible.

Los axiomas del modelo de utilidad esperada, a su vez, son los siguientes: completitud, transitividad, independencia y continuidad.

El primero dice que el agente tiene una lista completa y perfectamente ordenada de preferencias que le permite decidir entre dos o más opciones. La transitividad supone que si el agente prefiere a A sobre B y a B sobre C, entonces, si es racional, debe preferir A sobre C. Esto para subrayar el carácter coherente de la racionalidad. El axioma de la independencia postula que la formación de cada lotería (opciones con diferentes grados de utilidad y de probabilidad) son independientes de loterías anteriores y de la influencia de otros agentes. El axioma de la continuidad asume que entre las diferentes preferencias debe haber algún grado de continuidad, lo cual permite elaborar funciones de utilidad continuas, no discretas. Se dice que si el agente satisface todos estos axiomas entonces es racional (Morgenstern y Von Neumann, 1953; Ross, 2014).

En el modelo de elección racional se diferencia entre *actos, estados en el mundo* y *resultados* para poder modelar al agente racional y las situaciones de elección que este enfrenta. Cuando hay múltiples parámetros en las funciones de utilidad, se optimiza por medio del cálculo de derivadas parciales.

Este modelo, no obstante, involucra una serie de supuestos falsos que Milton Friedman y Leonard Savage (1952) defendieron a nivel metodológico por medio del uso de *supuestos "como si"*. Si bien los defensores de este modelo pueden reconocer que las personas de carne y hueso no usan los axiomas descritos arriba para razonar ni realizan cálculos probabilísticos complejos usando el teorema de Bayes, se puede suponer que estos actúan *como si* lo hicieran (Friedman y Savage, 1952). Esto es así porque en este modelo no interesa representar realísticamente a los agentes, sino facilitar la manipulación matemática para poder predecir sus comportamientos. Por ejemplo, si queremos predecir los tiros de un jugador experto de billar, se puede suponer que este es perfectamente racional, lo cual permite suponer que el agente hace sus tiros *como si* realizara cálculos matemáticos precisos. Ello nos permite usar geometría analítica y mecánica clásica para predecir el tiro que hará este jugador experto (Friedman y Savage, 1952). Para Friedman y Savage, el punto clave de un modelo no es si nos da o no una imagen más o menos realista, sino que sus predicciones sean correctas. El supuesto de la racionalidad perfecta es, por tanto, un asunto metodológico que facilita la elaboración de modelos predictivos. Y lo mismo pasa con el otro supuesto básico del modelo: el interés propio.

Si bien el interés propio fue planteado por primera vez en el siglo XVIII por Bernard Mandeville (1997) como un supuesto básico de la naturaleza humana, opuesto al altruismo, ya desde mediados del siglo XX no se le atribuye una realidad ontológica. Es decir, los economistas no dirían, con Mandeville, que los agentes solo actúan con amor propio (autoconservación), vanidad, ni que las acciones desinteresadas y altruistas no existen o que estas son hipócritas. Pero sí dirían que, si bien los seres humanos tienen múltiples motivaciones, se puede asumir que estos actúan *como si fueran egoístas* para poder construir modelos matemáticos (Becker, 1992; Kirchgässner, 2014; Segoviano Contreras y Morales Sánchez, 2021). El cálculo egoísta costo-beneficio está motivado por lo que le conviene de manera individual al agente, pensando solo en su propio interés individual y no en el interés de los demás.<sup>1</sup> Es decir, se trata de un reduccionismo a nivel metodológico, donde se asume un egoísmo metodológico muy parecido al de Maquiavelo, quien decía que en el diseño institucional podemos asumir que todos los agentes son bribones para de esta forma diseñar sistemas de premios y castigos que permitan reducir comportamientos oportunistas o corruptos (Bowles, 2014). Esto no significa que para Maquiavelo todas las personas sean, en efecto, oportunistas o bribonas, pero el supuesto permite prever hechos para así blindar las instituciones del gobierno de posibles oficiales que pueden aprovechar cualquier oportunidad para hacer actos corruptos.

El modelo de utilidad esperada tuvo una importante expansión en otras ciencias sociales durante la década de los ochenta y sobre todo durante los noventa, realizando aplicaciones a problemas que no son tradicionalmente económicos (Becker, 1992). Destaca el análisis de actos criminales, la corrupción, la discriminación racial y la elección de pareja (Becker, 1992). Algunos de los propios seguidores de Gary Becker han llamado a este proyecto de universalización del modelo de utilidad esperada el “imperialismo de la economía” (Lazear, 2000, p. 104) para enfatizar la idea de que el modelo de utilidad esperada está “colonizando” otras áreas de las ciencias sociales con métodos matemáticos rigurosos, y que, a la larga, puede presentar un lenguaje común y unificante para todas las ciencias sociales.

---

<sup>1</sup> Esta posición instrumentalista supuestamente agnóstica es ambigua porque, si bien los economistas ortodoxos negarían decir explícitamente, con Mandeville, que todos los seres humanos son egoístas por naturaleza, mucha de su retórica en favor de los libres mercados parece suponer que sí piensan eso.

Justo contra este tipo de modelos de elección racional es que Simon discutió tanto a nivel metodológico como teórico para abrir paso a su propuesta alternativa: la racionalidad acotada. Como veremos, la manera en que planteó la relación entre la economía y las otras ciencias sociales no es imperialista o expansionista (como sería en el caso del proyecto neoclásico de Becker y Lazear), sino que es, como lo plantea Uskali Mäki (2008, p. 27), como un humilde aprendiz: en vez de colonizar otras ciencias sociales, se aprende de ellas en un diálogo interdisciplinario para reelaborar una nueva noción de agente racional.

En resumen, podemos decir que el agente racional en su sentido neoclásico, i.e. el *homo economicus*, “es, en efecto, a la vez racional e interesado. Dicho de otra manera, sus elecciones reflejan la búsqueda racional de su interés” (Elster, 2011, p. 7). Así, podemos decir que los dos grandes supuestos del *homo economicus* son la racionalidad perfecta y el interés propio; contra esos dos supuestos, como veremos en las secciones 3, 4 y 5, Simon debate a lo largo de sus obras.

### 3. La crítica de Simon a la racionalidad perfecta y su modelo satisfaccionista

Herbert Simon, en sus primeros trabajos (1947 y 1955), se interesó en estudiar la toma de decisiones de consejos administrativos tanto de empresas como de gobierno. Sin embargo, vio que en esos consejos se tomaban decisiones que distaban de las predicciones hechas por el modelo neoclásico de utilidad esperada. Los tomadores de decisiones no buscaban maximizar su utilidad ni minimizar su desutilidad. Había otras motivaciones en la toma de decisiones que el modelo neoclásico tradicional no estaba captando. Para entender la toma de decisiones de los agentes reales, Simon requirió de elaborar un nuevo modelo. Y para ello partió de una serie de críticas al modelo neoclásico, donde dice que la economía neoclásica:

[...] postulates an “economic man”, who in the course of being “economic” is also “rational”. This man is assumed to have *knowledge of the relevant aspects of his environment which, if not absolutely complete, is at least impressively clear and voluminous*. He is assumed also to have a *well-organized and stable system of preferences*, and a *skill in computation that enables him to calculate*, for the alternative courses of action that are available to



him, which of these will permit him to reach *the highest attainable point on his preference scale* (Simon, 1955, p. 99; cursivas mías).

De lo anterior podemos señalar que la noción neoclásica de “agente racional”, según Simon, supone:

- Conocimiento casi completo de los aspectos relevantes del entorno (información casi perfecta), para cumplir así con el axioma de completitud.
- Un sistema de preferencias completo y estable, lo cual supone que se cumple con los axiomas de completitud y transitividad.
- Capacidad de cómputo, que le permite al agente racional “calcular” las diferentes opciones que tiene a su disposición para elegir la opción óptima que maximiza su utilidad, lo cual justificaría tanto el uso de teoremas bayesianos de probabilidad subjetiva como el del cálculo diferencial para calcular derivadas parciales.

Los tres puntos anteriores suponen que los agentes tienen *capacidades cognitivas ilimitadas*. Esto es: la memoria, el procesamiento de información y la capacidad de aprendizaje se consideran ilimitados, lo cual permite al agente tener una lista completa de preferencias perfectamente ordenadas, así como realizar todos los cálculos matemáticos y probabilísticos necesarios para optimizar sus funciones de utilidad.

Se atribuye capacidad de memoria ilimitada al agente porque, si se supone que este tiene un sistema de preferencias completo, ordenado y estable (como plantea el modelo neoclásico de utilidad esperada), se supone que el agente recuerda perfectamente sus preferencias tanto a corto como a largo plazo. Así, no viola el axioma de transitividad y mantiene su sistema de preferencias completo y ordenado. Se asume que la capacidad de procesamiento de información es ilimitada porque, si el agente es maximizador de utilidad además de optimizador, se supone que procesa toda la información sin cometer errores (o, como dicen los teóricos de las expectativas racionales, sin “desperdiciar información”; cfr. Muth, 1961). Finalmente, se asume que la capacidad de aprendizaje es

ilimitada porque el supuesto de que el agente es perfectamente racional lleva a comprometerse con la idea de que los agentes no cometen errores sistemáticos, sino solo errores aleatorios, más aún cuando se asumen métodos bayesianos para el cálculo de probabilidades (Muth, 1961; Friedman y Savage, 1952).

Los elementos arriba señalados llevaron a Simon (1947 y 1955) a decir que el modelo de elección racional era más normativo que descriptivo, pues este modelo habla sobre cómo los agentes racionales *deberían* tomar decisiones. No describe cómo los agentes, de hecho, toman sus decisiones. Entonces, la tarea que asume Simon es la siguiente:

[...] I shall assume that the concept of “economic man” (and, I might add, of his brother “administrative man”) is in need of *fairly drastic revision* [...] [:] the task is to *replace* the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man (Simon, 1955, p. 99; cursivas mías).

El pasaje anterior nos permite decir que, para Simon, el *homo economicus* requiere de una drástica revisión. La tarea consiste en reemplazar ese concepto por otra noción de racionalidad que sea más realista. Para ello, es necesario asumir supuestos con base psicológica, supuestos que estén validados por la evidencia empírica, no supuestos *a priori* (Simon, 1955). Con esta idea en mente, Simon inició su propuesta alternativa de racionalidad acotada, donde lo que hay de fondo es un proyecto de *naturalizar* la noción de “agente racional” (Heidl, 2016; Hortal, 2017; Hernández Cervantes, 2019).

Simon (1955 y 1956) primero consideró que ningún agente real, de carne y hueso, puede cumplir con los requisitos cognitivos que los axiomas del modelo de elección racional demandan. Por tanto, más bien hay que partir de reconocer que las capacidades cognitivas de los agentes son limitadas. El tiempo para decidir también es limitado, lo cual presiona a los agentes a tomar decisiones rápidas en contextos en que no se tiene toda la información a disposición y, además, las opciones son cambiantes.

Los agentes no pueden realizar cálculos precisos que los lleven a tomar decisiones óptimas si consideramos que tienen capacidades cognitivas limitadas, restricciones informacionales y de tiempo para decidir, así

como opciones cambiantes. Sus decisiones, por tanto, son rápidas y se apoyan en la intuición (Simon, 1955). Con estos elementos en mente, Simon (1955 y 1956) elaboró una propuesta alternativa para explicar la toma de decisiones de agentes reales: el modelo satisfaccionista. Este modelo asume que el agente busca satisfacer ciertos niveles de aspiración. Los agentes tienen preferencias *parcialmente ordenadas* que les permiten distinguir entre opciones que sí satisfagan sus niveles de aspiración y opciones que no. Esto representa un relajamiento del axioma de la completitud porque las preferencias no están totalmente ordenadas, sino parcialmente ordenadas. Mientras más altos sean los niveles de aspiración, más difícil será encontrar opciones que satisfagan dicho nivel; mientras más bajo sea, más fácil será encontrar alguna opción que lo satisfaga. No hay una única opción que sea la óptima, sino que más bien puede haber múltiples soluciones suficientemente buenas (*good enough*). Esto significa que, para Simon, el agente racional en realidad es satisfaccionista: no es maximizador de utilidad y tampoco busca tomar decisiones óptimas.

Es relevante remarcar este punto para enfatizar la idea de que el proyecto de Simon, a diferencia del de Thomas Sargent (1993) o Laibson y List (2015), no busca retener la idea de “optimización” como núcleo normativo, al cual habría que añadirle parámetros que nos expliquen las decisiones subóptimas de los agentes. Tampoco se trata de decisiones “poco racionales” (*misbehaviors*) —en el sentido de Richard Thaler (2015)— porque el agente, si bien no toma las mejores decisiones posibles (óptimas), sí toma decisiones suficientemente buenas que le permiten resolver problemas en un tiempo limitado. Simon, a diferencia de Thaler (2015) y de Kahneman y Tversky (1979), no enfatiza los comportamientos desviantes observados de los agentes porque no retiene el modelo de utilidad esperada como si fuera una lógica de la toma de decisiones.

La manera en que los agentes toman sus decisiones depende de su capacidad cognitiva de identificar pistas o señales (*cues*) en los entornos que pueden desencadenar respuestas rápidas que ayuden a descartar opciones que no son satisfactorias e identificar las opciones que sí son satisfactorias (Simon, 1955). Esto sin necesidad de procesar toda la información disponible. Un ejemplo para ilustrar este modelo es el de un trabajador que quiere vender su casa lo más pronto posible porque ha sido transferido. ¿Cómo decide esta persona si acepta o no una oferta hecha por alguno de los compradores? Simon (1955) señala que el vendedor de una casa puede fijar un precio —digamos, 15 000 dólares—,

el cual considera suficientemente bueno. Cualquier comprador que le ofrezca una cantidad menor será rechazado de manera inmediata. En cambio, cualquier cantidad que esté por encima de los 15 000 dólares será aceptada por el vendedor, más aún si tiene tiempo limitado para decidir, lo cual le impedirá comparar todas las opciones que hay disponibles en el mercado para tomar la decisión óptima, que en la economía neoclásica se asume como única.

En este caso, el agente no necesariamente elige la opción que optimiza su función de utilidad porque no compara las ofertas que todos los compradores de casas están dispuestos a hacer de manera simultánea. Este vendedor tiene que tomar decisiones en condiciones en que la información es incompleta, cambiante y oculta. Pero se contenta con tomar una decisión suficientemente buena que satisfaga su nivel de aspiración sin necesidad de tener grandes capacidades de cálculo: cualquier precio que esté por encima de los 15 000 dólares. Y lo mismo aplica cuando las personas van a los mercados a comprar sus bienes y servicios: una persona rara vez toma decisiones comparando todas las opciones que realmente hay en los mercados; más bien toma decisiones rápidas a partir de, por ejemplo, identificar señales en las mercancías que satisfacen sus niveles de aspiración. Una persona que va al mercado compra plátanos que se ven suficientemente buenos, no compra los mejores plátanos disponibles en el mercado comparando todas las opciones que ofrecen todos los vendedores.

Simon demostró, por tanto, que *no necesitamos retener alguna noción de optimización para dar cuenta de la toma de decisiones de los agentes*. Para Simon, el criterio principal para tomar decisiones es la resolución de problemas satisfaciendo niveles de aspiración. Esta idea, como veremos en el siguiente apartado, se profundizó en los trabajos pioneros de Simon en inteligencia artificial.

#### **4. Inteligencia artificial y las heurísticas orientadas a resolver problemas**

La resolución de problemas motivó a Simon a incursionar en el terreno de la inteligencia artificial en el contexto de la naciente revolución tecnológica de las computadoras de mediados del siglo XX. En esa época, Simon estudió fenómenos cognitivos por medio de simulaciones computacionales modelando la mente humana *como si* fuera una computadora (Simon y Newell, 1958 y 1962). Esta metáfora

computacionalista postula que la mente humana, igual que las computadoras, recibe insumos (*inputs*) y genera resultados (*outputs*) por medio del reconocimiento y manipulación de símbolos que representan objetos y estados en el mundo. Según Simon y Newell (1958 y 1962), es posible simular de manera computacional aspectos cognitivos como el aprendizaje, el reconocimiento de símbolos, la elaboración de composiciones musicales y las jugadas de ajedrez. Esto llevó a estos autores a decir que las máquinas literalmente “piensan” y que, al igual que los seres humanos, las computadoras pueden ser inteligentes porque resuelven problemas.

Los problemas por resolver pueden ser de dos tipos: los “problemas bien estructurados” (*well-structured problems*) y los “problemas mal estructurados” (*ill-structured problems*) (Simon y Newell, 1958 y 1962). Los primeros son problemas concretos que se pueden formular de manera explícita y de manera cuantitativa, de tal manera que se les puede dar una formalización matemática precisa. Luego se aplican algunas técnicas matemáticas, como la programación lineal, para resolver el problema mediante algoritmos o por medio de otras técnicas de optimización matemática. La solución en estos casos suele ser única, bien definida y estar garantizada si se sigue una secuencia bien ordenada de pasos. Sin embargo, este tipo de problemas no son los más comunes ni los más relevantes en la economía y en la sociedad. Más bien, los agentes se enfrentan, de manera cotidiana, a problemas que no pueden formularse de manera precisa en términos matemáticos porque *no siguen dinámicas lineales* y, dada la incertidumbre, presentan información incompleta, oculta y cambiante. Ese tipo de problemas son “los problemas mal estructurados” y, según Simon, son los más comunes y relevantes para la administración de organizaciones. Estos problemas involucran variables que son simbólicas o verbales, no solo variables numéricas.

Un ejemplo de este tipo de problemas es elegir cómo redactar un informe de política pública: lo importante no solo son los números o datos brutos, sino también la manera en que se presentan los datos, más aún cuando se quiere generar reacciones de cierto tipo en la percepción de los agentes. Para solucionar este tipo de problemas, según Simon, no es posible emplear técnicas de optimización matemática ni algoritmos, pues la solución de estos problemas no es única ni está garantizada. Se requiere, más bien, usar heurísticas de razonamiento (Simon y Newell, 1958 y 1962).

Para Simon es importante la distinción entre heurísticas y algoritmos. Recordemos que un algoritmo es un conjunto de instrucciones que pueden programarse en una computadora que permiten resolver mecánicamente un problema. Involucra una secuencia bien definida de pasos que garantizan la resolución de un problema (Martínez, 2003; Newell, 1983). En cambio, una heurística no garantiza los resultados y, cuando falla, lo hace con un sesgo (Martínez, 2003; Newell, 1983). Los algoritmos son infalibles; en cambio, las heurísticas son falibles.

Aquí un ejemplo puede ayudar a entender lo que Simon (1969) llama “razonamiento heurístico”: el comportamiento de las hormigas en la playa. A primera vista, se puede ver que las hormigas tienen comportamientos muy complejos, pues sus recorridos en la playa parecen poco geométricos y caóticos. Sin embargo, la complejidad de este comportamiento no está en la hormiga, sino en la *interacción* de la hormiga con su entorno irregular. Las hormigas pueden modelarse como organismos simples que siguen reglas heurísticas simples de búsqueda de comida (Simon, 1969). Parece que se mueven de forma zigzagueante porque buscan comida a un lado de las piedras, y como las piedras están distribuidas de manera caótica, las hormigas simplemente van siguiendo las piedras una a una. Para modelar este tipo de comportamientos, al igual que en el modelo satisfaccionista, lo único relevante es atribuir al agente una habilidad cognitiva: reconocer señales en el entorno que permitan al agente discriminar entre opciones binarias (hay o no un satisfactor).

Aquí no voy a polemizar con la manera computacionalista de Simon de entender la cognición, la cual ha sido criticada recientemente por diferentes enfoques contemporáneos de ciencias cognitivas (Chemero, 2011; Fonseca Patrón, 2019; Martínez, 2016). Sin embargo, me parece importante apuntar que Simon ya estaba abriendo la “caja negra de la cognición” (Arthur, 2015; Berg y Gigerenzer, 2010) con sus estudios, mostrando la relevancia de tomar en cuenta aspectos cognitivos para dar cuenta de la toma de decisiones de los agentes (tanto humanos como, en este caso, hormigas), lo cual lo distancia de los economistas neoclásicos.

Los trabajos de inteligencia artificial fueron llevando a Simon a interesarse cada vez más por el *proceso* de la toma de decisión y no tanto por el resultado, es decir, por responder a la pregunta sobre cómo los agentes toman sus decisiones, no así por estudiar el resultado de sus decisiones. Este tema llevó a Simón a establecer la distinción entre la “racionalidad substantiva” y “racionalidad procedimental”, donde:

The former is concerned only with finding *what action maximizes utility in the given situation*, hence is concerned with *analyzing the situation but not the decision maker* [...]. Procedural rationality is concerned with *how the decision maker generates alternatives of action and compares them. It necessarily rests on a theory of human cognition* (Simon, 1997, p. 18; cursivas mías).

Como podemos ver, Simon consideró que la racionalidad substantiva se centra solo en encontrar la acción que maximiza la utilidad en una situación específica. Por lo tanto, lo que interesa es estudiar los sistemas de incentivos que hay en una situación, no al tomador de decisiones. La racionalidad procedimental, en cambio, se interesa por estudiar cómo el tomador de decisiones genera alternativas de acción y las compara entre sí. Para dar cuenta de la racionalidad procedimental sí se necesita una teoría de la cognición humana, que Simon ya estaba desarrollando a partir de retomar elementos psicológicos que analizan la percepción, la intuición y la capacidad de los agentes para resolver problemas en las ciencias de lo artificial. Estos elementos que desidealizan la noción de “racionalidad perfecta” (Heidl, 2016), a su vez, fueron complementados por los trabajos de Simon sobre organizaciones, lo que lo llevó a cuestionar y superar el otro supuesto estructural básico de la economía neoclásica: el interés propio.

## 5. La teoría de las organizaciones de Simon y el interés propio acotado

Una de las críticas heterodoxas más comunes a la economía neoclásica es que carecen de una noción de “organización” (Estrada, 2007; Simon, 1994). Esto es así porque, en la economía neoclásica, entidades colectivas como las empresas, los gobiernos y las familias se representan por medio de funciones de utilidad. Estas entidades colectivas se asumen en la economía neoclásica: actúan *como si* fuesen individuos que, mediante una racionalidad instrumental, toman decisiones orientadas a la maximización de utilidad. Sin embargo, como bien dice Simon (1994), es falso asumir que las empresas siempre tienen comportamientos maximizadores de utilidad y que pueden, además, hacer cálculos precisos que los lleven a tomar los cursos de acción apropiados para optimizar. Por tanto, dice Simon, es necesario desarrollar una nueva teoría del mercado. Y para ello es fundamental

estudiar el comportamiento de las organizaciones porque desde ellas se estructuran las dinámicas de las sociedades (Estrada, 2007).

Para Simon, quien a su vez se consideraba un seguidor del empirista lógico Rudolf Carnap (Hortal, 2017), es fundamental que los supuestos teóricos de la economía tengan apoyo empírico. Es decir, Simon (1994) rompe con un apriorismo que consideraría suficiente el teorizar desde teoremas y supuestos tautológicos cuya verdad es independiente de la experiencia, o desde el cual se asumen supuestos “como si”. Simon más bien siguió una estrategia que lo llevó a naturalizar su noción de “agente” (y de “mercado”) tomando en serio la evidencia empírica. Así, la tarea que él asumió es la construcción de una noción de “organización” en la que agentes con racionalidad acotada toman decisiones.

Una empresa es una organización, y para entenderla se requiere de analizar su composición y cómo sus diferentes partes interactúan entre sí. En este caso, hablamos de individuos de carne y hueso que colaboran para lograr ciertos fines colectivos. Por esa razón, Simon (1947, 1979 y 1997) abordó las organizaciones como entidades colectivas complejas, compuestas por individuos que tienen jerarquías entre sí y división del trabajo, así como artefactos a partir de los cuales emergen rutinas organizacionales. Las rutinas organizacionales son patrones de comportamiento recurrentes que generan ciertas respuestas para problemas a los cuales se enfrentan las organizaciones (Estrada, 2007). Estos problemas organizacionales pueden resolverse de dos formas: por medio de decisiones programables y por decisiones no programables (Estrada, 2007). Las primeras abordan problemas bien estructurados, como vimos en el apartado anterior, que tienen un comportamiento lineal y con variables bien definidas. Según Simon, se pueden diseñar estrategias algorítmicas y de optimización para resolver estos problemas, los cuales se relacionan con rutinas organizacionales bien definidas y repetitivas, que, en cierto modo, ya garantizan la resolución de estos problemas rutinarios (Estrada, 2007). En cambio, los problemas no formalizables, los “mal estructurados”, requieren de otro tipo de estrategias *no rutinarias* que permitan a la organización hacerles frente (Estrada, 2007). Aquí entra, a nivel organizacional el razonamiento heurístico, donde lo que se plantea de fondo es apuntar a cierto tipo de rutinas organizacionales que generen en los agentes una capacidad de respuesta rápida que les permita abordar fenómenos imprevistos que surgen de la incertidumbre. Por esta razón, para Simon, las rutinas



organizacionales deben ser abiertas, flexibles y evitar en la medida de lo posible convertirse en meros patrones repetitivos.

Simon (1947) ya había cuestionado al *homo economicus* y había planteado sustituirlo por el *homo administrativus*, el cual se considera situado en un contexto organizacional. En estos contextos, al igual que en toda esfera humana, se resuelven problemas, pero en Simon el énfasis está en que los agentes reales, de carne y hueso, tienen (y deben tener) como objetivo tomar decisiones suficientemente buenas, no necesariamente decisiones óptimas (Estrada, 2007). El comportamiento humano, en estos contextos organizacionales no es reducible al interés propio porque, según Simon, es “empíricamente falso que los agentes económicos solo busquen ‘maximizar su utilidad’ y aumentar sus beneficios económicos” (Simon, 1993, p. 158). Si el interés propio fuera la única motivación de los agentes, entonces los temas por investigar a nivel organizacional serían el oportunismo y el conflicto de interés. Sin embargo, según Simon, los agentes económicos tienen muchos otros motivos que guían sus acciones, dentro de los cuales planteó la necesidad de reconocer la importancia del altruismo, la lealtad y la identificación grupal, las cuales también se tienen que estudiar para entender bien el comportamiento organizacional.

El altruismo se suele entender como acciones desinteresadas: el agente puede interesarse por la suerte de otros sin esperar alguna compensación económica. Esta idea está presente en Adam Smith cuando habla sobre la simpatía:

Por más egoísta que se pueda suponer al hombre, existen evidentemente en su naturaleza algunos principios que le hacen interesarse por la suerte de otros, y hacen que la felicidad de éstos le resulte necesaria, aunque no derive de ella nada más que el placer de contemplarla. Tal es el caso de la lástima o la compasión, la emoción que sentimos ante la desgracia ajena cuando la vemos o cuando nos la hacen concebir de forma muy vívida (2004, p. 49).

Como podemos ver, en Smith destaca la idea de que los agentes, por más egoístas que sean, se interesan por la suerte de los demás. La razón es psicológica: nuestras emociones nos llevan a interesarnos por el bienestar ajeno. Estas ideas también están presentes en Simon

(1993), quien ve al altruismo como algo estrechamente vinculado a las emociones.

Según Simon (1993, pp. 156-161), la lealtad es un comportamiento profundamente emocional que muestra el compromiso de un agente con respecto a una clase social, una etnia, un género (masculino o femenino) o una organización (empresarial, civil, etc.), y puede llevar al agente a tomar decisiones sin esperar un beneficio económico personal. Por ejemplo, un trabajador puede estar luchando por los intereses materiales de la clase trabajadora, salir a las calles y hacer huelgas para lograr la realización de demandas concretas, y no por ello esperar beneficios económicos personales inmediatos. En este caso hablamos de un comportamiento comprometido, leal con una consciencia de clase, que tiene como consecuencia el altruismo: actuar políticamente sin esperar beneficios económicos personales. Este elemento muestra que la lealtad no es reducible al interés propio. Y la razón es emocional: es lo que el agente siente a nivel emocional, no lo que decide a partir de un cálculo egoísta de costo-beneficio.

La noción de “lealtad” de Simon se parece a la noción de “compromiso” de Amartya Sen (1977 y 2005) en la medida en que ambos autores la presentan como una motivación diferente al interés propio para tomar decisiones. Sin embargo, Sen tiene una noción más amplia de “compromiso” porque, además de integrar elementos emocionales en su caracterización, que también recupera elementos de Adam Smith (en particular, la noción de “simpatía”), integra dos elementos adicionales: el compromiso con normas sociales y con principios morales. El comportamiento de un agente está comprometido con actuar de manera coherente con ciertos compromisos deónticos que le dicen qué debe hacer ante ciertas situaciones, llevando al agente a tener ciertas ideas morales que guían sus prácticas. Este elemento no está del todo desarrollado en Simon, aunque sí lo apunta, como lo mostraremos más abajo cuando hablemos de la importancia de las instituciones sociales.

Otra motivación, diferente del interés propio, es la de actuar con base en la *identidad organizacional* (Simon, 1993, 1995 y 1997). La identidad organizacional se genera en los agentes-individuos a partir de una interacción repetida que hace posible la existencia de organizaciones e instituciones gracias a que todos los individuos actúan de manera altruista entre sí. Esto establece confianza entre los agentes y, con ello, se forman identidades grupales. Lo que se busca es actuar para satisfacer el interés común de todos, no solo actuar para satisfacer intereses

propios individuales de algunos. Construir identidad organizacional es crucial para lograr mejorar el carácter adaptativo de una organización. Justo en este punto podemos ver un paralelismo con la noción de “compromiso” de Sen (1997 y 2005), donde los compromisos expresan también comportamientos diferentes a los intereses propios que llevan a los agentes a tomar decisiones que pueden fomentar la cooperación y la cohesión social.

Como pudimos ver en las secciones 4 y 5, Simon no niega que los agentes tengan una capacidad calculadora que les permita elegir entre diferentes medios para conseguir diversos fines, ni que existan los intereses propios. Pero Simon desidealiza ambos supuestos, en el sentido de que, tras tomar en serio la evidencia empírica disponible sobre el comportamiento humano, cuestiona el uso absolutista de ambos supuestos en los modelos neoclásicos, buscando presentar una versión más suave y realista de ambos.

## 6. La noción de Simon de “racionalidad históricamente acotada”

Hasta ahora hemos visto que las críticas de Simon a los supuestos neoclásicos de racionalidad perfecta e interés propio fueron llevando a que este autor enfatizara la relación agente-entorno, primero, destacando la importancia de las señales que le permitieran al agente reconocer satisfactores; luego, reconociendo en el mismo entorno problemas informacionales estructurales que imposibilitan el uso de técnicas optimizadoras y algorítmicas; después, reconociendo que en las organizaciones es necesario tomar en cuenta la relación entre agentes, la cual puede entenderse no solo basada en el interés propio, sino en la lealtad y la identificación grupal, que permiten el trabajo en equipo para lograr fines colectivos. Estas críticas y alternativas de Simon a la economía neoclásica se fueron incorporando en un marco teórico-metodológico más amplio y evolucionista, donde Simon enfatizó la interacción entre el agente y el entorno con su metáfora de la tijera: “Human rational behavior (and the rational behavior of all physical symbol systems) is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor” (Simon, 1990, p. 7). La metáfora de la tijera ayuda a ver que, si alguna de las partes falta, el todo no puede funcionar. Las tijeras, para cortar, necesitan de la *interacción* entre ambas hojas para funcionar.

Sin embargo, usualmente se interpreta la racionalidad acotada de Simon enfatizando solo una parte de las tijeras: las capacidades cognitivas

limitadas de los agentes. Se retoma muy poco su caracterización de los entornos sociales —la otra parte de la tijera—, que fue llamando su atención durante la década de los noventa. El mismo Simon (1990, pp. 6-7) lo pone así: “to describe, predict and explain the behavior of a system of bounded rationality, we must both construct a theory of the system’s processes and describe the environments to which it is adapting”.

Ya hemos visto en la sección pasada la teoría de las organizaciones de Simon, donde un entorno organizacional involucra problemas rutinarios y no rutinarios, con soluciones programables y no programables, así como elementos no reducibles al interés propio, como el altruismo, la lealtad y la identidad grupal. Sin embargo, Simon fue complejizando y profundizando el análisis del entorno, reconociendo la importancia de la historia en el acotamiento de la racionalidad humana. Lo presentó en estos términos:

A particularly important source of historicity in economics is that boundedly rational economic actors represent the economic scene in radically different ways from time to time, and these changes occur as a function of natural and social events, social influences on perception, and the molding of human motives by the social environment, which is itself time dependent (Simon, 1998, p. 241).

Simon (1998, p. 248) pensaba que la historia podría entenderse como un continuo proceso de cambio que puede tener efectos acumulativos sobre el sistema económico en su conjunto. Así, algunas variables que le dan dinamismo al sistema serían las siguientes:

- a) El cambio tecnológico, que tiene un impacto sobre procesos productivos, los cuales alteran cómo evoluciona un sistema económico al crear nuevas industrias, volver otras obsoletas, o reducir costos.
- b) El contexto institucional, el cual tiene un impacto regulatorio sobre los procesos productivos, como generar expectativas, confianza y credibilidad en los comportamientos de los agentes y establecer qué sí y qué no puede hacerse en un contexto dado.

- c) Variables exógenas institucionales. Simon tiene en mente cambios, paulatinos o repentinos, en las funciones de utilidad de los agentes, que pueden alterar la demanda y la oferta, así como cambios en las leyes de propiedad.

A estas variables, que parecen operar en un nivel macrosocial, Simon (1998, p. 251) busca agregar otras, también dinámicas, que parecen operar en un nivel más microsocial, que impacta directamente en el comportamiento de agentes con racionalidad acotada:

- 1) El conocimiento y la información sobre el mundo y sobre cómo funciona la economía. Simon (1998, pp. 251-253) destaca tanto las creencias que tienen los economistas adscritos a diferentes escuelas de pensamiento económico (desde los mercantilistas, los fisiócratas, los economistas clásicos) como las creencias populares sobre cómo se cree que funcionan los mecanismos de la economía, lo cual puede guiar tanto la formación de políticas públicas como la estructuración de fenómenos económicos (por ejemplo, si se cree que habrá escasez de comida, ello hace que haya agentes que demanden más productos, lo cual subirá el precio en los mercados).
- 2) La habilidad humana para estimar las consecuencias de acciones. Simon (1998, p. 254) destaca las habilidades humanas que pueden usarse para derivar implicaciones precisas sobre diferentes decisiones. Destacan también herramientas técnicas que pueden mejorar estas habilidades.
- 3) El entorno institucional en que el comportamiento económico toma lugar. Simon (1998, p. 254) dijo que las decisiones económicas están fuertemente influenciadas por el entorno gubernamental, legal, de empresas e instituciones sociales, el cual es cambiante.

- 4) El foco de atención de las creencias y expectativas. Simon (1998, p. 255) menciona cómo algunas variables juegan un papel más importante que otras y cómo ello va cambiando a lo largo del tiempo. Un ejemplo es cómo, durante los cincuenta y sesenta, la población no prestaba atención al nivel de precios, pero sí al nivel de empleo y de productividad, pero luego, cuando se dio el fenómeno del estancamiento con inflación de los setenta, empezó a preocupar más el rápido incremento de precios.
- 5) El altruismo humano. Simon (1998, p. 257), como en otros trabajos, no niega que exista el interés propio, pero sí cree que existen otras motivaciones para actuar que no tienen que ver con obtener beneficios propios. Un ejemplo es la “docilidad”, que se puede entender como la tendencia de los agentes a aceptar instrucciones, consejos y autoridad cuando esta emana de fuentes socialmente legitimadas. Un aspecto de la docilidad, según Simon, es la tendencia a que las personas se identifiquen y sean leales a un grupo social, incluidas las organizaciones a las cuales pertenecen.
- 6) La identificación grupal. Simon (1998, pp. 257-258) destaca, como hemos visto antes, cómo el comportamiento humano está muy ligado a la lealtad que se siente por grupos sociales (la familia, la clase social, el género o una nación). A nivel organizacional, Simon destaca que las motivaciones para actuar no necesariamente pueden reducirse a recompensas, pero sí a lealtades. Si en una organización hay identificación grupal, entonces habrá más facilidad para impulsar el objetivo general, colectivo, que en organizaciones donde no hay identificación grupal.

Todas estas variables hacen que la visión de Simon sobre el comportamiento económico sea más dinámica y compleja que la visión

neoclásica, y que, además, el comportamiento económico pueda variar de forma radical, según Simon, en diferentes momentos históricos y en diferentes grupos sociales, ya sea por las ideas económicas dominantes del momento, las condiciones técnicas disponibles, el contexto socio-institucional vigente (en el que destacan tanto leyes jurídicas como costumbres del momento), o por la existencia de diferentes grupos sociales. Simon, me parece, tampoco negaría la importancia de la política y de las relaciones de poder entre diferentes grupos sociales, aunque este no es un tema que desarrolle en sus trabajos.

Es importante recordar aquí que Simon (1998, p. 248) está intentando convencer a los economistas de que deberían incorporar la historia de una forma más fuerte en sus planteamientos teóricos. De ahí que él sugiera que estas variables se pueden endogeneizar en modelos dinámicos, como los sistemas de ecuaciones diferenciales, donde se puede dar cuenta de los efectos acumulativos que generan todas estas variables en el comportamiento y trayectoria del sistema total.

Para Simon es fundamental recuperar la dimensión histórica de la economía: que los datos recogidos y analizados por los historiadores nos ayuden a entender los mecanismos que alteran el comportamiento económico por medio de cambios en las seis variables arriba descritas. Ello compromete a Simon con una visión *evolucionista* porque, a diferencia de la economía neoclásica, Simon está concibiendo al sistema social como algo vivo, en constante transformación, no situado en un punto de equilibrio o de reposo. La trayectoria del sistema es altamente dependiente del pasado. Las decisiones de los agentes que estructuran comportamientos organizacionales, así como el cambio tecnológico y las decisiones de gobierno, le imprimen dinamismo al sistema social en su conjunto.

Esta visión histórica de la racionalidad, lejos de contradecir lo que Simon pensaba sobre las invarianzas del comportamiento humano, parecen complementarlo:

Many of the invariants we see in behavior are *social invariants*. And since they are social invariants, many are invariant only over a particular society or a particular era, or even over a particular social or professional group within a society. Social variables must be introduced to set the boundaries of our generalizations (Simon, 1990, p. 16; cursivas mías).

Si bien Simon (1990 y 1998) pensaba que la ciencia busca invarianzas tanto en la naturaleza como en la sociedad, como podemos ver en el pasaje citado, Simon mismo reconoce que muchas de las invarianzas que están presentes en el comportamiento son sociales, y estas son solo aplicables a una sociedad particular, o a una época o incluso a un grupo social en particular. Estas particularidades abren paso, por tanto, a la posibilidad de que las invarianzas en el comportamiento de los agentes puedan ser válidas solo en contextos limitados, ya sea al nivel de una sociedad, de una época, o de un grupo social. Se trata, por tanto, de invarianzas social e históricamente acotadas. Lo importante es que Simon (1990, p. 18) buscó estas invarianzas del comportamiento en los mecanismos de adaptación de los agentes a su entorno, mecanismos que pueden cambiar según diferentes circunstancias históricas y sociales. Una estrategia que Simon (1990, p. 18) usó para buscar estas invarianzas fue analizar la toma de decisión de expertos, que retienen información adaptativa que les ayuda a resolver problemas cotidianos. Pero estos mecanismos, siguiendo a Simon pueden variar en cada grupo de la sociedad a lo largo del tiempo.

Si bien Simon logró avanzar hacia una perspectiva más histórica de la economía presentando intuiciones sobre cómo modelar procesos históricos en sistemas de ecuaciones diferenciales, así como al presentar una visión de la racionalidad como históricamente acotada, sus avances en esta dirección quedaron trancos. En Simon no hay una visión sistemática que explique cómo pueden darse estas modificaciones de los mecanismos de adaptación a cambios en el entorno, ni tampoco una visión que dé cuenta del carácter socialmente distribuido de los mecanismos adaptativos. Esto, en gran medida, es atribuible a que Simon siguió asumiendo, como veremos en la siguiente sección, un supuesto en común con la economía neoclásica que le impidió tener una visión más histórica y social que le permitiera orientar su investigación hacia temas más históricamente sustantivos: una variante del individualismo metodológico que llamaremos “supuesto de cognición individualista”.

## **7. La cognición individualista y sus límites**

El supuesto de cognición individualista consiste en asumir que los procesos cognitivos involucrados en la toma de decisiones de los agentes, entre ellos el razonamiento heurístico, ocurren al interior de la mente de agentes individuales (Martínez, 2016). Según Zerubavel y Smith (2010), el individualismo cognitivo busca explicar los procesos cognitivos —



como percibir, vigilar, recordar, contextualizar, generalizar, clasificar, interpretar, el reconocimiento del tiempo, etc. — solamente en términos de procesos mentales que ocurren dentro de la cabeza de individuos. Si bien esta manera de ver la cognición ha generado importantes avances en las ciencias cognitivas (como las neurociencias y la psicología de la toma de decisiones), ha llevado a centrar el análisis solo en el individuo aislado, prestando una insuficiente atención a la interacción de los agentes con su entorno social (Zerubavel y Smith, 2010).

Simon parece haber asumido un individualismo cognitivo en sus estudios sobre cognición, pues piensa que la mente es una especie de computadora, y que esta está situada en la cabeza de individuos (Hernández Cervantes, 2019). Si bien quería enfatizar la interacción agente-entorno, este supuesto de cognición individualista hizo que Simon abordase poco los efectos que esas interacciones generan sobre la estructuración del razonamiento heurístico. En ese sentido, el entorno no juega un papel sustantivo en la estructuración de comportamientos (Bonome, 2009, p. 98). Por ejemplo, en el caso del vendedor de casas o el de la hormiga en la playa, el razonamiento de ambos agentes ya está estructurado y lo único que hace el entorno es dar señales al agente que activa respuestas ya preestablecidas y que están ya “en la cabeza” de los agentes.

El entorno no juega un papel sustantivo en la formación y estructuración de la racionalidad porque Simon mismo no vislumbró todas las consecuencias que se siguen de su propia metáfora de la tijera: que agente y entorno son una unidad inseparable. Simon parece no tener una categoría intermedia que ayude a ver el carácter incrustado (*embedded*) entre el agente y el entorno, que en la metáfora de la tijera sería justo el punto de intersección entre ambas hojas. Es decir, no tiene una noción relacional que sea equiparable a los *habitus* (Bourdieu, 2000; Merleau-Ponty, 2013) o los hábitos (Dewey, 2002; Cohen, 2007) que, por referir a predisposiciones a actuar que se forman mediante complejos procesos de socialización, visibilizan los andamiajes que se establecen entre el agente y su entorno, pues los hábitos se adquieren mediante experiencias repetidas en entornos sociales estables a lo largo del tiempo.

La falta de una noción relacional que juegue el papel de andamiaje entre el agente y el entorno llevó a Simon a estudiar el razonamiento heurístico de forma individualista. Estudió, mediante programas de simulación computacional, lo que él pensaba que era el razonamiento heurístico que intentó equiparar a las mentes humanas individuales.

La cognición, según Simon, solo se puede atribuir a individuos, no a entidades colectivas, y las heurísticas están situadas en la mente de individuos que piensan como si fueran computadoras, pues estas poseen capacidades de procesamiento, memoria, cálculo y predicción. El razonamiento heurístico estaría esencialmente ligado a las capacidades cognitivas de agentes individuales, que desencadenan respuestas que solucionan problemas cuando reconocen señales en el entorno. Pero esta manera de ver el razonamiento heurístico no es relacional, pues no enfatiza la relación agente-entorno; con ello, no puede dar cuenta de cómo el propio razonamiento heurístico evoluciona, cambia y se moldea por medio de diferentes prácticas que están situadas y que están socialmente distribuidas (Martínez, 2003 y 2016). La racionalidad procesual es ella misma invariante aunque Simon mismo haya dicho que esta puede variar de grupo en grupo y de época en época; el punto es que él nunca explicó cómo se dan estos cambios en los mecanismos de adaptación al entorno. En Simon (1990), las heurísticas ya están situadas en la mente humana y operan como si fueran un programa de computadora, que desata respuestas automáticas, presentando así una posición mecanicista que contrasta con su posición evolucionista más general.

En resumen: a Simon le hizo falta romper con el individualismo cognitivo para hablar de los procesos cognitivos, en particular el razonamiento heurístico, como una capacidad cognitiva moldeable por la práctica social (González, 2003). Entender la racionalidad como algo meramente individual hizo que la ruptura de Simon con la economía neoclásica fuera limitada, lo cual hizo que su visión histórica de la racionalidad fuera limitada también. Sin embargo, en los últimos trabajos de Simon, como hemos visto, se apuntó hacia la posibilidad de desarrollar una noción de “racionalidad” más social e históricamente acotada, que puede seguir siendo desarrollada. En el siguiente apartado mostraremos cómo desarrollar más esta perspectiva histórica de la racionalidad recuperando la noción de “hábito” de Dewey.

## **8. Hacia una noción más social de la cognición: la relación entre el razonamiento heurístico y los hábitos**

El supuesto de cognición individualista se puede cuestionar porque este, al centrarse en los individuos, le quita carácter explicativo al entorno social (Echeverría y Álvarez, 2008; Hernández Cervantes, 2019).

Una manera de alejarse del supuesto de cognición individualista y de destacar la relación de andamiaje en que se ve al agente y al entorno como una unidad inseparable dentro de un marco totalmente evolucionista es recuperando la noción de “hábito” de John Dewey.<sup>2</sup> Para Dewey (2002, p. 12), “all conduct is *interaction* between elements of human nature and the environment, natural and social”. Cuando estas interacciones son repetidas en entornos estables a lo largo del tiempo, según Dewey, se van formando predisposiciones a actuar que forman el comportamiento de los agentes, donde “the nature of habit is to be assertive, insistent, self perpetuating” (Dewey, 2002, p. 49). También los agentes adquieren, mediante un complejo proceso de socialización, algunas habilidades que les permiten adaptarse a cambios a los entornos, pues el hábito “is an ability, an art, formed through past experience” (Dewey, 2002, p. 52).

Para Dewey, si bien los hábitos tienen una tendencia hacia la perpetuación, no son meras repeticiones automáticas y rígidas. Los hábitos no están muertos porque también son habilidades adquiridas que los hacen ser herramientas cognitivas que sirven para resolver los diferentes problemas a los cuales los agentes se enfrentan en sus procesos de socialización. En ese sentido, los hábitos son herramientas adaptables a diferentes prácticas sociales que se adquieren cuando un agente entra a formar parte de esas prácticas. Incorporar la idea de “hábito” al análisis del razonamiento heurístico puede llevar a desarrollar una idea a la que Simon apuntó: que la racionalidad puede ser social e históricamente acotada, y que está socialmente distribuida en diferentes prácticas sociales (Martínez, 2003).

Pero los hábitos —y esta es la clave que le hizo falta a Simon— forman actitudes, creencias y razones para actuar en los agentes (Cohen, 2007; Dewey, 2002). Los hábitos forman voluntades porque pueden ser ellos mismos herramientas adaptables que pueden modificar entornos sociales. Por ser herramientas adaptables, los hábitos nunca son invariantes, sino que son productos históricos que pueden modificarse a lo largo del tiempo, pero contienen información adaptativa que puede transmitirse a otras generaciones (Bardone, 2011; Hernández Cervantes, 2019). Es

---

<sup>2</sup> Las nociones de *habitus* de Merleau-Ponty (2013) y de Bourdieu (2000) también son convergentes con el punto que se va a desarrollar a continuación, pues, al igual que Dewey, estos dos autores enfatizan el carácter incrustado de la relación agente-entorno, donde cambios en el entorno generan cambios en el agente, y los *habitus* son la mediación.

decir, los hábitos contienen información de lo que otros agentes hicieron durante un tiempo en un entorno social estable. A Simon le interesó retener esta idea de transmisión de información adaptativa, como vimos en la sección pasada; esta puede desarrollarse aún más retomando la noción de “hábito” de Dewey, así como otras caracterizaciones que se hacen desde las ciencias cognitivas contemporáneas (Chemero, 2011; Fonseca, 2019; Martínez, 2016).

La implicación de poner en el centro la noción de “hábito” es que se da cuenta de las variaciones de la racionalidad humana, lo cual, a su vez, puede explicar la gran pluralidad de racionalidades que pueden existir en las diferentes sociedades, en diferentes momentos históricos y en diferentes grupos sociales. En ese sentido, la noción de “hábito” sí permite dar cuenta de cómo la historia moldea diferentes formas de razonar en los agentes, además de que permite entender el carácter socialmente distribuido de las diferentes maneras de razonar y de actuar en el mundo, pues los hábitos están estrechamente relacionados con las prácticas de diferentes grupos sociales. Como dice Dewey (2002, p. 100): “man is a creature of habit, not of reason nor yet of instinct”.

Las costumbres son a las colectividades lo que los hábitos son a los individuos (Dewey, 2002). Esta idea muestra que el razonamiento está vinculado con el entorno social, y que este mismo se moldea de manera social. En ese sentido, las capacidades cognitivas del agente no serían una hoja separada del entorno social, como lo pensó Simon con su metáfora de la tijera. Además, se evita tener una relación tensa entre una posición mecanicista en el terreno cognitivo y una visión evolucionista en el terreno social, tal y como Simon la tuvo. Más bien, siguiendo a Dewey, tendríamos una posición evolucionista en ambos niveles: en el del agente y en el del entorno social.

La interacción entre el agente y el entorno es *constitutiva* de la racionalidad del agente porque estos son inseparables. Ya lo vimos con la metáfora de la tijera, en la que las capacidades cognitivas del agente y la estructura informacional del entorno son, al igual que las tijeras, inseparables. En ese sentido, agente y entorno están incrustados y tienen una relación interactiva, donde ambos se codeterminan a lo largo del tiempo. La mediación/andamiaje entre el agente y el entorno es la noción de “hábito”, y ello nos permite ver el carácter adaptativo y, sobre todo, la “plasticidad” del razonamiento heurístico (Bardone, 2011). De esta forma podemos evitar pensar que las heurísticas son una clase natural fija y ya dada, situadas en la cabeza de individuos (Hernández

Cervantes, 2019), para más bien dar cuenta de su carácter evolutivo y cambiante.

## 9. Comentarios finales

En este artículo analizamos lo que Simon entiende por “racionalidad acotada”, expusimos su crítica a la noción neoclásica de “agente racional” y revisamos sus trabajos sobre satisfaccionismo, inteligencia artificial, su teoría de las organizaciones y su manera de plantear el carácter históricamente acotado de la racionalidad. Luego, al identificar que mantiene una versión del individualismo metodológico que llamamos “supuesto de cognición individualista”, pasamos a criticar ese supuesto y terminamos mostrando una posibilidad de desarrollo de la noción de “racionalidad históricamente acotada” por medio de la recuperación de la noción de “hábito” de Dewey.

A partir de los elementos presentados en este trabajo, una primera conclusión es que la crítica de Simon a la noción neoclásica de “agente racional” lo llevó a formular una noción de “racionalidad acotada” donde lo acotado está en negar los dos supuestos neoclásicos básicos de su noción de “agencia”: que la racionalidad de los agentes es perfecta y que su interés propio sea ilimitado. No hay necesidad de retener alguna noción de “optimización” para poder modelar el comportamiento económico. Una segunda conclusión es que la crítica al *homo economicus* llevó a Simon a apuntar hacia una noción de “racionalidad históricamente acotada” donde sus trabajos fueron subrayando el análisis de las interacciones entre el agente y un entorno social evolutivo, sujeto a constantes transformaciones y que puede llevar a cambios en el comportamiento y razonamiento del agente. Sin embargo, una tercera conclusión apunta a que Simon, pese a sus intentos por elaborar una noción históricamente acotada de la racionalidad, se mantuvo cercano a la visión neoclásica porque siguió asumiendo una versión del individualismo metodológico para estudiar la racionalidad humana: el supuesto de cognición individualista. Pero vimos que una forma de alejarse del individualismo cognitivo para acercarse a una cognición más social es por medio de retomar el concepto de “hábito” de Dewey, que puede dar cuenta del carácter moldeable y adaptable del razonamiento heurístico.

Futuras investigaciones podrían ir en la dirección de elaborar y profundizar más la relación entre razonamiento heurístico y hábitos para avanzar hacia una noción de “racionalidad” social e históricamente

acotada. En este punto, los trabajos contemporáneos de ciencias cognitivas (Chemero, 2011; Fonseca Patrón, 2019; Martínez, 2016), así como los trabajos en ciencias sociales sobre hábitos (Cohen, 2007; Dewey, 2013) y *habitus* (Bourdieu, 2000; Merleau-Ponty, 2013) pueden ser útiles para entender la cognición como un sistema complejo que emerge a partir de la interacción repetida entre el sistema sensorio-motriz de los agentes, sus sistemas neuronales y sus entornos materiales, que involucran tanto artefactos tecnológicos como instituciones socioculturales. La interacción repetida entre estos sistemas forma andamiajes entre los agentes y sus entornos, que luego pueden formar hábitos cognitivos cambiantes y distribuidos en diferentes prácticas sociales.

## Referencias

- Angner, E. y Loewenstein, G. (2012). Behavioral Economics. En U. Mäki (ed.), *Philosophy of Economics*. (pp. 641-690). Elsevier. DOI: <https://doi.org/10.1016/B978-0-444-51676-3.50022-1>.
- Arthur, W. B. (2015). *Complexity and the Economy*. Oxford University Press.
- Augier, M. y March, J. G. (2004). Herbert A. Simon, Scientist. En M. Augier y J. G. March (eds.), *Models of a Man: Essays in Memory of Herbert A. Simon*. Volumen 1. (pp. 3-32). The MIT Press.
- Becker, G. (1993). Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy*, 101(3), 385-409. DOI: <https://doi.org/10.1086/261880>.
- Berg, N. y Gigerenzer, G. (2010). As-if Behavioral Economics: Neoclassical Economics in Disguise? *History of Economics Ideas*, 18, 1000-1033.
- Bonome, M. G. (2009). *La racionalidad en la toma de decisiones: análisis de la teoría de la decisión de Herbert A. Simon*. Netbiblo.
- Bourdieu, P. (2000). *Pascalian Meditations*. R. Nice (trad.). Stanford University Press.
- Bowles, S. (2014). Niccolò Machiavelli and the Origins of Mechanism Design. *Journal of Economic Issues*, 48(2), 267-278. DOI: <https://doi.org/10.2753/JEI0021-3624480202>.
- Chemero, A. (2011). *Radical Embodied Cognitive Science*. The MIT Press.
- Cohen, M. (2007). Reading Dewey: Reflections on the Study of Routine. *Organization Studies*, 28(5), SAGE Publications.
- Dewey, J. (2002). *Human Nature and Conduct: An Introduction to Social Psychology*. Prometheus.

- Elster, J. (2011). *El desinterés. Tratado crítico del hombre económico. (I)*. R. Ancira (trad.). Siglo XXI.
- Echeverría, J. y Álvarez, J. F. (2008). Bounded Rationality in Social Sciences. En E. Agazzi, J. Echeverría y A. Gómez Rodríguez (eds.), *Epistemology and the Social*. (pp. 173-191). Brill.
- Estrada, F. (2007). Herbert A. Simon y la economía organizacional. *Cuadernos de Economía*, 26(46), 169-199.
- Fonseca Patrón, A. L. (2019). Cognición situada y racionalidad. Hacia una ecología interactiva del razonamiento. *Diánoia*, 64(83), 103-131. DOI: <https://doi.org/10.22201/iifs.18704913e.2019.83.1618>.
- Friedman, M. y Savage, L. (1952). The Expected-Utility Hypothesis and the Measurability of Utility. *Journal of Political Economy*, 60(6), 463-474. DOI: <https://doi.org/10.1086/257308>.
- Gigerenzer, G. (2008). *Rationality for Mortals: How People Cope with Uncertainty*. Oxford University Press.
- González, W. J. (2003). Racionalidad y Economía. De la racionalidad de la economía como ciencia a la racionalidad de los agentes económicos. En W. J. Gonzalez (ed.), *Racionalidad, historicidad y predicción en Herbert A. Simon*. (pp. 65-96). Netbiblo.
- Heidl, S. (2016). *Philosophical Problems of Behavioral Economics*. Routledge.
- Hernández Cervantes, J. I. (2019). *Racionalidad acotada, heurísticas y hábitos. Implicaciones de diferentes caracterizaciones del razonamiento heurístico para la formulación de políticas públicas*. [Tesis doctoral]. UNAM.
- Heukelom, F. (2014). *Behavioral economics: A history*. Cambridge University Press.
- Hortal, A. (2017). Empiricism in Herbert Simon: Administrative Behavior Within the Evolution of the Models of Bounded and Procedural Rationality. *Brazilian Journal of Political Economy*, 37, 719-733. DOI: <https://doi.org/10.1590/0101-31572017v37n04a04>.
- Kahneman, D. y Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica: Journal of the Econometric Society*, 47(2), 263-291.
- Kirchgässner, G. (2014). On Self-Interest and Greed. *Journal of Business Economics*, 84(9), 1191-1209. DOI: <https://doi.org/10.1007/s11573-014-0737-1>.
- Laibson, D. y List, J. A. (2015). Principles of (Behavioral) Economics. *American Economic Review*, 105(5), 385-390. DOI: <https://doi.org/10.1257/aer.p20151047>.

- Lazear, E. (2000). Economic Imperialism. *The Quarterly Journal of Economics*, 115(1), 99-146. DOI: <https://doi.org/10.1162/003355300554683>.
- Mäki, U. (2008). Filosofía y metodología de la economía. *Cátedra Bolívar. Temas de teoría económica y su método 2*. J. J. Jardón Urrieta (trad.). Universidad de Santiago de Compostela. URL: <https://www.google.com/url?client=internal-element-cse&cx=007758419574493913536:b3wgmoma76g&q=https://www.usc.es/economet/aeeadepdf/aeade104.pdf&sa=U&ved=2ahUKEwiS-t5uaaBAxVDFBAlHZELB7UQFnoEACAcQAQ&usg=AOvVaw3RkhzPe3CBeD40LDBoLsHs>.
- Mandeville, B. (1997). *The Fable of the Bees and Other Writings*. Hackett Publishing.
- Martínez, S. F. (2003). *Geografía de las prácticas científicas. Racionalidad, heurística y normatividad*. UNAM.
- (2016). Cultura material y cognición social. En P. Hernández, J. García y M. Romo (ed.), *Cognición. Estudios multidisciplinares*. (pp. 247-264). Centro de Estudios Filosóficos, Políticos y Sociales Vicente Lombardo Toledano.
- Merleau-Ponty, M. (2013). *Phenomenology of Perception*. C. Smith (trad.). Routledge.
- Mill, J. S. (2010). *The Basic Writings of John Stuart Mill: On Liberty, the Subjection of Women and Utilitarianism*. Modern Library.
- Morgan, M. S. (2012). *The World in the Model: How Economists Work and Think*. Cambridge University Press.
- Morgenstern, O. y Von Neumann, J. (1953). *Theory of Games and Economic Behavior*. Princeton University Press.
- Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica: Journal of the Econometric Society*, 29(3), 315-335. DOI: <https://doi.org/10.2307/1909635>.
- Newell, A. (1983). The Heuristic of George Polya and Its Relation to Artificial Intelligence. En R. Groner, M. Groner y W. F. Bischoff (eds.), *Methods of Heuristics*. (pp. 195-243). Lawrence Erlbaum.
- Ross, D. (2014). *Philosophy of Economics*. Palgrave Macmillan.
- Sargent, T. J. (1993). *Bounded Rationality in Macroeconomics: The Arne Ryde Memorial Lectures*. Oxford University Press.
- Segoviano Contreras, L. E. y Morales Sánchez, M. A. (2021). El principio de interés propio en el análisis y el diseño económico. *Interdisciplina*, 9(25), 185-208.
- Sen, A. K. (1977). *Rational Fools: A Critique of the Behavioral Foundations*



- of Economic Theory. *Philosophy & Public Affairs*, 317-344. URL: <https://www.jstor.org/stable/2264946>.
- \_\_\_\_ (2005). Why Exactly Is Commitment Important for Rationality? *Economics & Philosophy*, 21(1), 5-14. DOI: 10.1017/S0266267104000355.
- Simon, H. A. (1947). *Administrative Behavior. A Study of Decision-Making Processes in Administrative Organization*. The Free Press-Collier Macmillan.
- \_\_\_\_ (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1), 99-118. DOI: <https://doi.org/10.2307/1884852>.
- \_\_\_\_ (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, 63(2), pp. 129-138. DOI: <https://psycnet.apa.org/doi/10.1037/h0042769>.
- \_\_\_\_ (1969). *The Sciences of the Artificial*. The MIT Press.
- \_\_\_\_ (1979). Rational Decision Making in Business Organizations. *The American Economic Review*, 69(4), 493-513.
- \_\_\_\_ (1990). Invariants of Human Behavior. *Annual Review of Psychology*, 41(1), 1-20.
- \_\_\_\_ (1993). Altruism and Economics. *The American Economic Review*, 83(2), 156-161.
- \_\_\_\_ (1994). Testability and Approximation. En D. M. Hausman (ed.), *The Philosophy of Economics: An Anthology*. (pp. 214-216). Cambridge University Press.
- \_\_\_\_ (1995). Rationality in Political Behavior. *Political Psychology*, 16(1), 45-61. DOI: <https://doi.org/10.2307/3791449>.
- \_\_\_\_ (1997). *An Empirically Based Microeconomics*. Cambridge University Press.
- \_\_\_\_ (1998). Economics as a Historical Science. *Theoria: An international Journal for Theory, History and Foundations of Science*, 13(2), 241-260.
- Simon, H. A. y Newell, A. (1958). Heuristic Problem Solving: The Next Advance in Operations Research. *Operations Research*, 6(1), 1-10. DOI: <https://doi.org/10.1287/opre.6.1.1>.
- \_\_\_\_ (1962). Computer Simulation of Human Thinking and Problem Solving. *Monographs of the Society for Research in Child Development*, 27(2), 137-150. DOI: <https://doi.org/10.2307/1165535>.
- Smith, A. (2004). *La teoría de los sentimientos morales*. C. Rodríguez Braun (trad.). Alianza.
- Thaler, R. (2015). *Misbehaving: The Making of Behavioral Economics*. Norton.

Zerubavel, E. y Smith, E. R. (2010). Transcending Cognitive Individualism. *Social Psychology Quarterly*, 73(4), 321-325. DOI: <https://doi.org/10.1177%2F0190272510388998>.