

# EPITOME SOBRE RACIONALIDAD

Jesús MOSTERIN\*

El presente epítome no pretende sino presentar, desde mi perspectiva, un breve resumen de los principales contenidos, resultados y problemas de la teoría de la racionalidad individual, tal y como ésta se ha desarrollado en el siglo XX.

## 3 sentidos de *racionalidad*

La palabra racionalidad se emplea en al menos tres sentidos distintos:

(1) Racionalidad como capacidad lingüística. En este sentido -el más débil- un ser racional es un ser capaz de comunicarse lingüísticamente. Y el humano es el animal racional.

(2) Racionalidad como razonabilidad. En este sentido alguien es racional si da (o está dispuesto a dar) razones para decir o hacer lo que dice o hace. Y toda persona razonable y bien educada es racional.

(3) Racionalidad en sentido fuerte, que presupone la capacidad lingüística y la razonabilidad, pero va más allá, incluyendo típicamente procesos de evaluación y optimización.

A partir de ahora sólo hablaré de racionalidad en este sentido fuerte.

## En qué situaciones se plantea el tema

Hablamos de racionalidad en contextos en los cuales un agente elige o toma decisiones en función de ciertas preferencias u objetivos. Por tanto, sólo tiene sentido hablar de racionalidad en situaciones que reúnan estas dos condiciones:

(1) Que haya diversas alternativas entre las que elegir; es decir, que la salida de la situación no esté unívocamente determinada.

(2) Que no todo dé igual. Que unas salidas a la situación sean preferibles a otras.

En situaciones deterministas, donde no hay nada que decidir o elegir, no se plantean problemas de racionalidad. Y en situaciones de indiferencia o frivolidad, donde no hay objetivos que alcanzar ni referencias que satisfacer, tampoco se plantean problemas de racionalidad.

Un cierto margen de maniobra, y, algún tipo de preferencias u objetivos han de estar dados, para que tenga sentido siquiera hablar de racionalidad.

Cuando decidimos qué cosas creer, hablamos de racionalidad teórica o credencial. Cuando decidimos qué cosas hacer, hablamos de racionalidad práctica.

Tanto sobre la racionalidad teórica, como sobre la práctica, se ha desarrollado una teoría formal (de carácter básicamente matemático y respecto a la que reina

amplio consenso) y una teoría material (que trata de reducir la infradeterminación de las creencias y acciones por la teoría formal mediante un cierto anclaje en la realidad y una cierta conexión con nuestra naturaleza, y respecto a la cual hay menos consenso).

### **Teoría formal de la racionalidad teórica**

La teoría formal de la racionalidad creencial o teórica indaga las condiciones formales que tiene que satisfacer el conjunto de todas las creencias de un agente (o creyente) dado  $x$  para que digamos que  $x$  es racional en sus creencias. Puesto que las creencias varían con el tiempo, las condiciones han de ser relativizadas a un instante determinado  $t$ .

Las condiciones formales de la racionalidad teórica del agente (o creyente)  $x$  en el instante  $t$  son las siguientes:

(1) **Coherencia:** el conjunto de todas las creencias de  $x$  en  $t$  ha de ser consistente.

(2) **Clausura:** el contenido de todas las creencias de  $x$  en  $t$  ha de estar clausurado respecto a la relación de implicación, es decir, si  $x$  cree que  $A$ , y  $A$  implica  $B$ , entonces  $x$  cree que  $B$ . En otras palabras,  $x$  cree todas las consecuencias de sus creencias.

(3) **Probabilidad:** Si  $x$  asigna probabilidades subjetivas a sus creencias, lo ha de hacer de un modo compatible con la teoría de la probabilidad. Por ejemplo, si  $x$  asigna la probabilidad  $1/3$  a  $B$ , entonces tiene que asignar  $2/3$  a  $\text{no-}B$ .

La condición (1) es la esencial. Lo primero que prohíbe la racionalidad teórica es que nos contradigamos en nuestras creencias, que creamos a la vez una cosa y su contraria,  $B$  y  $\text{no-}B$ . A pesar de todo, tal y como está formulada, resulta quizás demasiado fuerte, pues basta con que varias de las creencias impliquen una contradicción para que el conjunto sea inconsistente, aunque esa implicación sea muy poco obvia e incluso sumamente difícil de descubrir. Por ello esa condición es a veces sustituida por algo más débil, pero más realista, tal como:  $x$  está dispuesto a revisar su sistema de creencias, siempre que detecte en él alguna contradicción.

La condición (2) es excesivamente fuerte e irrealista. De hecho nadie conoce todas las consecuencias de sus creencias. Por ello puede ser sustituida por algo así como:  $x$  está dispuesto a aceptar como creencia suya cualquier consecuencia de sus creencias.

La condición (3) sólo tiene aplicación si el creyente asigna probabilidades. En ese caso también resulta fuerte y puede ser sustituida por:  $x$  está dispuesto a revisar su distribución de probabilidades subjetivas, siempre que se detecte alguna incompatibilidad con la teoría de la probabilidad.

### Teoría material de la racionalidad teórica

En definitiva la noción formal de racionalidad teórica se reduce a la de consistencia lógica, y es compatible con cualesquiera contenidos de creencia, por muy lunáticos que éstos puedan ser. (Si creo que soy Napoleón y que Napoleón nació en Córcega, tengo que creer -por racionalidad formal- que yo nací en Córcega; pero ni yo soy Napoleón ni nací en Córcega; la consistencia es compatible con la falsedad más obvia). Los constreñimientos formales (incluso en sus versiones iniciales, irrealistamente fuertes) son demasiado débiles por sí mismos para caracterizar completamente la racionalidad teórica. De algún modo tenemos que tocar tierra, de algún modo tenemos que atar ese globo consistente de creencias a la realidad.

Los animales necesitamos información sobre el entorno para sobrevivir. Un conjunto inconsistente de creencias contiene información cero acerca del entorno. Pero no todo conjunto consistente de sentencias contiene información positiva acerca del entorno. La descripción de una ciudad ficticia puede ser tan consistente como la de nuestra ciudad real.

Las ataduras materiales principales que podemos exigir del conjunto de creencias de un agente racional son la atadura a la percepción y la atadura a la ciencia (es decir, el encaje con la racionalidad teórica colectiva).

El realismo (pragmático) de la percepción sensible está garantizado por su éxito evolutivo. Nuestras sensaciones nos suministran información veraz sobre la realidad del mundo, aunque no nos suministren toda la información objetivamente presente, sino sólo aquella que es (o ha sido) relevante para nuestra supervivencia. (La distinción entre información objetiva presente e información subjetivamente detectada y codificada es comparable a la distinción termodinámica entre energía libre, es decir, aprovechable).

Las raras ilusiones perceptivas (ilusiones ópticas y otras) no cambian nada en el hecho de la fiabilidad básica de nuestro aparato sensorial. La desconfianza retórica de algunos filósofos respecto a los sentidos no deja de ser bastante teatral. De hecho estoy más seguro de la existencia de mi camisa que de la existencia de mi yo, aunque no sea más que porque la camisa me viene dada, mientras que el yo es una construcción problemática y elusiva.

De todos modos, nuestro lenguaje y nuestro pensamiento y nuestras creencias van mucho más allá del mundo relativamente limitado de las percepciones. Por eso la atadura a la percepción no basta. Otras ataduras son requeridas, sobre todo la atadura a la ciencia. Si el agente tiene suficiente tiempo e interés por un tema como para analizarlo por su cuenta, entonces puede examinar la evidencia matemática y empírica correspondiente, pero si no, si ha de apoyarse en opiniones de tercera mano, entonces lo mejor que puede hacer es aceptar la opinión consensuada de la comunidad científica pertinente. (No se trata de aceptar acríticamente autoridad alguna, sino de, puestos a aportar, apostar por el mejor caballo).

En cualquier caso, el requisito de la racionalidad formal exige que nuestras creencias no se contradigan entre sí y, en especial, que las creencias no perceptivas no contradigan a las perceptivas. Un sistema materialmente racional de creencias debe ser al menos compatible con las percepciones del sujeto, e incluso tratar de dar cuenta de ellas.

### **Teoría formal de la racionalidad práctica**

La teoría de la decisión o teoría formal de la racionalidad práctica considera tres tipos distintos de situaciones: situaciones en las que el agente ha de tomar una decisión sabiendo con certeza cuáles serán las consecuencias de cada una de las posibles acciones entre las que tiene que elegir, situaciones en las que el agente decide bajo riesgo y, finalmente, situaciones de incertidumbre. Recordemos brevemente los tres problemas de decisión:

(1) **Decisión bajo condiciones de certeza.** Se trata de maximizar o minimizar un parámetro, bajo ciertos constreñimientos. El ejemplo típico es el de la confección de un menú que trate de minimizar el coste, bajo constreñimientos dietéticos sobre mínimos de calorías, proteínas, hidratos de carbono y vitaminas. La técnica formal para resolver este tipo de problemas es la programación lineal.

(2) **Decisión bajo condiciones de riesgo.** El agente tiene que decidir entre un conjunto de acciones alternativas, de cuyas consecuencias no está seguro, aunque se atreve a asignarles probabilidades subjetivas. También suponemos que el sujeto puede asignar utilidades a las diversas consecuencias posibles. (Si el agente tiene una relación binaria de preferencia entre consecuencias, que satisfaga ciertas condiciones idealizadas razonables, entonces se pueden definir escalas ordinales de utilidad compatibles con ella). Se trata del tipo de situación más estudiado. La técnica formal para resolver este tipo de problemas es la regla de Bayes:

*actúa de tal modo que maximices tu utilidad esperada.*

La utilidad esperada de una acción es la suma ponderada por la probabilidad de las utilidades de sus diversas consecuencias posibles. Si llamamos  $C_1$  a las consecuencias posibles de una determinada acción,  $u$  a la función de utilidad y  $p$  a la de probabilidad, la utilidad esperada de esa acción será

$$\sum u(c_1) \cdot p(c_1)$$

(3) **Decisión bajo condiciones de incertidumbre.** El agente está tan inseguro respecto a las consecuencias que pueda tener cada acción, que ni siquiera se atreve a asignarles probabilidades subjetivas. En este tipo de situaciones (al contrario de lo que ocurría en los anteriores) no hay ninguna regla unívoca y comunmente aceptada para resolver el problema. Hay una pluralidad de reglas propuestas, que corresponden a otros tantos temperamentos o actitudes distintos. Dos reglas famosas son *MAXIMIN* (Actúa de tal manera que maximices la mínima

## EPITOME SOBRE RACIONALIDAD

utilidad -o, equivalentemente, actúa de tal manera que minimices el máximo riesgo-) y *MAXIMAX* (Actúa de tal manera que maximices la máxima utilidad). La primera es una regla de prudencia especialmente atractiva para temperamentos conservadores, pesimistas, que van a lo seguro. La segunda es una regla de audacia, atractiva para los temperamentos aventureros, optimistas, arriesgados, deseosos de jugar fuerte.

La teoría formal de la racionalidad práctica supone que el sujeto sabe lo que quiere (o lo que prefiere) en todas las circunstancias. Eso es una idealización poco realista. Aunque la regla de Bayes es difícilmente atacable, no siempre es aplicable, pues a veces el sujeto no sabe exactamente lo que quiere. La programación lineal y la regla de Bayes formalizan nuestra intuición en la consistencia práctica. Allí donde son aplicables, es imposible entenderlas y no estar de acuerdo con ellas, sin contradecirse en un sentido práctico, es decir, sin reconocer que no queremos aquello que decíamos querer.

### **Teoría material de la racionalidad práctica**

Con frecuencia no está nada claro lo que yo quiero, ni siquiera para mí. El yo es una construcción hipotética a partir de múltiples episodios dispersos de consciencia. En cualquier caso, es la punta apenas entrevista de un iceberg cerebral. Nuestro cerebro, a su vez, es el resultado chapucero de la yuxtaposición de sistemas distintos de procesamiento de la información, sistemas surgidos en épocas diferentes para resolver problemas dispares. A veces parece una empresa mal avenida, donde distintos comités toman decisiones opuestas, lo que puede conducir a la parálisis práctica. El yo con planes y voluntad propia no es algo dado, sino algo construido o por construir.

Cuando nuestros yos descoyuntados llegan a un compromiso, no siempre el resultado es lo que llamaríamos racional. La racionalidad práctica bayesiana o formal se reduce también a la consistencia, y es compatible con cualquier conducta y sistema coherente de fines, por lunático que este sea. (Si la automortificación es uno de mis fines últimos, será formalmente racional que busque cilicios cada vez más lacerantes; lo cual no es óbice para que tanto el fin como el medio sean irracionales en un sentido intuitivo no formal, sino biológico). Para que nuestro sistema de fines merezca ser llamado racional en un sentido material hay que atarlo a algo no formal, hay que anclarlo en alguna realidad material. Una atadura, la más sólida, es la que lo liga a nuestro sistema de fines y necesidades biológicamente dados, heredado genéticamente y plasmado en nuestros diversos sistemas encefálicos. De hecho, nuestro encéfalo ha llegado evolutivamente a ser lo que es como un utensilio fundamentalmente adaptado a satisfacer nuestras necesidades biológicas.

En general (aunque no siempre) el placer y el dolor son sistemas retroalimentativos que nos informan sobre nuestros errores y aciertos prácticos.

La felicidad tiene dos componentes principales: (1) el hedonista, los placeres; y (2) la satisfacción por la consecución de nuestras metas.

En último término, la racionalidad formal se reduce a consistencia y su estudio es una parte de las matemáticas, mientras que la racionalidad material hunde sus raíces en nuestras estructuras biológicas y su estudio está íntimamente ligado al de la biología.

### **Racionalidad parcial y global**

Con frecuencia observamos que algunas personas son muy racionales en ciertas áreas restringidas de sus creencias o de su conducta, y muy irracionales en otras. Por ejemplo, el famoso físico Michael Faraday era tan racional en sus investigaciones como irracional en sus creencias religiosas, que lo llevaron a afiliarse a una oscura secta que lo aterrorizaba y maltrataba.

La racionalidad limitada a ciertos campos o situaciones, con exclusión de los demás, constituye el objeto típico de estudio de la teoría formal de la racionalidad, que en realidad es una teoría de la racionalidad parcial.

La racionalidad práctica parcial, restringida a un ámbito, presupone la racionalidad teórica, al menos en relación a ese ámbito. Para aplicar la programación lineal o regla de Bayes tenemos que indagar previamente las funciones implicadas o asignar probabilidades. Incluso la determinación de utilidades depende de nuestra información acerca del mundo. En efecto, la función de utilidad mide lo mucho o poco deseable que algo es para nosotros. Pero, según lo que pensemos sobre el mundo, desearemos más unas cosas u otras. La información de que una determinada persona tiene el SIDA disminuirá drásticamente nuestro deseo de copular con ella. El inversionista o el agente de bolsa, si se comportan racionalmente en sentido práctico, tratarán de obtener la mejor información disponible acerca de las empresas cuyas acciones pretenden comprar o vender. El enfermo, si quiere curarse, deberá informarse bien al menos sobre su propia enfermedad y su posible terapia.

Si extendemos la racionalidad a la vida entera, entonces apuntamos hacia la racionalidad global, lo que es una empresa tan ambiciosa que ya desborda el ámbito matemático de la teoría de la decisión y se acerca a las sabidurías clásicas y orientales. La racionalidad práctica global es la estrategia de maximización de la felicidad a lo largo de toda nuestra vida. En cualquier caso, la completa racionalidad práctica presupone la completa racionalidad teórica. Sólo en función de un sistema de creencias fiable y de gran alcance acerca del universo podemos orientar la nave de nuestra vida de un modo óptimamente satisfactorio.