

Cognición y Racionalidad

Calanda. 1 Julio 2014
Fernando Broncano

El objetivo es presentar un panorama amplio, sin profundizar, de lo que es el campo de los tratamientos contemporáneos de la racionalidad. Obviamente se trata casi de una lista de encabezamientos, pero la idea es presentar el poliedro de aproximaciones a la racionalidad teniendo en cuenta que en un extremo la teoría filosófica normativa (que se ha tratado sobre todo en contextos de teoría moral y teoría de la normatividad) y en otro extremo la psicología experimental que estudia el comportamiento y las decisiones humanas reales.

1. Introducción al concepto de racionalidad

La idea intuitiva que tenemos del concepto de racionalidad, y que se incorpora a nuestras prácticas discursivas habituales está constituida por un pequeño núcleo de ideas:

- Los hombres son racionales
- Las emociones perturban la racionalidad
- Las tres almas de Platón (racional, afectiva, visceral): sólo la mente lingüística es racional
- Dualismo mente-cuerpo
- Dualismo entre razones internas y circunstancias externas

Una gran parte de esta idea intuitiva se ha incorporado a la Ciencia Cognitiva tradicional y a la psicología experimental cercana a la teoría económica de la elección racional. Frente a esta concepción tradicional, en las dos últimas décadas se ha ido extendiendo un modelo de cognición diferente caracterizado por el rechazo del dualismo y por la importancia de los patrones corporales y por la situacionalidad en el medio. Los marcos teóricos sobre los que se apoya esta concepción son distintos al procedimentalismo de la Inteligencia Artificial clásica y se traducen básicamente en modelos dinámicos sobre variables que no tienen por sí mismos una composición semántica basada en símbolos.

Cognición “encarnada” y racionalidad

Cognición: capacidad de resolver problemas (estímulos externos o internos que se enfrentan a ciertos fines del agente) mediante la información contenida en el organismo

Cognición Encarnada: línea de investigación posterior a la metáfora del ordenador. La información emerge de la complejidad y dinámica de numerosas rutinas de interacción corporal con el medio.

Racionalidad: forma de ejercer la cognición que está sometida a ciertos requisitos entre los que destaca la consistencia y la responsividad a razones

Racionalidad fácil y difícil

Cualquier sistema que resuelva con éxito una clase de problemas que impliquen una explotación efectiva de información puede ser considerado racional. Pero si la

clase de problemas es **estable** y **recurrente** y el sistema está provisto de rutinas efectivas, tendríamos un modo fácil de ser racional
La racionalidad difícil es la que nos permite explicar la alta proporción de éxito ante **problemas nuevos** (nuevos contextos, nuevos objetivos, nuevos obstáculos... con respecto a la trayectoria cognitiva del sistema)

Razones externas vs. internas

Denominamos razones internas a todas aquellas que forman parte de la motivación del agente. Razones externas son las que podrían apoyar o justificar una decisión o acción.

El internismo garantiza

- Motivación
- Diferencia entre moralidad y racionalidad
- Pero hace demasiado fácil ser racional

El externismo garantiza "ser sensible a razones objetivas"

- La controversia conduce a un dilema:
- Toda decisión es trivialmente racional,
- Pero los agentes ideales pueden ser seres racionales,

Dos actitudes filosóficas ante la racionalidad:

Trascendental: : Se considera la racionalidad como una condición a priori:

- Para ser sujeto (kantianos: C. Korsgaard, neo-pragmatistas: R. Brandom)
- Para conseguir autoridad en primera persona
- Para poder interpretarnos mutuamente (D. Davidson)

Naturalista:

- Ser racionales es una adaptación para la supervivencia
- La racionalidad explica el éxito práctico y teórico (suponiendo que las creencias son verdaderas)

La actitud naturalista que proponemos sostiene que la racionalidad deriva de las formas de reconocimiento mutuo en nuestras prácticas cotidianas, por lo que acepta en cierto modo de los trascendentalistas que es el modo en el que nos reconocemos como sujetos, en el que adquirimos una cierta autoridad en primera persona y una condición para interpretarnos mutuamente, pero no acepta que sea una condición a priori del ser humano, al menos no más que cualesquiera otras condiciones derivadas de las trayectorias evolutivas.

2. Racionalidad y normatividad

La normatividad de la racionalidad desde el punto de vista filosófico:

John Broome ha propuesto recientemente una concepción de la racionalidad en el que el criterio mínimo el ajuste o consistencia entre nuestras actitudes. Para él la racionalidad tiene que ver con lo que creemos que estamos obligados y cómo esta creencia nos lleva a intentar o a creer algo. Puede ser discutible si la creencia es en términos de obligación o de permiso, pero la fórmula Broome es muy interesante para captar lo que consideramos *racionalidad mínima*

Se formula de acuerdo a varios requisitos, el más importante es la **enkrasia**:

*La racionalidad te exige que si **crees** que estás obligado a hacer algo entonces **intentas** hacerlo*

Consistencia como:

Si crees que p y si crees que p implica q , entonces crees que q

Si crees que no p entonces no crees que p

Racionalidad instrumental:

Si crees que hacer A es un medio para conseguir B y crees que debes conseguir B entonces intentas A

Racionalidad como logro

Conocimiento y racionalidad son dos términos de éxito o logro:

Conocimiento implica un vínculo robusto de la cognición con la verdad (la información correcta que emplea el agente no se obtiene por azar, sino debido a ciertas características (habilidades, competencia, carácter) del agente

Racionalidad es un logro práctico debido a capacidades de gestión adecuada de las transiciones entre actitudes o entre actitudes y acciones.

Normatividad y Explicación

La racionalidad es **explicativa** en un doble sentido:

- Permite explicar la conducta
- Explica evolutiva y u ontogenéticamente la supervivencia

Es **normativa** en tanto que

- Implica una valoración de ciertos modos de razonar y comportarse
- Nos concierne cuando nos damos cuenta de que está en juego ser racionales

El **fracaso** se puede considerar de dos maneras:

- Llevaría implícita la exclusión de los límites de una comunidad
- O, por comparación, lleva a una pérdida de estatus respecto a un modelo ideal

Teorías de la racionalidad

En ciencias sociales, psicología y ciencia cognitiva encontramos a lo largo del pensamiento contemporáneo teorías que muchas veces están implícitas y no se desarrollan más que como reacciones a ciertos experimentos realizados a lo largo de décadas

- Racionalidad ideal (Piaget, teoría económica)
- Racionalidad acotada (Herbert Simon)
- Racionalidad imperfecta (Kahneman)
- Racionalidad como caja de herramientas (Psicología evolucionista)

3. Racionalidad perfecta y acotada

Racionalidad ideal

Estaría definida por teorías formales del razonamiento correcto

- Que establecen una solución efectiva una vez que el problema se ha formulado en el lenguaje de la teoría y
- Contienen una teoría de la prueba que demuestra que cualquier otra solución es menos eficiente o, en el peor de los casos, inconsistente

Racionalidad ideal

La racionalidad ideal viene definida históricamente por las teorías normativas que dan cuenta de las formas correctas de razonamiento y argumentación

dependientes de la forma de representación del pensamiento: lenguaje natural, lenguajes formales, matemáticas.

- Silogística
- “Critical Thinking” (teoría de la argumentación)
- Lógica formal
- Teoría de la probabilidad
- Teoría de la decisión
- Teoría de juegos
- Teoría de la computabilidad
- Teoría de la prueba

Problemas de la racionalidad ideal

La cuestión en discusión es si estas teorías definen lo que es la racionalidad humana o, por el contrario, representan instrumentos parciales que usamos para calibrar, mejorar o afinar la argumentación o el pensamiento en contextos artificiales, mediados por representaciones formales que suponen medios externos como la escritura. No se juzga aquí la validez de estos instrumentos, sino su capacidad explicativa de la mente humana. Es claro que presentan varios problemas:

- Problemas para la “traducción” del problema al formato representacional de la teoría formal
- Problemas de tiempo de resolución
- Problemas de adecuación de la solución: puede que en situaciones reales, no idealizadas tal como se formulan en la teoría, se desprecien numerosas variables no formalizables que cambiarían el resultado

Ejemplo: en teoría de la decisión y juegos se supone que los sujetos tienen información completa y no tienen sesgos cognitivos ni intrusiones emocionales (envidia)

Teoría de la racionalidad acotada

La teoría de la racionalidad acotada es una restricción de la teoría ortodoxa a ciertos límites de decisión del agente dentro de los cuales actúa racionalmente, aunque fuera de ellos no lo haga. Se establece un umbral de aspiración (satisfacción) dentro del cual la teoría formal/ideal es válida (Herbert Simon) El umbral de satisfacción puede incluir límites a la utilidad o límites a la información (modelos parciales de la realidad) (Johnson-Laird)

Problemas de la racionalidad acotada

- Confía en los instrumentos ideales dentro de los límites impuestos por la aspiración del sujeto (en realidad es una teoría ortodoxa con constricciones)
- Por ello contiene los mismos problemas que la teoría ideal más el problema añadido de la racionalidad de los umbrales de aspiración (¿los umbrales de satisfacción son racionales o irracionales?)

4. Racionalidad imperfecta

Límites de la racionalidad

Los seres humanos tienen serias limitaciones de recursos:

Tiempo de acceso a la información

Comprensión de reglas efectivas de procesamiento de la información

Mutua constitución dinámica entre los sistemas cognitivos y afectivos

El pensamiento rápido

La solución a los límites informacionales que desarrolló la trayectoria evolutiva y cultural de los humanos ha sido la reutilización de rutinas y patrones de lenguaje, pensamiento y acción que se han mostrado exitosos en el pasado:

- Prototipos
- Narrativas (refranes, cuentos populares)
- Rituales
- Rutinas de acción

Dos sistemas de pensamiento:

Los psicólogos experimentales Daniel Kahneman y Amos Tversky, son quienes promovieron la idea de que nuestro pensamiento se produce en dos modos. El sistema 1 explicaría lo que aparentemente son errores, mientras que el sistema 2 estaría regido por la racionalidad ideal.

Sistema 1

- Asociativo
- Holístico
- Automático
- Basado en la experiencia personal
- Actúa basándose
- Estados afectivos
- Propensiones causales y rutinas de acción
- Prototipos

Sistema 2

- Basado en reglas
- Analítico
- Controlado
- Lento, serial
- Exigente en información
- Adquirido por formación
- Constituido por procesos neutrales estadísticos y abstractos

El precio de la rutina

Los límites de información se resuelven usando rutinas rápidas que confían en patrones estadísticos en el pasado. son medios que combinan elementos emocionales con cognitivos

Pero que en contextos de decisión estrictos producen cegueras a veces muy importantes.

En analogía con la percepción, son "ilusiones" cognitivas y emocionales

Problema de las cuatro tarjetas

En el contexto del pensamiento piagetiano, que proponía que la madurez cognitiva coincide con el dominio del pensamiento lógico, y especialmente de la relación condicional “si-entonces” (incluyendo la negación), se desarrollaron experimentos que trataban de poner a prueba la hipótesis. El más conocido es el problema de las cuatro tarjetas, que muestra las dificultades universales para manejar y entender la relación condicional cuando se expone en contextos abstractos, aunque la habilidad mejora notablemente en contextos “situados”, lo que llevaría a la conclusión de que el pensamiento se adapta mejor a regularidades concretas que a reglas abstractas.

Sesgos y heurísticas

Kahneman y Tverski desarrollaron una enorme cantidad de experimentos que mostraban que las operaciones fundamentales de pensamiento estaban “sesgadas” o que se producían bajo la activación de ciertas rutinas o heurísticas. Lo importante de sus experimentos es que se aplican tanto a expertos como a legos. Lo que sigue es una lista representativa, aunque no exhaustiva, de lo que se han ido considerando como imperfecciones o fallos generalizados de la racionalidad.

Sesgos en el juicio

- Heurística de la representatividad
- Heurística de la disponibilidad
- Insensibilidad a la tasa-base
- Anclaje (en la primera información recibida)
- Distorsión retrospectiva (hindsight)

En una ciudad de un millón de habitantes hay cien terroristas escondidos. Se han instalado cámaras de seguridad que detectan el rostro de un terrorista con 99% de aciertos

Falsos negativos: si detecta un terrorista, saltará en un 99% y fallará el 1%.

Falsos positivos: si detecta un no-terrorista, no saltará el 99% de los casos y saltará el 1% de los casos

Una cámara ha señalado un rostro como probable terrorista:

¿Qué probabilidad hay de que esa persona sea terrorista?

Explicación desde el teorema de Bayes:

Imaginemos que el millón de personas pasa por delante de la cámara, dado que hay 100 terroristas: ¿cuántas veces saltará la alarma?

La disparará 99 veces de los 100 terroristas,

Pero también 9.999 de los 999.900 que no son terroristas,

Es decir, 10.098 veces

Luego la probabilidad de que sea una alarma correcta es aproximadamente el 1%

Sesgos en la argumentación

- Incomprensión de la relación condicional (problema de las cuatro tarjetas)
- Dificultades para coordinar teoría y evidencia
- Perseveración en las creencias
- Sesgos de las creencias previas
- Afirmación del consecuente

- Sesgo de la confirmación
- (en general: falacias comunes)

Sesgos en la atribución causal y contrafactual

- Covariación y causación
- Correlaciones ilusorias
- Sesgos en la atención a ciertos tipos de información (disponibilidad)
- Sesgos en la atribución causal (prominencia)
- Sesgos contrafactuales (culpa y contrafácticos)

Sesgos en las decisiones bajo riesgo e incertidumbre

- Aversión al riesgo (Teoría de las perspectivas)
- Sesgos en las preferencias y elecciones
- Violaciones de la transitividad
- Violaciones de la invariancia de la preferencia

Preferencias dependientes de contexto

- Costos invertidos (efecto Concorde)
- Calibración de las capacidades
- Sobreconfianza
- Optimismo injustificado
- Ilusiones de control

Teoría de los “mecanismos” mentales

Jon Elster, en la tradición davidsoniana de la mente dividida (la idea de que la mente es ortodoxamente racional en los planos conscientes, pero con subestructuras que perturban la racionalidad) propuso la idea de la existencia de “mecanismos” o subestructuras que producen una interacción deseos-creencias que distorsiona la teoría ortodoxa de la decisión. Algunos mecanismos conocidos son:

- **Wishful-Thinking:** o distorsión cognitiva producida por la fuerza del deseo de un cierto estado de cosas
- **Uvas verdes:** distorsión motivacional producida por la evidencia de que el objetivo presenta

Sesgos de la identidad:

Conocidos y discutidos desde la antigüedad clásica, están los modos de estructuración del pensamiento que se anclan en el carácter agencial del sujeto y pueden llegar a conformar de manera importante la toma de decisiones

Auto-engaño: se trata de un grave déficit en el autoconocimiento del sujeto.

Tradicionalmente se ha discutido en estos términos

- Imposibilidad del “autoengaño” consciente
- Efecto emocional
- tensión entre el yo ideal y el yo percibido

Akrasia: La debilidad de la voluntad es un déficit volitivo. Se trata de que el sujeto actúa contra lo que considera que debería hacer.

Procastinación: es una forma de akrasia muy usual. Consiste en dejar para el futuro acciones que uno sabe que tiene que hacer. Es un Sesgo en la gestión del

tiempo de la vida. En el mismo tipo están las dificultades en la previsión de estados afectivos futuros que conducen a formas de angustia o euforia irracionales.

Decisiones en grupo

La teoría de juegos estudia la racionalidad dependiente, es decir, la racionalidad de decisiones cuyo resultado depende de las decisiones de otros. En los contextos colectivos, sin embargo existen “distorsiones” producidas por los mismos vínculos que hacen de esos entornos colectividades. Algunos “sesgos” difícilmente tendrían que ser considerados como tales, por ejemplo el altruismo y la actitud cooperativa, o el sentimiento de justicia, a pesar de que la teoría de juegos así lo considere. Sin embargo, algunos sí son importantes:

- Obediencia a la autoridad
- Polarización
- Influencia de las minorías
- Decisiones influidas por el grupo

Racionalidad como caja de herramientas adaptativas

Esta teoría es una formulación evolucionista de la teoría de las heurísticas y de los sesgos. No difiere en los aspectos observacionales (en general, aunque ha presentado numerosas variaciones propias de los experimentos) pero sí en la *valoración* del significado de los sesgos. Se trata de una teoría de inspiración evolucionista que no considera como “imperfecta” la racionalidad observable, sino como un resultado de las trayectorias evolutivas. De acuerdo a los autores que defienden esta teoría, el pensamiento rápido evolucionó en la etapa de formación de la mente humana (Pleistoceno) como un conjunto de rutinas rápidas dirigidas a resolver los problemas básicos de supervivencia en un entorno abierto.

- Preservación del grupo, prototipos y frecuencias (regularidades) fueron los principales roles funcionales de ciertos mecanismos que se habrían estabilizado evolutivamente.
- La mente se diseñó como un multisistema de módulos cognitivos especializados

Problemas de la “caja de herramientas”

- No hay ninguna evidencia de cuáles podrían haber sido los retos evolutivos en el Pleistoceno
- Nada indica que haya un sistema modular masivo en el cerebro humano.
- Por el contrario, todas las pistas hablan de la alta plasticidad y de la capacidad de estabilización cultural de estructuras

5. Racionalidad naturalizada

La concepción naturalizada que propongo hace compatible el carácter normativo que tiene la racionalidad con los datos experimentales y las hipótesis sobre la evolución de la mente. Significa que la racionalidad es una propiedad que se asienta en otras características algunas de especie y otras culturales. Algunas de estas características son constitutivas y otras, más contingentes, forman parte del desarrollo personal del individuo.

Historia natural de la racionalidad

Racionalidad básica:

Delimitada por las affordances.

Implicada en representaciones “pushmi-pullyu” (empuja-tira)

Que permiten al organismo maximizar sus impulsos (protodeseos) dentro del marco de las affordances.

Logra algún grado de robustez y flexibilidad.

Racionalidad informacional:

Explota la información contenida en los estados internos (que pueden ser también patrones y rutinas sensorio-motoras)

Es necesaria para todo agente que sea capaz de grabar un número amplio de hechos pasados sin interés inmediato para el agente.

La consistencia entre es la propiedad más relevante (supone algún medio representacional)

Posibilidades intermedias de algunas operaciones protológicas y no lingüísticas. Proto-inferencias (negación, disyunción, modus ponens y modus tollens)

Auto-regulación

La racionalidad es una característica de un tipo particular de organismos y/o sistemas que están constituidos como un sistema de sistemas de auto-regulación en diferentes modos de ensamblamiento, aunque en último extremo deben incluir una reacción a un modelo de yo en relación con el entorno.

La intencionalidad individual emerge en **entornos poco previsibles** donde se presentan problemas que ni son similares a los ya resueltos ni son completamente nuevos

El cerebro evoluciona en estos ambientes como un **sistema auto-regulativo** que interactúa con el medio por “aprendizaje popperiano”: la hipótesis muere, yo no)

La principal habilidad es la de crear un **modelo virtual** del mundo sobre el que se hacen inferencias

Las otras dos habilidades son el **control** y la **monitorización on line** de las interacciones con el medio

En estos entornos se promueven ciertas capacidades relevantes que son vectores evolutivos, en el sentido de que fueron líneas estables de ventaja adaptativa.

Atención relevante

La racionalidad ocurre en agentes o sistemas que tienen una relación perceptiva con el entorno, es decir, que disponen de sensores y formas de interactuar con las *affordances* del medio, aquellas propiedades que ordenan el medio como un conjunto de posibilidades de acción para el agente. Heidegger ha postulado la idea de cuidado o preocupación como característica del ser humano, y Simone Weil la de atención a lo que importa. Ambas están relacionadas con la racionalidad, en tanto que relacionan al yo con su situación inmediata de forma que organiza tal situación distinguiendo:

- Relevancia: Posibilidades
- Positivas (oportunidades)
- Negativas (obstáculos)

Esquematización de la información

La experiencia solamente es útil en contextos novedosos si es posible esquematizarla para poder desarrollar modelos parciales que por un lado clasifiquen tipos y por otro lado reconozcan casos (Teoría de situaciones: estructuras parsimoniosamente abstractas)

- Interpretación: “Ésta es una de esas”
- Imaginación off-line, basada en imágenes
- Representación de posibilidades alternativas como producto de la combinación de variaciones en el esquema

Lo que los grandes simios pueden lograr

La racionalidad emerge de y sobre características compartidas con numerosas especies, y particularmente con aquellas más próximas en nuestro pasado evolutivo. Las más cercanas, los grandes simios, de los que nos separan unos millones de años (entre 6-9) tienen comportamientos que nos permiten adscribirles capacidades de proto-razonamiento:

- Pensamiento proto-causal
- Proto-negación
- Proto-condicional
- Inferencias elementales sobre la conducta de los otros (que implican regularidades mentales)

Los experimentos sugieren que la racionalidad evolucionó muy pronto en los primates (Tomasello)

El diseño de la mente del homo sapiens

No conocemos todavía cuáles fueron las variaciones genéticas específicas del homo sapiens, pero sí es razonable pensar en tres adaptaciones centrales para el desarrollo de la racionalidad.

Las tres grandes fuerzas que diseñan la mente humana son:

- El lenguaje
- La socialidad
- La técnica

Las tres reutilizan capacidades primates gracias a una nueva capacidad para la recursividad y combinatoria

La “imaginación combinatoria” es muy exigente en recursos informacionales y ello crea problemas de interferencia entre subsistemas.

Desacoplamiento

El invento de la racionalidad es un producto de la la cognición off-line, es decir la génesis de modelos plásticos, combinatorios adaptativos.

Pensamiento relacional (conexión múltiple semántica): la capacidad para incluir en los modelos de mundo información heterogénea produce una explosión de posibilidades de inferencia.

Orígenes de la socialidad:

Desarrollo en entornos de segunda persona:

- Compartición de estados mentales (afectivos, cognitivos, volitivos)
- Yo-Tú
- Desarrollo normativo por referencia al tú
- Actitudes compartidas

Pensar “es como improvisar jazz en una habitación a solas. Es una actividad solitaria, pero el instrumento y la música han sido posibles por la cooperación social”

Capacidad para acciones conjuntas y actitudes “nosotros...”

Microinstituciones (H. G. Mead, Wittgenstein):

- Atención conjunta
- Acción conjunta
- Testimonio

Racionalidad reflexiva en entornos sociales.

Sólo los humanos son capaces de una representación abstracta off-line de una situación que

- Se representan una y la misma situación bajo el punto de vista de diferentes individuos (objetividad intersubjetiva)
- Hacen inferencias auto-reflexivas y sociales recursivas (reflexividad)
- Evalúan los pensamientos de acuerdo a estándares compartidos

Ello indica que el medio social es determinante en la conformación de las operaciones del pensamiento y de los estándares de evaluación de estas operaciones.

Orígenes de la socialidad: Fase 2

El entorno evolutivo de los homínidos condujo a una forma muy particular de cultura, la que está basada en símbolos y sobre todo en convenciones que convierten a las bandas en sociedades:

La cultura crea entornos de desarrollo próximo (Vigotsky, Bajtin) donde se adquieren habilidades que de otro modo no serían posibles por la expresión de las capacidades transmitidas genéticamente

Intencionalidad colectiva (formación de instituciones y sistemas simbólicos compartidos)

Prácticas y significados: la conducta se convierte en acción simbólica

Unbounded Rationality (Racionalidad desacotada)

Si consideramos “racionalidad” como un término de logro, deja de tener sentido la discusión sobre la “verdadera” naturaleza de la racionalidad humana.

- La racionalidad tiene componentes de diseño cerebral, pero también de trayectoria histórica
- La racionalidad se ejerce en nichos cognitivos
- La racionalidad evoluciona
- La racionalidad se configura por cambios que tienen “efecto trinquete”
- Racionalidad y responsabilidad coevolucionan culturalmente

Los nichos culturales humanos

Por lo que el lenguaje, la socialidad y la técnica se constituyen como fuerzas evolutivas es porque crean *nichos culturales* que modelan las propias presiones evolutivas. Los nichos culturales permiten “efectos trinquete”, es decir, asimetrías históricas y evolutivas que producen efectos direccionales. Así podemos entender estas fuerzas como creadoras de tres tipos de nicho:

- Lenguaje como nicho cognitivo
- Cooperación y socialidad

- Capacidades técnicas y entornos artificiales

La hipótesis de los nichos culturales: los miembros de la especie alcanzan logros que naturalmente no conseguirían en ausencia de esos nichos

Las trayectorias históricas de los nichos culturales preservan los logros a través de un sistema social de prácticas

El nicho cognitivo: la ingeniería informacional

Los homínidos no solamente modifican el entorno como otras especies animales (nidos, madrigueras, etc.) con artefactos funcionales, también modifican el entorno llenándolo de artefactos simbólicos que tienen un efecto retroactivo sobre la configuración del cerebro

El más importante por los profundos cambios que induce en el cerebro es el lenguaje

Pero hay otros entornos simbólicos: gestos, pinturas, totems, etc.

El nicho social: la ingeniería de la cooperación

Los homínidos crean estructuras estables de cooperación

En el caso de la especie humana se traduce en convenciones que introducen un orden social donde se reproducen:

- Parentesco
- Ritos
- Mitos

El nicho artificial: artefactos técnicos

Las técnicas humanas son composicionales

- La combinatoria permite una explosión creativa
- Los entornos técnicos crean posibilidades: affordances
- Los entornos definen la cultura material que configura las identidades

En estos entornos la racionalidad evoluciona como sistema de control de la acción en la forma que llamamos agencia: acción intencional comprensible y sujeta a explicaciones en forma de razones. Pensemos en un modelo de niños jugando aislados donde sus acciones son erráticas e imprevisibles y en juegos basados en convenciones que ellos mismos se dan.

6. Esferas de racionalidad

La racionalidad mide la calidad de la agencia

La racionalidad implica habilidades de “descentramiento” del agente, para situarse en una perspectiva sobre sí mismo

Permite al agente identificarse con los motivos propios de la acción.

Propongo considerar la calidad de agencia como un concepto inclusivo de racionalidad.

- La agencia como una relación de determinación entre actitudes (y entre actitudes y acciones).
- Agencia como un vínculo robusto con el mundo (externo e interno)

La agencia divide a los organismos en aquellos capaces de determinar cambios debido a su motivación y aquellos que simplemente reaccionan a las demandas del medio

Requisitos de la calidad de agencia

Robustez: la relación entre motivos y decisiones

- Transciende la mera relación de probabilidad
- Requiere excluir la suerte (es decir, otras causas que las incluidas en el conjunto de motivaciones del agente)

Flexibilidad:

- Correspondencia entre diversidad de medios y un fin
- entre un medio y varios fines
- entre medios y fines (algunos medios devienen fines en el curso de planes de acción)

Valora el cómo el agente se enfrenta a nuevos retos que provienen del mundo externo o de sus fuentes internas.

Mediante al robustez y flexibilidad la agencia capacita al agente como fuente de causalidad en el mundo.

- La condición de robustez aísla el lazo especial de ajuste mente-mundo.
- La flexibilidad extiende el campo de la agencia y así incrementa el ámbito de la actitud intencional hacia el mundo.

Racionalidad y responsabilidad: esferas de racionalidad

La evolución de la racionalidad en los entornos de los nichos culturales crea transformaciones en varios niveles:

- La gestión de la información interna
- La gestión comunicativa de la información bajo lazos de dependencia social
- La gestión de las consecuencias a través del control del medio

Los tres entornos se constituyen también como entornos de responsabilidad:

Esfera del control interno de contenidos

En el nivel personal, la conciencia debe estar constreñida por las propiedades del contenido: holismo, normatividad, flexibilidad de aplicación.

Por ello describimos a los agentes como seres que están sometidos al *Principio de Generalidad*.

Es una restricción que capacita para generalizar atribuciones de predicados..

La racionalidad es una facultad que implica alguna suerte de auto-conocimiento.

Responsabilidad en el entorno interno

Metacognition: control y monitorización de la interacción con el medio

Coimplicación entre *epistemología y racionalidad:* la racionalidad implica una revisión continua de las creencias en la dinámica de la interacción con el entorno

La epistemología implica un reforzamiento de los lazos cognitivos basándose en las capacidades racionales para la comprensión y justificación

Hay una evolución no solamente de los contenidos sino también de los formatos de control

Esfera de racionalidad situada y ampliada:

La calidad de la agencia se mide por la disposición a conseguir todos y sólo los objetivos de la motivación propia.

Esto constituye una suerte de vínculo de ajuste entre la mente y el mundo que insta al agente como causa singular.

La racionalidad implica el ajuste de medios y fines: organizar los medios para que sean adecuados a los fines.
Esto implica que hay una cierta responsabilidad por las consecuencias de la acción.
Los entornos de acción son físicos, sociales y personales

Esfera de racionalidad en el entorno social

La racionalidad es parte de nuestra existencia en dependencia de los otros
Aunque la racionalidad es endógena, la preservación de los lazos sociales exige la capacidad de ser interpretado por otros
La "transparencia" depende de que las razones y motivos de la acción sean accesibles a otros
Junto a la transparencia, el segundo pilar de la racionalidad en el entorno social es la justificación, es decir, la capacidad para dar razones concluyentes de la acción
La racionalidad en el entorno social es pues la responsabilidad por la manifestación pública de nuestras razones

Dos sentidos de responsabilidad

Responsabilidad atributiva:

Depende del nivel personal: se atribuye la decisión a un sujeto (o agente) autónomo

Cuyas decisiones son endógenas y voluntarias

Responsabilidad como "dar cuenta"

La racionalidad implica una extensión de la agencia en el tiempo

De forma que el sujeto se hace cargo de las consecuencias de sus decisiones

En un sentido que puede ser moral o simplemente racional (dependiente de las normas y convenciones sociales)

Racionalidad trágica

La racionalidad es un rasgo humano que valoramos, pero no es un rasgo permanente

No hay razones incondicionales para ser racional

La vida humana está sometida a constricciones trágicas:

Los héroes de la tragedia clásica provocan ellos mismos su propia destrucción

Las consecuencias no queridas de las acciones pueden hacer que las acciones más racionales produzcan trayectorias irracionales.

El esfuerzo de autoconocimiento puede producir mala fe y sumisión a modelos ideales del yo

7. Apéndice

En este apéndice se exponen en boceto algunas ideas filosóficas que están en el trasfondo de lo anterior, pero que no han sido tratadas. Considero aquí temas que han sido objeto de discusión en contextos de teoría de la acción, teoría de la normatividad y metaética, y que son centrales en la idea de racionalidad, fundamentalmente en lo que respecta al problema de si hay algo así como una propiedad de racionalidad o solamente una sensibilidad o responsabilidad a razones. Por otro lado se plantea también la fuerza de la racionalidad: si consiste en un conjunto de obligaciones o en un conjunto de licencias o permisos fundados para una decisión o una inferencia. Igualmente, se expande n algunas consideraciones sobre la primera esfera de responsabilidad, la de la gestión de contenidos.

Tal como planteamos aquí la idea de racionalidad, se sitúa en un dominio que llamamos el nivel personal, más allá de la pura constitución físico-mental, en un entorno de relaciones sociales reactivas a la propia conducta en forma de interpretaciones y reconocimientos.

La racionalidad del filósofo

Un componente usual del concepto de racionalidad en filosofía es ***ser responsivo a razones***

Por razones entendemos o bien estados mentales (creencias, deseos, motivaciones, planes, valores) o bien hechos que ***respaldan una actitud*** (mental): creencias, intenciones o ***una acción***.

Por sensibilidad a razones entendemos el tomar decisiones basadas en la ***conciencia de las razones***.

El nivel personal

La división personal/subpersonal es más interesante que la dicotomía conceptual/no-conceptual en lo que respecta al problema de la racionalidad, una dicotomía en la que generalmente se han situado los filósofos de origen kantiano (McDowell, Korsgaard)

La racionalidad depende tanto de los niveles sub-personales como de las actitudes personales. En los niveles subpersonales encontramos explicación de las constricciones cognitivas con las que opera el nivel personal. En el nivel personal se producen interacciones que pueden incorporar artefactos de “mente extendida” como instrumentos de razonamiento (escritura, cálculo, etc.)

Bilgrami: hay un nivel meta-cognitivo necesario para la agencia. No hay posibilidad de agencia sin algún nivel básico de autoconocimiento, de saberse implicado en las decisiones que se están tomando. Un nivel primero de implicación es el de activación de emociones reactivas que sitúan al agente (como tal) en el curso de los hechos.

Estas reacciones evaluativo-cognitivas convierten los meros estados informacionales en agencias causales.

Al pasar de lo sub-personal a lo personal el agente se convierte en un todo que se confronta con el mundo en forma de relaciones medio-fin.

Esto permite al agente emprender planes que involucran su ser completo (distinto a la mera satisfacción de deseos) en forma de compromisos y trayectorias de vida. La racionalidad estriba menos en el formato representacional que en la capacidad de situarse a sí mismo en un mundo de causas.

¿Qué constituye una razón?

Tradicción inferencialista: solamente pueden constituir razones relaciones inferenciales.

Argumento circular: para aceptar algo como dependiente inferencialmente de otra cosa el agente ya debe suponerse racional,

Teoría de la licencia (entitlement): lo que constituye una razón es todo aquello que licencia al agente para hacer una transición entre actitudes.

La licencia es ella misma una transición a un nivel personal que sostiene un punto normativo:

Al atender a algo en la percepción estamos licenciados para formar una creencia, al notar un deseo estamos licenciados para formar un motivo, etc.

Las relaciones inferenciales son licencias que tienen los agentes que dominan las relaciones conceptuales, sin embargo hay otras fuentes de licencias como las actitudes reactivas, los estados perceptivos, etc.

Las fuentes de autoridad que constituyen la licencia, en el caso de la racionalidad, tienen su sede en cómo reconoce una persona la autoridad de algo como una razón para algo.

Esto implica una relación compleja entre una fuente objetiva de autoridad y un nivel personal que la reconoce como un resorte con autoridad para una transición.

Responsividad a razones

Una autoridad objetiva por sí misma no constituye una licencia sin el reconocimiento del nivel personal.

Solamente una fuente endógena de motivación constituye esta autoridad de licencia

Solo una mente bien sintonizada que responda a razones objetivas puede caracterizarse como racional.

El agente racional tiene una capacidad de sumisión receptiva a las fuentes de autoridad.

En la racionalidad práctica, sólo el reconocimiento de la autoridad de los propios estados es la fuente de autoridad.

El mecanismo de generación endógena de transiciones racionales debe ser necesariamente meta-cognitivo.

El agente anticipa una decisión que va a tomar al tiempo que se identifica con la decisión que anticipa que va a tomar

Es un mecanismo de feedback que tiene tanto un carácter epistémico como motivacional.

Ser racional es así tener una suerte de carácter virtuoso que explica por qué el mecanismo funciona adecuadamente.

Relevancia del formato representacional:

La información es un bien valioso en términos metabólicos (constricciones de los flujos de energía)

El diseño del sistema está constreñido por el objetivo de salvar la información,

El nivel del yo como lugar natural de la racionalidad nos permite explicar el lugar de la referencia en las transiciones racionales.

La preservación de la referencia garantiza la reutilización de la información

Preservación de tokens (referencia individual)

Preservación de tipos (conceptos y clases).

La referencia funciona como un puntero que unifica muchas representaciones.

El nivel personal es la sede de este ensamblamiento.

Racionalidad y "licencias" (entitlements)

En la racionalidad son más importante las "licencias" para transiciones de estados mentales que las inferencias para provocar cambios o acciones del sistema entero

Que puede suponer, como condición necesaria un cierto nivel de conciencia suficiente para gestionar el sistema en su totalidad (el yo)

Aunque previamente haya un gran número de transiciones cognitivas

La racionalidad tienen que ver con la autoridad sobre sí mismo (autoridad en primera persona)

Que nace del desacoplamiento del yo y el mundo
Y tiene que ver menos con el formato en el que se producen las transiciones.
El nivel conceptual sería el mejor y más potente de los medios para servir al objetivo primario de gestionar el yo.

Racionalidad centrada en el agente

La racionalidad es un rasgo del *carácter del agente* que se manifiesta en cómo toma las decisiones

La normatividad nace de *las disposiciones fiables a formar las razones* adecuadas
Sensibilidad o responsividad a razones: motivos que un agente fiablemente racional forma como resorte de sus actitudes o acciones.

Racionalidad como autonomía

Ser racional es algo más que obedecer a un deseo (Hume-Williams)

Tendríamos que preguntarnos por el proceso completo de formación de decisiones, por su adecuación, etc., para contrastar la fiabilidad racional

La racionalidad podría ser entendida como un tipo de virtud de una mente capaz de tomar decisiones en condiciones normales.

Del mismo modo, una decisión racional sería el logro de tal mente racional.

Racionalidad como autonomía personal

Una condición suficiente de racionalidad es que el agente tenga una razón para la decisión.

Tener una razón significa un tipo particular de auto-determinación causal.

Al tener una razón el agente se transforma en una singularidad causal en las cadenas causales de la naturaleza.

La autonomía personal sería la fuente de la normatividad personal.

Racionalidad y auto-determinación se interdefinen.

La perspectiva naturalista:

La autonomía se da en grados:

Existe un espectro desde los animales (y agentes autónomos) con cierto grado de conciencia y capacidad para tomar decisiones hasta las políticas de acción auto-reflexivas.

Formar una razón no implica necesariamente formarla *conceptualmente*.

Podría darse también la posibilidad de una perspectiva continua sobre cómo emerge la racionalidad en la naturaleza.

De otra forma, el “espacio de las razones” sería un milagro evolutivo.

Racionalidad como desacoplamiento normativo

La racionalidad es una propiedad que poseen los organismos o sistemas que tienen algún grado de desacoplamiento de la realidad.

Estar desacoplado significa vivir en una “segunda naturaleza” realizada por un medio rico en información (interna/externa)

El desacoplamiento es un tipo estructural de propiedad que explica cómo las decisiones y acciones tienen un origen endógeno basado en actitudes hacia el mundo

El desacoplamiento abre la posibilidad de errores debidos a las actitudes del agente.

El puzzle de Kripke como mala suerte epistémica

Pierre es incapaz de captar la inconsistencia de dos de sus creencias: “Londres is jolie”, una creencia que aprendió como niño cuando vivía en Paris, y “London is ugly”, que ha aprendido viviendo en Londres:

Debemos acusar a Pierre de irracionalidad? Si no lo hiciéramos estaríamos violando el criterio de consistencia; si lo hacemos estamos siendo injustos con la trayectoria cognitiva de vida de Pierre.

La acusación internista: una teoría externista del contenido pone en peligro el autoconocimiento en tanto que condición aceptada de autoridad en primera persona.

Dado que el contenido se fija simultáneamente por las relaciones causales internas y por los estados internos, el agente no sabría qué es lo que está pensando.

En la primera esfera de responsabilidad

Una posible solución representacionalista: postular un desconexión de archivos no normal (mala suerte informacional)

O bien: el puzzle apoya la idea de que existe una separación de niveles personal/sub-personal

La desconexión sub-personal es compatible con un postulado de consistencia en el nivel personal,

Que incluye como condición necesaria la conciencia de que London y Londres son términos co-referentes

Esto nos lleva a aceptar un primer nivel normativo: la racionalidad puede haber emergido como una propiedad normativa de la gestión de contenidos.

Tendríamos así la base de la esfera primera de responsabilidad .

No cartesianismo

No deberíamos confundir el teatro cartesiano de la conciencia con un modelo de autoridad en primera persona.

La responsabilidad en primera persona no implica la privacidad de contenidos.

La autoridad de primera persona supone solamente que debemos aceptar la asimetría entre primera y tercera persona.

Esta asimetría no se basa en el acceso privilegiado sino en la diferencia de las esferas de responsabilidad.

El contexto de la práctica es el lugar adecuado para discutir esta asimetría.

En el caso de las transiciones racionales hay una diferencia análoga en la responsabilidad de la decisión y el carácter normativo del contenido que obedece a consideraciones publicitarias

Responsabilidad en la gestión de contenidos

Es la responsabilidad de gestionar adecuadamente los archivos informacionales que se activan en las transiciones. Así, por ejemplo,

Un requisito de racionalidad tal como el *modus ponens* nos licencia (obliga) a que, si entendemos que $p \rightarrow q$, y estamos licenciados (obligados) a creer que p , entonces tenemos que creer que q .

Somos responsables por ello en tanto que conozcamos la norma y sepamos que se aplica a p y q . Está bajo nuestra responsabilidad el responder a las demandas de la norma.

¿Emerge la responsabilidad de un estado de segundo orden o simplemente de un rasgo de nuestro carácter el reconocer la licencia para hacer la transición? Estar implicado en prácticas inferenciales de seguir reglas no impide (de hecho supone) también tener responsabilidades prácticas y epistémicas.

El carácter público de las reglas explica la objetividad de las transiciones. Pero la diferencia entre seguir una regla y hacer transiciones porque uno se somete a la regla es la frontera que señala el ámbito de la responsabilidad en primera persona y el reino de la racionalidad.

Sin embargo este reconocimiento puede darse como una habilidad encajada en los procesos atencionales a los propios estados.

Ser racional sería una suerte de destreza en la gestión de contenidos.

La diferencia entre esta posición y el inferencialismo de Brandom en este punto es que el reconocimiento de la soberanía de las reglas no tienen por qué tomar la forma de un compromiso donde dar y pedir razones (tampoco lo impide) sino con un compromiso con el contenido y las propias actitudes. Toma la forma de una identificación con los propios estados mentales y las licencias de transición.

8. Referencias (para ampliar algunos de los puntos de la presentación)

- Brandom, R. (1994) *Hacerlo explícito. Razonamiento, representación y compromise discursive*. Barcelona: Herder (2013)
- Broome, John (2013) *Rationality through reasoning*. Oxford: Wiley Blackwell
- Buss, David M. (2008) *Evolutionary Psychology. The new science of the mind*, Needham Heights, MA: Allyn & Bacon.
- Gigerenzen, Gerd (2008) *Rationality for mortals: How people cope with uncertainty*, Oxford: OUP
- Gilovich, T., D Griffin, D. Kahneman (Eds) (2002) *Heuristic and biases: The psychology of intuitive judgment*, Cambridge: Cambridge University Press
- Hardman, David (2011) *Judgment and decision. Psychological perspectives*. Oxford: Blackwell
- Kahneman, Daniel (2013) *Pensar rápido, pensar despacio*. Madrid: DeBolsillo
- Skorupski, John (2010) *The Domain of reasons*, Oxford: OUP
- Sterelny, Kim (2004) *Thought in a hostile world*, Oxford: Blackwell
- Stich, Stephen (1990) *The fragmentation of reason*. Cambridge (MA): MIT Press
- Tomasello, Michael (2014) *A natural history of human thinking* Cambridge (MA): Harvard University Press
- Walton, D.N, (1996) *Informal logic: a handbook for critical argumentation*. Cambridge: CUP