



INVITACIÓ
A LA TEORIA
ECONÒMICA

Edició a càrrec de Ramon Marimon i Xavier Calsamiglia

Ariel

CAPÍTOL 12

ALGUNS MODELS DE COMPORTAMENT RACIONAL EN ECONOMIA*

per SALVADOR BARBERA

BARBERÀ, Salvador (Barcelona, 1946). Començà els estudis de llicenciatura en Ciències Econòmiques a Barcelona l'any 1964, i els acabà a Bilbao l'any 1968. Entre el 1970 i el 1975 realitzà estudis de posgrau a la Universitat de Northwestern, on obté els graus de *Master of Arts* (1972) i de *Doctor* (Ph. D., 1975). Durant els deu anys següents és professor a les universitats Autònoma de Madrid (1975-77) i a la del País Basc (1977-86), i professor visitant a la Stanford Graduate School of Business (1984-85). Des del 1986 és novament a Catalunya, com a catedràtic a l'Universitat Autònoma de Barcelona. A més dels temes de decisió individual, als quals fa referència aquest article, ha treballat sobretot en teoria de l'elecció social i aplicacions de teoria de jocs a l'anàlisi d'incentius, temes sobre els quals ha publicat, entre altres, els articles «The manipulation of social choice mechanisms that do not leave "too much" to chance», *Econometrica*, 1977, «Implementability via protective equilibria», (amb Bhaskar Dutta) *Journal of Mathematical Economics*, 1982, i «Voting by Committees», (amb H. Sonnenschein i L. Zhou), *Econometrica*, 1991. Alguns dels seus treballs a revistes espanyoles són panoràmiques de diversos aspectes en economia normativa: «Justicia, Equidad y Eficiencia», *Hacienda pública española*, 1978; «Teoría de juegos y métodos de votación», *Cuadernos económicos de ICE*, 1989. És director de la *Revista española de Economía*, i membre del consell de redacció d'*Econometrica*, *Social Choice & Welfare*, *Mathematical Social Sciences*, i *Economic Theory*. El 1988 va ser elegit Fellow de l'Econometric Society.

1. Introducció

La racionalitat dels agents sovint es pren com a hipòtesi de partida de l'anàlisi econòmica. Una possible raó per adoptar-la és que d'ella se'n deriven regularitats en els comportaments individuals, algunes de les quals es preserven en l'agregat. Des d'aquest punt de vista, constitueix una hipòtesi senzilla, unificadora i, en principi, contrastable, des de la qual

* El treball de Salvador Barberà ha estat realitzat pels projectes PB 89-0294 de la DGICYT del Ministerio de Educación y Ciencia.

és possible explicar molts fenòmens econòmics. De totes maneres, ni un comportament agregat té perquè heretar les regularitats d'un comportament individual ni és necessari que sigui regular el comportament individual perquè ho sigui l'agregat. Raonaments de tipus estadístic podrien, potser, justificar comportaments col·lectius perfectament estables i subjectes de teorització sense necessitat de suposar que procedeixen de decisions individuals racionals. No es tracta així doncs d'una hipòtesi imprescindible per a tota teoria positiva, però és atractiva per la seva simplicitat i la seva capacitat productiva. A més, la hipòtesi de racionalitat individual apleix altres papers analítics. L'economista desitja comprendre la realitat, i també modificar-la. Dificilment podríem parlar dels objectius dels subjectes econòmics si aquests no fossin racionals, i poc es podrien valorar les conseqüències socials d'una determinada política si els agents individuals no sabessin què volen. Suposar que cada agent és capaç de pronunciar-se sobre les diferents opcions a les quals s'enfronta sembla essencial per jutjar polítiques alternatives: la hipòtesi de racionalitat apareix com a base per formular criteris normatius. Finalment, hem de tenir en compte que tota mesura de política econòmica incideix sobre el sistema mitjançant els seus agents, i que les reaccions d'aquests influenciaran les possibles conseqüències de la mesura en qüestió. No es pot pensar correctament en l'impacte d'una política sense tenir en compte aquestes reaccions, i per preveure les seves característiques, resulta d'utilitat suposar que els agents són racionals. Per tant, a més de ser una hipòtesi unificadora i de base dels criteris normatius, la hipòtesi de racionalitat és també una restricció a tenir en compte en el disseny de sistemes i mesures econòmics.

En haver destacat fins ara l'interès que té el fet de suposar que els agents són racionals, aquestes notes estaran dedicades a palesar que la racionalitat no és un concepte perfectament demarcat, sinó que sota aquesta denominació s'hi amaguen molts significats diferents. Això no tan sols s'esdevé amb el concepte de racionalitat, sinó amb gairebé qualsevol de les paraules clau que utilitzem en parlar d'economia: també l'eficiència, la justícia, la desigualtat, l'equitat, la pobresa són termes que tenen multitud de significats. I no es tracta tampoc d'escampar la confusió insistint en això, sinó que pretenem assenyalar-ho per entendre les seves causes, que no són altres que la riquesa de contingut del terme, i la varietat de situacions a les quals s'aplica. Per exemple: si diem que ser racional és actuar de la forma més adequada donades les circumstàncies, queda per definir què entenem per adequat, i quin conjunt de circumstàncies contemplem en cada cas. Només en intentar de concretar, precisament, aquesta vaga i discutible definició, haurem generat ja multitud d'accepcions per a aquest terme.

Així doncs, aquest treball es proposa explicar amb un cert detall com s'ha concretat la idea de racionalitat en alguns models utilitzats pels economistes. Es comença pel cas més elemental, en què un agent aïllat pot escollir entre un nombre finit d'alternatives finites, i s'examinen diferents definicions de racionalitat en aquest context. Tot partint d'aquest marc, es compliquen les descripcions dels objectes a escollir i de les circumstàncies de l'elecció, tot intentant en cada cas justificar la connexió entre el model formal adoptat i les seves interpretacions. A banda de l'interès que pugui tenir cada model, i del possible contingut informatiu d'aquest repàs parcial a la literatura, aquest treball vol emfasitzar la tensió entre el llenguatge ordinari i el llenguatge formalitzat, afirmant l'interès de mantenir l'ús de l'un i l'altre en la nostra activitat com a economistes. No podem renunciar a la riquesa de significats que inclouen els grans termes en la seva utilització habitual, ni tampoc al rigor ni a la claredat del llenguatge formalitzat: per beneficiar-se de tots dos, no queda cap altre remei que entendre l'un i l'altre, interpretar correctament cada model en la seva vàlua, i comprendre el seu abast i les seves limitacions.

El treball no té cap pretensió exhaustiva i la selecció de models correspon als gustos

personals de l'autor. Espero, però, que serveixi per il·lustrar el punt de vista general expressat sobre la interacció entre llenguatge formal i interpretacions. Tan sols es consideren situacions en què un agent individual adopta decisions les conseqüències de les quals són ben definides, tot i no ser-ho necessàriament amb una total certesa, sense referència a les accions d'altres agents. S'ha de reconèixer que en la majoria dels casos, les conseqüències de les accions d'un individu depenen de les adoptades per altres, amb els quals interacciona directament o indirectament mitjançant el sistema econòmic. La formulació d'una noció adequada de racionalitat per a aquests casos exigeix l'ús de models i d'un instrumental més poderosos que els utilitzats en aquest treball i, d'això, en parlaré breument en l'apartat de conclusions. Malgrat tot, els models aquí analitzats són un punt de partida imprescindible per a estudis més ambiciosos: poc podríem entendre què és un comportament racional en situacions complexes si prèviament no fóssim capaços de definir-lo en altres de més senzilles. Partirem, doncs, d'un model elemental, i introduïrem complicacions i matisos a partir d'ell, dins els límits esmentats.

2. El model bàsic

Considerem una situació altament idealitzada, on *un sol agent* es troba en condicions de decidir, per sí mateix, entre diverses opcions mútuament excloents x, y, z, \dots , de les quals en direm les *alternatives*.

Per descriure el comportament d'aquest agent és tradicional recórrer a dues formalitzacions, diferents però complementàries. D'una banda, podem fer-ne una *descripció externa*, tot indicant quins serien els resultats observables de la seva actuació. Aquesta descripció externa vindria donada per una *funció d'elecció*, que indica què escolliria l'agent segons quines fossin, entre totes les alternatives, les accessibles o factibles en un determinat moment. Formalment, si A es el conjunt de totes les alternatives, i \mathcal{A} el conjunt de tots els subconjunts no buits d'alternatives, una funció d'elecció serà de la forma $e: \mathcal{A} \rightarrow A$, amb la propietat que, per a tot subconjunt $B \subseteq A$, $e(B) \subseteq B$. Si $e(B)$ té un sol element, ens indica l'alternativa, dins de B , que l'agent escolliria si les úniques opcions obertes fossin les de B . Si $e(B)$ té més d'un element, considerarem que l'agent pot a l'últim triar-ne un qualsevol dins de $e(B)$.

Així doncs, una funció d'elecció $e(\cdot)$ ens descriu el comportament de l'agent sense explicar ni com, ni per què, tria el que tria. Com podem dir si aquest comportament és racional o irracional? Una forma seria entrar en la descripció explícita de quines són les alternatives, i prendre partit: per exemple, si x representa conduir per la dreta en una autopista, i y conduir per l'esquerra, serem molts els disposats a dir, sense més discussió, que y és irracional. Però aquest mateix exemple ens mostra la relativitat del concepte: abans de contestar, més ens val preguntar si l'agent es troba a Catalunya o a Anglaterra! Si, un cop esclarit això, podem determinar quin acte és racional és perquè es tracta d'una situació extrema, on implícitament se suposa que tothom, entenent bé la situació, coincidiria en no voler tenir accidents. En la majoria dels casos, però —i és d'aquests dels que ens ocuparem—, conèixer quines són les alternatives no ens ajuda a definir què és racional, ja que poden haver-hi diferències d'opinió: si x és viure en una gran ciutat, y en un poble, i z enmig del bosc, qualsevol decisió entre les tres, o entre una parella d'alternatives, pot ser perfectament racional. Depèn de les preferències de cadascú. Podem diferenciar, en aquest cas, entre un agent racional i un que no ho és? La distinció que proposarem estarà lligada a la coherència interna de les decisions.

Suposem que l'agent, quan pot triar entre les tres opcions, decideix en favor de z , viure al bosc. Suposem que l'opció y —viure al poble— deixés de ésser possible, i que això el

porta a triar x —la ciutat— en lloc de seguir al bosc. Un comportament així no admet certs tipus d'explicació. No permet, en particular, pensar que l'agent té preferències estables sobre les alternatives i que tria la millor entre les disponibles. Perquè si fos així, la seva elecció de z quan totes les opcions són al seu abast voldria dir que z és millor que x , i així doncs que hauria de escollir encara z quan y ja no és possible. No serà, doncs, racional, si creiem que el model de decisió implícit en l'argument anterior representa una forma racional de fer. L'estratègia relativista que descriurem en aquest apartat, com a punt de partida per presentar diferents conceptes de comportament racional, consisteix a adoptar models que representin una *descripció interna* del comportament de l'agent que sigui indicativa del tipus de procés de presa de decisions que nosaltres, exògenament, volguem identificar com a racional. I el model més tradicional consisteix a distingir entre dos aspectes: les preferències de l'agent, que li permeten ordenar les diferents alternatives, i el criteri de decisió pròpiament dit, que consisteix a triar aquelles alternatives més preferides entre les possibles en cada situació. Formalment, les preferències de l'agent es representen per una relació binària R , completa i transitiva sobre el conjunt d'alternatives interpretada de manera que, per a $x, y \in A$, xRy vol dir que x és almenys tan bo com y . I se suposa que, per a cada conjunt $B \subseteq A$, l'agent escollirà les alternatives maximals $C(R, B) = \{x \in B \mid xRy \text{ per a tot } y \in B\}$, és a dir, aquelles que són si més no tan bones com qualsevol altra possible.

Si identifiquem el procés intern de decisió amb una descripció de comportament racional, ser racional voldria dir a) tenir preferències completes i transitives sobre les alternatives, i b) escollir el millor possible en relació a les preferències en qüestió. Atés que moltes preferències diferents són possibles sobre un mateix conjunt d'alternatives, molts comportaments poden ser racionals. Però no tots! Podem observar que, un cop fixades unes preferències, el criteri esmentat genera una funció d'elecció tal com l'hem descrit abans: ens diu un conjunt escollit per a cada subconjunt no buit d'alternatives. Inversament, direm que una *funció d'elecció* $e(\cdot)$ és *racionalitzable* si existeixen unes preferències R sobre les alternatives tals que, per a tot subconjunt no buit d'aquestes, l'escollit per la funció coincideixi justament amb les millors alternatives possibles segons les preferències. És a dir, si $e(b) = C(R, B)$ per a tot $B \subseteq A$.

De fet, hem canviat una mica la terminologia, tot suavitzant-la d'acord amb la posició relativista que representa. Més que dir si un agent és o no és racional, i en termes absoluts, parlarem de si el seu comportament és o no racionalitzable respecte a un model concret d'elecció. En els termes del model inicial proposat, si una funció d'elecció és racionalitzable voldrà dir que representa el comportament d'un agent *com si* aquest triés el millor possible, guiat per una relació binària completa i transitiva. De fet, que hagi pres o no les seves decisions tot seguint aquest procés o un altre queda fora de la nostra modelització.

3. Racionalitat individual. Variants en el criteri de coherència

En l'apartat anterior el model proposat per representar la racionalitat suposava que el comportament maximitzador de l'agent era guiat per una relació binària, completa i transitiva, definida sobre el conjunt d'alternatives. Diferents modificacions d'aquest model tradicional donen lloc a diferents definicions implícites de comportament racional. Examinem més de prop les hipòtesis emprades, les seves conseqüències i possibles relaxaments.

a) Hem descrit les preferències dels agents com a relacions binàries. Essencialment, això vol dir que donem sentit a afirmacions de preferència entre dues alternatives, del tipus

« x és almenys tan bo com y », sense referència a cap altra opció. Això dona estructura al model. Però hem suposat, a més, que les preferències són transitives i que, per tant, un agent que consideri « x almenys tan bo com y », i « y almenys tan bo com z », sempre creurà que « x és almenys tan bo com z ». La hipòtesi de transitivitat reforça molt l'estructura del model, més enllà del caràcter binari de les relacions de preferència, i restringeix considerablement el nombre de comportaments —de funcions d'elecció— que són racionalitzables. La transitivitat s'identifica, en gran mesura, amb la «coherència» de l'agent. Hi ha, però, restriccions més febles que es poden imposar a una relació binària, a cadascuna de les quals correspondria una concepció de «coherència» també més feble. Per donar-ne dos exemples convé definir les relacions de preferència estricta, P , i d'indiferència, I , associades a una relació binària R . Definim la relació I , «és indiferent a», de forma que xIy sempre i quan xRy i yRx . És a dir: x és indiferent a y si x és tan bo com y , i a la vegada y és tan bo com x . La relació de preferència estricta, «és millor que», es defineix naturalment així: xPy sempre i quan x sigui tan bo com y , mentre que y no sigui tan bo com x .

Una relació binària pot exhibir graus ben forts d'«incoherència», com és el cas de les *relacions cícliques*. Una relació R és cíclica si existeixen alternatives $x_0, x_1, x_2 \dots x_i$ per a les quals $x_0Px_1, x_1Px_2 \dots$ i x_iPx_0 . Per concretar, suposem que el cicle es dona entre tres alternatives x, y i z : és a dir, que $xPyPzPx$. Si la interpretem com a expressió d'unes preferències, aquesta relació representaria un agent el qual tot podent triar entre dues alternatives cada cop, tindria sempre una elecció justificable. Però podríem posar en dubte la coherència del seu discurs quan ens digués que prefereix x a y , i y a z , però que z és millor que x ! Un test addicional a un agent amb preferències com aquestes seria demanar-li què triaria, entre x, y i z , considerades alhora. No hi ha cap alternativa que sigui millor que totes les altres: cap elecció no seria racionalitzable, tret de no triar res, segons aquestes preferències. Així doncs hi ha dificultats per interpretar el comportament d'un agent amb preferències cícliques com un comportament racional.

Hi ha graus de «coherència» intermedis entre la transitivitat i la ciclicitat, formalitzables mitjançant altres propietats de les relacions binàries. Una d'aquestes propietats és la *quasi-transitivitat*: una relació R és quasi-transitiva si la relació estricta P associada és transitiva. Observem que aquesta relació té sentit perquè, si bé la transitivitat de R implica automàticament la de P i la de I , pot ésser que P sigui transitiva sense que la indiferència ho sigui. Si bé la definició és tècnica, pot ser útil per descriure preferències en casos en què les característiques de les alternatives varien imperceptiblement i gradual. Així, un agent que vulgui expressar preferències sobre diferents pressupostos del Ministeri d'Educació, i pensi que com més grans millor, però que no està en condicions de distingir entre aquells que només són diferents per unes poques pessetes, pot declarar-se indiferent entre cada parell de pressupostos successius dins una llista x_0, x_1, \dots, x_T , tals que $x_i = x_{i-1} + 1$ mentre que per a T prou gran considerarà clarament millor x_T que x_0 . Les seves preferències no seran transitives, però sí quasi-transitives.

Un segon relaxament de la transitivitat és la *aciclicitat*. Una relació R és acíclica si no té cicles —és a dir, si per a tota cadena d'alternatives x_1, x_2, \dots, x_n , tenim que $x_1Px_2, x_2Px_3, \dots, x_{n-1}Px_n$ implica que no es compleix x_nPx_1 (la qual cosa donaria lloc a un cicle)— i que així doncs x_1Rx_n . L'aciclicitat és, doncs, un altre requisit de «coherència», més feble que la transitivitat o que la quasi-transitivitat, però suficient per evitar algunes de les dificultats que suposaria declarar «racional» un agent amb preferències cícliques. Concretament, l'aciclicitat és la condició mínima, necessària i suficient, per garantir que els conjunts maximals (sobre conjunts finits d'alternatives) mai no seran buits.

Amb aquestes definicions, podem ampliar i relativitzar la definició de racionalitat considerada abans. Si reservem el terme «racionalitzable» a l'existència d'una relació binària que generi les eleccions apropiades, i el qualifiquem segons les propietats de la relació en qüestió, podem parlar alternativament de funcions d'elecció que siguin «quasi-transitivament racionalitzables» o «acíclicament racionalitzables», o bé (corresponent al model de la secció anterior) «transitivament racionalitzables».

b) Relaxar les condicions de coherència d'una relació binària que volguem associar amb la racionalitat és una forma d'anar obrint les possibles significacions d'aquest terme. Un altre enfocament més radical, consistiria en la renúncia a identificar racionalitat amb relació binària —a deixar de parlar d'una descripció interna del procés de decisió basat en preferències—, i associar-la exclusivament amb propietats de coherència sobre la funció d'elecció. Els models d'elecció basats en preferències imposen restriccions sobre les funcions d'elecció; i aquestes restriccions es poden «factoritzar», poden subdividir-se en altres més senzilles, que podrien exigir-se directament de les funcions. N'esmentarem una: la independència de subdivisions del procés (*path independence*). Suposem que hem de triar entre els elements d'un conjunt A d'acord amb una funció d'elecció. Podem fer-ho directament, i triar $e(A)$. Però, també podríem dividir A en conjunts disjunts A_1, \dots, A_k , triar el millor de cada un d'ells, i triar després entre els escollits. És a dir, podríem buscar $e(e(A_1) \cup e(A_2) \dots \cup e(A_k))$. O triar, per a una partició diferent B_1, B_2, B_3, B_4 , els millors elements de B_1 , afegir-los a B_2 i triar de nou, afegir l'escollit a B_3 i tornar a triar, fins a calcular $e(e(e(e(B_1) \cup B_2) \cup B_3) \cup B_4)$. Si per qualsevol d'aquests mètodes sempre arribem a la mateixa elecció, direm que la funció és independent de subdivisions del procés. Les etapes en què, arbitràriament, s'ha pogut subdividir la tria de l'element escollit, en un procés intern no descrit per la funció $e(\cdot)$, no són rellevants. Si una funció d'elecció té aquesta propietat, podem pensar en fer-la operativa mitjançant algorismes de busca sistemàtica, tot subdividint decisions complexes, etc., aspectes que aquí només apuntem. En tot cas, la quasi-transitivitat està íntimament lligada a l'independència de subdivisions del procés: una funció d'elecció racionalitzable és independent en aquest sentit sempre i quan sigui quasi-transitiva la preferència que la racionalitza. Però aquesta propietat d'independència és perfectament ben definida respecte a la funció d'elecció $e(\cdot)$, i pot satisfer-se sense necessitat que cap relació binària racionalitzi $e(\cdot)$. Hi ha autors que prefereixen associar la racionalitat a propietats «desitjables» del $e(\cdot)$, tant si $e(\cdot)$ pot explicar-se com a resultat d'una maximització com si no es pot.

c) Un tercer aspecte en què pot matisar-se el model senzill de l'apartat anterior es refereix als subconjunts sobre els quals és definida una funció d'elecció. Ja hem dit que representa una descripció exògena de comportaments. Podem interpretar-la com un recull sistemàtic de quines serien les decisions observables d'un agent que triés, en diverses ocasions, sobre conjunts variables d'alternatives. Si volem representar circumstàncies en què mai no s'oferiran a l'agent determinats subconjunts, llavors haurem de modificar la definició de funció d'elecció, admetent que pot ser definida només sobre determinada classe dels subconjunts d' A , $B \subseteq A$. Un exemple clàssic ens el dona l'estudi de les funcions (o correspondències) de demanda. Efectivament, una correspondència de demanda ens diu quins vectors de consum són escollits, per part d'un consumidor, per a cada restricció pressupostària. Els conjunts d'alternatives sobre els quals és definida no són, doncs, tots els subconjunts de l'espai de consum, sinó només aquells determinats com a accessibles i possibles per la renda, els preus dels béns, i el conjunt de consum.

En general, una funció d'elecció sobre una família de subconjunts B és racionalitzable (per a una determinada classe de relacions binàries R) si existeix alguna relació de pre-

ferències $R \in \mathcal{R}$ tal que $e(B) = C(R, B)$ per tot $B \in \mathcal{B}$. En particular, els estudis sobre preferència revelada són exercicis sobre la racionalitzabilitat de determinades correspondències entre conjunts pressupostaris (determinats per preus i rendes) i vectors de consum (demandes). El model clàssic del consumidor li atribueix preferències completes i transitives, i un comportament maximitzador. Aquest model, restringit a eleccions sobre conjunts pressupostaris, genera funcions d'elecció a les quals anomenem funcions (o correspondències) de demanda. Inversament, podem preguntar-nos quan una correspondència d'aquest tipus és racionalitzable, ja sigui transitivament, o acíclicament. Aquestes són les preguntes canòniques en allò que tradicionalment s'ha anomenat teoria de la preferència revelada. Veure-ho dins d'aquest marc general ha de deixar ben clar que no es tracta d'una teoria alternativa del comportament del consumidor, sinó d'un tipus d'exercici invers al de derivar funcions de demanda tot partint de relacions de preferència.

d) Fins ara no hem estat gens explícits respecte a l'estructura de les alternatives, ni quant a la possibilitat que aquestes es puguin expressar en formes que facilitin el seu tractament matemàtic. Tampoc no hem esmentat, si bé és un tema familiar per a tot economista, que les preferències acostumen a representar-se mitjançant funcions d'utilitat. Considerar aspectes concrets lligats a l'estructura específica de l'espai d'alternatives i/o al tipus de preferències admissibles per racionalitzar comportaments ens introdueix en refinaments possibles entorn de les preguntes genèriques sobre la racionalitat.

Prenem el cas concret de la teoria del consumidor. És tradicional identificar l'espai de consums amb el conjunt de vectors no negatius en \mathbb{R}^n on n seria el nombre de béns de consum considerats pel consumidor i un vector $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ ens indicaria les quantitats de consum x_i de cada bé i associades a una possible decisió. Cada $x \in \mathbb{R}^n$ és, doncs, una alternativa, i la identificació del conjunt d'alternatives amb un espai matemàtic ens permet parlar de nous conceptes i problemes. Atès una restricció pressupostària ens ve donada un cop coneixem la renda del consumidor i els preus de cada bé, podem veure que una correspondència de demanda ens associa a cada punt de \mathbb{R}^{n+1} (n nombres per als preus de cada bé, i un més per a la renda), un conjunt de \mathbb{R}^n , que interpretem com els vectors de consum escollits. O, si per cada restricció pressupostària tenim un sol vector de consum escollit parlarem d'una funció de demanda. En un espai prou estructurat com és aquest, té sentit parlar, per exemple, de la continuïtat de les funcions de demanda, o de la seva diferenciabilitat, o de propietats similars en el cas més general de correspondències. I això obre camí a noves preguntes. Per exemple: ¿quan serà racionalitzable una funció continua com a funció de demanda, i quina mena de preferències hi haurà al darrera? Clarament, com més específics siguin els comportaments a racionalitzar, més restrictives seran també les condicions sobre el tipus d'agent que els generi i més estret el concepte de racionalitat implícit. És qüestionable, doncs, si pot considerar-se que un requisit de continuïtat té gaire a veure amb la racionalitat d'un agent. Però sense entrar en qüestions semàntiques, cal assenyalar que, en afegir estructura als models, s'obren noves qüestions lligades estretament a la pregunta bàsica de la racionalitzabilitat.

També de la banda de les preferències es poden plantejar refinaments i precisions. Donada qualsevol relació binària és possible plantejar-se la possibilitat de recórrer a representacions numèriques que ens descriguin els seus aspectes rellevants. El cas més clàssic el dona la representació de relacions binàries completes i transitives sobre un espai d'alternatives mitjançant funcions d'utilitat. Donada una relació R sobre A , direm que una funció $f: A \rightarrow \mathbb{R}$ representa a R si els valors numèrics que f li atribueix per a cada alternativa $x \in A$ (la utilitat de x) ordenen A de la mateixa forma que R . És a dir, si per a tot $x, y \in A$, $f(x) \geq f(y)$

sempre i quan xRy . Depenent de l'estructura de l'espai d'alternatives i de la relació binària R , poden o no existir funcions f que representin R . Si f representa R , i φ és una transformació monòtona creixent, $\varphi \circ f$ clarament també representa R . També pot —o no— ser possible que R admeti representacions contínues, o diferenciables, o quasi-còncaves, i cadascuna d'aquestes restriccions matemàtiques admet interpretacions econòmiques. Novament, aquest enriquiment del possible llenguatge sobre preferències, unit a la varietat de comportaments possibles representats per funcions d'elecció, obren camp per plantejar-se qüestions més matisades sobre la relació entre comportaments observats i preferències subjacents.

e) En els anteriors apartats hem fet esment de diverses modificacions possibles en el tipus de preferències que racionalitzen una funció d'elecció. Però les preferències són només un dels ingredients dins el model canònic de racionalitat que ens ha servit de base. L'altre és la hipòtesi de maximització, el supòsit que l'agent triarà les millors alternatives al seu abast. També podríem posar-lo en qüestió, i de fet aquest ha estat un terreny abonat pel debat sobre la concepció de racionalitat, potser massa simplista, que hi ha darrera dels models econòmics. Formalment no és difícil generar hipòtesis alternatives de comportament compatibles amb el mateix format preferències-criteri amb què treballem. Per exemple, podríem descriure el comportament d'un agent que tria sistemàticament alternatives classificades en tercer lloc, i no en primer. Però deixant de banda altres dificultats tècniques (què és el «tercer lloc» no sempre és ben definit), seria potser difícil trobar interpretacions atractives d'un model d'aquest tipus, que ens permetessin associar-lo a una idea acceptable de «racionalitat». Hi ha altres arguments, a la vegada més rics i més difícils de formalitzar, per posar en dubte una actitud estrictament maximitzadora per part dels agents. Uns es basen en la complexitat de les decisions, les quals durien a prendre accions només «satisfactòries» aproximadament, però no del tot òptimes. D'altres depenen de la inevitable ambigüïtat associada a tota modelització, la qual sempre deixarà fora alguns aspectes que, des d'algun punt de vista, podríem considerar rellevants per a un agent totalment racional. Així, per exemple, en un model en què només es recullen decisions per un període poden quedar excloses, per una definició estreta de maximització, aquelles consideracions de llarg termini que facin que a un comportament maximitzador dins un període no ho sigui des d'una perspectiva intertemporal. O, en un model en què no es reconeixin explícitament les interaccions entre diversos agents, poden no quedar recollides consideracions estratègiques naturals en el comportament de cadascun d'ells. El que vull assenyalar és que molts cops es posa en dubte la rellevància de la hipòtesi de maximització dins un model de comportament racional, quan el que s'hauria de qüestionar és el model de base: la descripció de les alternatives, la pertinència de tractar l'agent decisor de manera aïllada, la forma de tenir en compte les restriccions imposades per limitacions informacionals i de còmput en decisions complexes, etcètera. Això no vol dir que s'hagi d'acceptar la maximització com un dogma, però sí que per qüestionar-la convé fer-ho en el context de models més rics que el que hem pres com a punt de partida. Per exemple, discussions sobre la relació entre la hipòtesi de maximització i la complexitat de les decisions només tenen sentit en contextos en què la descripció de l'espai d'alternatives permet una definició rigorosa de complexitat, i poden proposar-se comportaments alternatius ben definits, tenint en compte la descripció del problema. Atès les dificultats d'aquest programa, la formalització de criteris de racionalitat que impliquen comportaments no maximitzadors és encara un camp important de recerca.

4. Racionalitat individual. Decisions enfront de la natura

Un cas particular en què interessa analitzar el comportament dels agents i la seva racionalitat és aquell en què es prenen decisions sense saber amb exactitud quines en seran les conseqüències. Hi ha diferents formes de modelar un entorn amb aquestes característiques, i l'oportunitat de cadascuna dependrà de la situació que volguem descriure. El punt de partida, en tot cas, consistirà a reconsiderar quines són les alternatives entre les quals pot triar l'agent, i com representar-les.

Un primer cas tradicional es dona quan un agent pot triar entre diferents accions, cada acció té conseqüències possiblement diferents segons la realització d'un fenomen de caràcter aleatori que queda fora del seu control, i la distribució de probabilitats associada al fenomen aleatori en qüestió és coneguda. És el cas de decisions en condicions de risc, en la terminologia proposada ja fa cinquanta anys per F. Knight. Un exemple en seria la decisió d'un camperol entre sembrar dos tipus de vegetals, x o y , tot sabent que l'un rendeix millor que l'altre, i viceversa, segons el clima de l'any sigui sec o no, i tot disposant d'estimacions que considera fiables respecte a les probabilitats d'una i altra situació climàtica. O la decisió d'un jugador de comprar o no loteria, i quanta, tenint en compte els diners que té i les diferents probabilitats que li toquin premis. Suposem, per simplificar, que les possibles realitzacions del fenomen aleatori del què depenen les conseqüències de les accions són finites; els direm estats de la natura. Cada acció pot identificar-se amb l'expressió de les seves conseqüències per a cada estat de la natura, és a dir, una n -pla de conseqüències. En particular, si aquestes poden expressar-se numèricament (i aquest seria el cas si les conseqüències fossin quantitats de diners disponibles, com a la loteria), un vector ens serviria per descriure alternatives. La seva interpretació seria ben diferent que en el cas d'un vector de consum. De tota manera, un cop hem interpretat convenientment què és una alternativa, podem començar novament a preguntar-nos, en els mateixos termes que en seccions anteriors, què són comportaments racionals.

Una funció d'elecció, així, ens triarà llistes ordenades de conseqüències representatives cadascuna d'una acció. I un comportament serà transitivament racionalitzable, per exemple, si existeix una relació binària transitiva sobre aquestes n -plas de conseqüències i el qual tingués com a elements maximals per a cada subconjunt admissible precisament aquells que són la imatge del conjunt segons la funció d'elecció.

Ara bé: com ja hem assenyalat abans, un cop l'espai d'alternatives té una determinada estructura, sorgeix la possibilitat de plantejar-se noves qüestions, més específiques, que no podrien ni tans sols formular-se en situacions més generals. Aquest és el cas, en particular, dins de la teoria de decisions en condicions de risc. Les alternatives consisteixen a triar accions, però aquestes accions tenen conseqüències: podem parlar d'un nou conjunt, el conjunt C de les conseqüències possibles d'una acció per un estat de la natura determinat. I sorgeixen qüestions naturals, específiques per a aquest context. Una d'elles és la següent: siguin $a, b, c \in C$ conseqüències possibles d'accions, en una situació on hi ha només dos estats de la natura rellevants, i en què així doncs cada acció alternativa pot descriure's com un parell ordenat. Suposem que un agent té preferències (potser revelades com a racionalització d'una funció d'elecció) segons les quals (a, a) és preferit a (b, b) , però (b, c) és preferit a (a, c) . Podem qualificar de racional aquest agent? Certament, aquestes dues observacions són perfectament compatibles amb una relació binària, fins i tot transitiva, definida sobre els parells de conseqüències. Però potser, en aquest cas, queda recança. Al capdavall, (a, a) i (b, b) representen accions en què, amb tota certesa, independentment de l'estat de la natura

que s'acabi produint, s'obtindran les conseqüències a i b , respectivament. Si bé, fins ara, no n'hem tingut necessitat, sembla natural parlar de preferències sobre les conseqüències, com a relació binària diferent (sobre C) a les preferències sobre accions, que són relacions sobre $A = C \times C$. I, lògicament, si l'acció (a, a) és preferida a l'acció (b, b) , i les preferències entre accions i conseqüències han de ser mínimament consistents, la conseqüència a serà preferida a la conseqüència b . Si això és així, i fent ara el camí invers, sembla coherent que (a, c) es preferís a (b, c) , ja que, al capdavant, són accions només diferents en les conseqüències per a un determinat estat de la natura, i la primera té una conseqüència millor que la segona, en aquest cas. Aquest raonament, plausible però de cap manera indiscutible, es pot formalitzar en termes d'un axioma sobre les ordenacions d'accions, segons el qual (a, a) preferit a (b, b) requeriria que (a, c) es preferís a (b, c) , per a tot c . És l'axioma d'independència, que pot o no considerar-se com un requisit atractiu o, fins i tot, com una forma de coherència i de racionalitat.

Tan si té a veure o no amb la racionalitat, l'axioma d'independència és l'element clau, juntament amb altres axiomes de caràcter més tècnic, per a un resultat important respecte a les preferències individuals sobre eleccions en condicions de risc: ens permet, quan se satisfà, representar aquestes preferències sobre accions mitjançant una funció lineal d'utilitat obtinguda de la següent manera: atribuir, primer, valors numèrics (utilitats) a cada conseqüència, i calcular per a cada acció la seva esperança matemàtica, com a suma dels productes de la utilitat de la conseqüència que l'acció associa a cada estat de la natura, per la probabilitat que aquest estat es produeixi. Sense entrar en detalls tècnics, veiem que es tracta d'un resultat de representació. Si tenim preferències completes i transitives sobre les alternatives (accions) i, a més, aquestes preferències satisfan requisits addicionals (essencialment la independència), aleshores existirà una funció numèrica sobre l'espai de les conseqüències amb la propietat que l'esperança matemàtica dels seus valors, calculada per cada acció, ens donarà utilitats apropiades que representen la relació binària inicial. Aquestes utilitats, per construcció són lineals, i es coneixen com a funcions d'utilitat de Von Neuman-Morgenstern. Els valors numèrics definits sobre les conseqüències no poden modificar-se per qualsevol transformació monòtona i mantenir les seves propietats, sinó tans sols per transformacions afins, de la forma $\varphi(x) = a + bx$, amb b no negatiu. Cal notar que, entre les hipòtesis tècniques a les quals no ens hem referit explícitament, hauríem d'haver inclòs prou estructura a l'espai d'accions (i de conseqüències) perquè sigui un espai lineal, sense la qual cosa algunes de les afirmacions anteriors no tindrien sentit.

El cas de decisions en condicions de risc no és sinó un dels contextos en què les accions dels agents no especifiquen del tot les seves conseqüències. Si bé hem parlat de les decisions com a n -plas de conseqüències, una per a cada estat de la natura, la descripció d'una situació amb risc exigeix també que les probabilitats de cada un d'aquests sigui donada. Una descripció més pobra de la situació es limitaria a descriure una acció associant una conseqüència a cada estat, però sense especificar-ne la probabilitat. Aquest és el punt de partida comú de dos tipus de literatura: la basada en nocions de probabilitat subjectiva i la que considera condicions d'incertesa, on els agents no tenen cap coneixement sobre les probabilitats de cada un dels estats.

Comencen per descriure noves qüestions, lligades al pensament de Savage, que poden plantejar-se si tenim una relació de preferències completa i transitiva sobre accions, i sense menció de cap mena a les probabilitats dels estats. L'argument que ja hem fet abans, i que ens ha portat a la condició d'independència és encara tan plausible com abans. Però ara podríem afegir un altre argument, lligat a la coherència del comportament de l'agent, tot fixant-nos en

la possibilitat que les preferències sobre accions també siguin un reflex, no tan sols de les seves preferències sobre conseqüències, sinó també expressió implícita de les seves creences sobre la probabilitat amb què es donaran els diferents estats de la natura. Suposem, com en l'exemple anterior, que a, b, c, d són conseqüències possibles, que hi ha dos estats de la natura (ara amb probabilitats desconegudes), i que (a, a) és preferit a (b, b) , tot indicant així que la conseqüència a és millor que la conseqüència b . Si, ara, tenim que (a, b) es considera millor que (b, a) , quina conclusió en podem treure? Novament, i sense que això sigui exigint en absolut pel fet que les preferències siguin relacions binàries transitives sobre A , sembla plausible que un agent «coherent», que prefereixi a sobre b , només triarà (a, b) sobre (b, a) si el primer estat té més probabilitats de donar-se que el segon. Si això és així, i (c, c) és preferit sobre (d, d) , un altre argument de «coherència» exigiria que (c, d) fos preferit a (d, c) . De no ésser així, les «probabilitats subjectives» atribuïdes als estats primer i segon mitjançant la comparació de (a, b) i (b, a) serien diferents que les corresponents als mateixos estats comparant (c, d) i (d, c) .

Tampoc aquesta dependència dels estats seria, en sí mateixa, incompatible amb preferències transitives sobre les accions, però si volem respectar el tipus de «coherències» que acabem de descriure caldrà imposar una nova condició sobre les preferències, un axioma «d'independència de les probabilitats subjectives respecte a les conseqüències», en certa forma dual a la independència abans assenyalada, i que tenia un valor «d'independència de les preferències sobre conseqüències respecte a les probabilitats». De fet, i estenent el tipus d'estudis sobre representabilitat a l'estil de Von Neuman-Morgenstern al cas menys estructural de què parlem ara, es pot demostrar que quan una relació de preferències sobre les accions alternatives (n -plas de conseqüències associades a n estats de la natura) satisfà els dos tipus assenyalats d'independència —més algun altre requisit de tipus tècnic— admet una representació numèrica interessant, basada en dues funcions induïdes per l'estructura de la preferència. D'una banda, es pot extreure una funció que valori cada conseqüència, i d'altra una que li atribueixi una probabilitat a cada estat, amb la propietat conjunta que els valors esperats corresponents a cada acció siguin, de fet, una funció d'utilitat que representa les preferències de partida. Des d'aquest punt de vista, les preferències que satisfan aquests axiomes —ja siguin preses com a primitives o induïdes com a racionalització d'una funció d'elecció— admeten una justificació molt elaborada, com a conseqüència de la ponderació de probabilitats subjectivament estimades i de valoracions de les conseqüències de cada acció per a cada estat de la natura. Si aquest nivell d'elaboració correspon o no a un grau superior de racionalitat a l'implícit en altres ordenacions de les accions que no satisfan les condicions d'independència és una qüestió semàntica. Però podem constatar com s'integren al discurs elements addicionals de «racionalitat», en un sentit cada cop més ampli, a mesura que l'espai de les alternatives agafa estructura formal i es dota d'interpretacions específiques.

Un altre cas en què partim de la mateixa estructura formal (ordenacions d'accions enteses com a n -plas de conseqüències) però que ens formulem qüestions d'un altre tipus és el lligat a la idea d'incertesa —novament en terminologia de Knight—, és a dir, absència total de coneixement de les probabilitats lligades a cada estat de la natura. Certs autors negarien sentit a la possibilitat d'ignorància de les probabilitats, admetent que tot agent té, en cada moment, avaluacions subjectives de probabilitats, per molt que siguin revisables, a mesura que obtingui millor informació: és el punt de vista bayessià, i el nostre marc s'acostaria a suposar que les hipòtesis d'independència —que com hem vist permeten parlar amb precisió de probabilitats subjectives—, són essencials en preferències racionals. D'altres, però, tot acceptant que la idea d'incertesa té sentit, han proposat fer ordenacions concretes de les possibles accions, segons

críteris específics, i n'han discutit l'atractiu (la «racionalitat», en sentit ampli) en termes axiomàtics. Concretament, en un marc en què cada conseqüència tingui atribuït un valor d'utilitat, i en què cada acció sigui doncs representada per un vector n -dimensional, les accions poden ordenar-se segons el criteri *maximin* (atenent a l'ordre de preferències entre conseqüències i comparant aquelles que s'obtindrien de produir-se l'estat més desfavorable en cada cas), o segons el criteri *maximax* (similar a l'anterior però posant-nos, per a cada acció, en el millor dels casos), o segons combinacions i variants de l'un i l'altre. Caracteritzar axiomàticament aquestes ordenacions que disposen de cert atractiu, com ho prova el fet que s'hagin proposat de forma recurrent en molts àmbits i contextos diversos, és una forma de presentar-ne formalment les propietats i, per omissió, els defectes, com a críteris «racionals» de decisió.

Una idea atractiva, que s'ha explorat recentment, parteix del fet de reconèixer que les hipòtesis de risc i d'incertesa representen casos extrems de situacions en què no es controla la connexió entre accions i resultats. Els casos intermedis en què possiblement se situen moltes decisions serien aquells en els quals els agents no coneixen del tot les probabilitats dels estats de la natura, ni en són del tot ignorants. Una forma de concretar aquesta idea consisteix a suposar un marc general en el qual els agents poden atribuir un interval (o, més en general, un conjunt de valors possibles) a cada estat de la natura. Casos extrems d'aquest model serien el risc (interval degenerat a un sol valor per a cada estat) i la incertesa (en què l'interval $[0, 1]$ vindria associat a les probabilitats possibles per a cada estat). Dins d'aquest marc més ampli poden reconciliar-se formes de comportament tradicionalment associades al risc i a la incertesa, com són el càlcul d'utilitat esperada i el criteri *maximin*, respectivament. Per exemple, sota hipòtesis apropiades, les preferències sobre accions quan aquestes són descrites per les conseqüències en cada estat de la natura, i a cada estat li correspon un interval de possibles probabilitats, poden representar-se com si els agents associessin a cada acció els intervals de valors d'utilitat esperada derivats d'un determinat índex d'utilitat per a cadascuna de les probabilitats possibles, i triessin, després, aquelles accions que maximitzen la mínima esperança garantida.

5. Racionalitat individual. Preferències sobre conjunts d'alternatives

En l'apartat anterior hem vist com una de les qüestions associades amb la racionalitat d'un agent pot ésser la coherència entre diverses relacions valoratives (de preferència, o de probabilitat) que naturalment s'encadenen entre si: per exemple, coherència entre la valoració relativa dels actes, de les conseqüències i de les probabilitats dels estats de la natura que lliguen a uns i altres. Diversos models plantegen qüestions de coherència anàlogues en un format molt senzill però que, como de seguida veurem, admet moltes interpretacions alternatives interessants.

Suposem un agent del qual coneixem les preferències sobre un conjunt d'alternatives A . ¿Podem dir alguna cosa quant a quines seran les seves preferències respecte als diferents subconjunts d'alternatives? Per exemple, si sabem que $x, y \in A$, i que x és preferit a y , ¿quines serien les preferències de l'agent si se li demana de comprar els conjunts $\{x, y\}$ i $\{x\}$? La resposta més prudent és dir que depèn. Depèn de què vulguem representar pel conjunt $\{x, y\}$. Si, per definició, tant x com y són estats mútuament excloents, un o altre acabarà sortint escollit, i convé saber com. Per exemple, podria ésser que

- 1) si s'escull $\{x, y\}$, una de les dues alternatives es triarà aleatòriament sense cap intervenció ulterior de l'agent; o bé que
- 2) si s'escull $\{x, y\}$ l'agent podrà triar en una etapa posterior, i dins d'aquest cas que
 - 2a) les seves preferències no puguin canviar d'un moment a l'altre o que
 - 2b) l'agent consideri possible un canvi de preferències.

Cadascun d'aquests casos és interessant en sí mateix, però és clar que el que és raonable, sota cada interpretació, serà ordenar diferentment $\{x, y\}$ respecte a $\{x\}$. Sota la interpretació 1), $\{x\}$ garanteix que aquesta alternativa sigui escollida, i així doncs $\{x\}$ serà millor que $\{x, y\}$, ja que aquest segon conjunt no exclou l'elecció de l'alternativa menys preferida y . Sota la interpretació 2a), $\{x\}$ i $\{x, y\}$ serien indiferents, ja que l'agent pot encara triar, i triarà $\{x\}$ en tot cas, en la segona etapa. En canvi, sota 2b), un agent preferirà $\{x, y\}$ sobre $\{x\}$, ja que això li dóna més gran flexibilitat per decidir en el futur, i s'admet la possibilitat que canviïn les seves preferències. No hi ha, doncs, una forma «racional» de lligar les preferències sobre alternatives amb les preferències sobre conjunts d'alternatives, o, en una interpretació diferent, les preferències sobre conseqüències. Però sí que, per a cada interpretació d'aquests conjunts, podem preguntar-nos pels possibles lligams entre unes i altres. Comentarem breument algunes d'aquestes interpretacions.

Fins ara, tant si hem parlat de risc com d'incertesa, hem admès implícitament que coneixíem els estats rellevants de la natura, i que això ens permetia descriure cada acció, associant una conseqüència a cada estat (amb o sense una probabilitat). Una descripció més primitiva d'una acció consistiria a enumerar el conjunt de les seves possibles conseqüències, sense lligar-les explícitament a cap estat de la natura. Ordenar accions, aleshores, és ordenar conjunts de conseqüències, i un agent pot ser més o menys racional segons quins siguin els lligams entre la seva valoració de conseqüències i la seva ordenació d'accions. Un exemple de possible ordenació de conjunts de conseqüències, que ha estat estudiat sistemàticament de forma axiomàtica, consisteix a ordenar-los lexicogràficament respecte a un indicador d'utilitat de les conseqüències.

Un altre cas interessant és aquell en què els agents ordenen, avui, els conjunts d'alternatives sobre els quals podran triar demà, sense tenir certesa de quines seran les seves preferències en el moment de l'elecció, sinó només una distribució de probabilitat sobre quines poden ésser. De fet, en esmentar aquest cas comencem a introduir un nou element no considerat fins ara: la possibilitat que els agents no tinguin preferències constants al llarg dels diversos actes d'elecció contemplats en discutir la racionalitat dels seus actes. Si un agent no està segur de què voldrà en el futur, el que és racional és deixar la màxima flexibilitat oberta per al futur, que es tanquin el mínim de portes. Aquestes són idees molt arrelades en el pensament econòmic: la preferència per la liquidesa, que porta a retenir actius poc productius o el valor d'opció d'un recurs no renovable, que fa moderar les seves taxes d'explotació, són exemples clàssics de preferència per la flexibilitat. Poden expressar-se formalment, en el nostre llenguatge, com a problemes d'ordenació de conjunts d'opcions en funció de preferències no del tot certes, però respecte a les quals es busca ser coherent. Una vegada més, a un marc més complex li corresponen noves qüestions possibles, les quals sempre poden posar-se en relació amb una idea general de racionalitat.

6. Eleccions intertemporals. Variacions de les preferències

Amb l'excepció de l'últim cas esmentat, fins ara hem parlat d'eleccions i de preferències considerant que un agent racional ha de ser guiat per una relació invariable al llarg dels diversos actes respecte als quals avaluem si el seu comportament és o no coherent. No hem donat importància explícita al pas del temps entre unes decisions i altres, ni introduït en cap forma la possibilitat que el criteri d'elecció s'alteri, endògenament o exògenament. La introducció del temps no és, necessàriament, una font de variacions en les preferències. Tot el que pugui dir-se, en general, d'eleccions racionals en espais d'alternatives, pot dir-se, en particular, del cas en què aquestes alternatives comprenen decisions datades: per exemple, en teoria de la demanda, si les alternatives són vectors indicatius de les quantitats de cada bé a consumir durant els diversos períodes d'anàlisi, una funció d'elecció estarà definida sobre aquests vectors, és a dir, sobre seqüències de consums per als diversos períodes, i si és racionalitzable ho serà en termes d'una relació binària sobre aquestes seqüències. Es poden plantejar, donada l'estructura intertemporal, noves qüestions, per exemple, sobre la possibilitat d'expressar les preferències sobre seqüències com a funció de preferències més elementals sobre els consums en cada període, o els tipus de descompte intertemporals implícits en una preferència determinada. Però moltes d'aquestes qüestions no plantegen problemes conceptuals addicionals.

Més complicat és el cas en què volguem reconèixer que els mateixos actes d'elecció fets en un període poden tenir incidència sobre els criteris d'elecció futurs. Un exemple clar en serien les preferències d'un consumidor que desenvolupa un hàbit per aquells productes que ja ha consumit prèviament. Així podríem observar un consumidor que destina quantitats creixents dels seus recursos a comprar tabac. La seva conducta pot no ser racionalitzable si exigim com a condició de racionalitat que la seqüència d'actes sigui maximal respecte a una relació única sobre totes les seqüències possibles, però podria admetre's una racionalització més feble, segons la qual cada decisió en la seqüència temporal és òptima respecte a una relació sobre els consums del període, i el seguit de relacions d'un període a l'altre representa el canvi de preferències induït pels consums dels períodes anteriors. Admetre aquests canvis, sense qualificació, ens podria portar a considerar qualsevol seguit de decisions com a racionalitzable. Però si imposem restriccions sobre quins canvis de preferències són admissibles en funció dels consums anteriors podem restablir estructura en el model, i intentar redefinir quan els comportaments d'un agent amb preferències variables són racionals. La tasca, però, no és senzilla, perquè existirà sempre una tensió entre les preferències de l'agent abans d'iniciar les seves decisions —les quals, per exemple, poden portar-lo a decidir no començar a fumar per no crear-se'n l'hàbit— i els interessos del fumador potencial que hi ha en ell i que, de poder manifestar-se, optaria per consumir tabac, per tal de poder-ne gaudir cada cop més. Amb quines de les múltiples seqüències de preferències possibles cal avaluar la racionalitat d'un agent es converteix en una qüestió interessant, però difícil de contestar inequívocament.

Una formulació molt diferent de la possibilitat de canvis en les preferències consisteix a suposar que aquestes es modifiquen de forma estocàstica. És a dir que, en cada circumstància en què l'agent hagi d'escollir, existeix una probabilitat p_1 perquè ho faci d'acord amb certes preferències R_1 , una probabilitat p_2 perquè esculli el millor segons unes preferències R_2 , etcètera. Des del punt de vista de la seva descripció interna, a un agent ja no li vindria associada una relació de preferències, sinó una distribució de probabilitat sobre les preferències possibles, donades les alternatives. Direm que té preferències estocàstiques si la

hipòtesi de maximització pot mantenir-se suposant que, un cop determinada aleatòriament la preferència que guiarà l'elecció, es tria el millor d'acord amb tal preferència.

La descripció externa de les accions a les quals aquest model dóna lloc no es pot fer mitjançant funcions d'elecció. Suposem, per exemple, que un agent es regeix per unes preferències estocàstiques segons les quals, amb una probabilitat $\frac{1}{2}$ té la preferència $zPyPz$, i amb probabilitat $\frac{1}{2}$ la preferència $yPzPx$. Quan se li presenta l'oportunitat d'escollir x, y, z , escollirà x o y , cadascuna amb una probabilitat $\frac{1}{2}$, perquè una o l'altra serà la millor alterna-

tiva segons una o altra preferència. Quan tan sols pugui escollir entre y o z sempre es quedarà amb y , mentre que, en escollir entre x o z ho farà novament amb una probabilitat $\frac{1}{2}$ per a cadascuna. Per tal de reflectir sistemàticament comportaments d'aquest tipus podem estendre la seva descripció externa mitjançant el concepte de funció estocàstica d'elecció: per a cada subconjunt d'alternatives aquesta funció ens donarà una distribució de probabilitat sobre l'esmentat subconjunt. Si $c(\cdot, \cdot)$ és una funció estocàstica d'elecció, per a cada $B \subseteq A$, $c(\cdot, B)$ serà una distribució sobre B . I, per a cada x , $c(x, B)$, ens indicarà la probabilitat que x sigui l'alternativa escollida quan B representa el conjunt de totes les alternatives possibles. Naturalment, s'haurà de complir que $c(x, B) = 0$ si $x \notin B$ (la probabilitat d'escollir una opció no disponible és nul·la) i que $\sum_{x \in B} c(x, B) = 1$ (sempre s'escollirà alguna cosa entre el que hi ha de disponible).

Podríem pensar que, un cop admesa la possibilitat de preferències canviants, qualsevol funció estocàstica d'elecció seria racionalitzable, però no és així. Fins i tot en aquest context més ampli les hipòtesis de maximització i de transitivitat imposen restriccions importants. Per exemple, si un agent escull x sobre y en un 60 % dels casos, i y sobre z en un altre 60 %, haurà d'escollir x sobre z com a mínim amb una freqüència del 20 % perquè el seu comportament sigui racionalitzable estocàsticament amb preferències transitives. Efectivament, les eleccions indicades exigeixen que amb probabilitat 0,6 s'obtinguin preferències en què xPy , i també amb probabilitat 0,6 preferències en què yPz . Això només és possible si s'atribueix probabilitat almenys 0,2 a preferències en què simultàniament xPy i yPz , per a les quals la transitivitat implica xPz . Així doncs, fins i tot si admetem canvis aleatoris en les preferències individuals, es pot parlar encara de diferents graus de coherència en les eleccions.

El model de preferències estocàstiques és compatible amb la hipòtesi que els agents tenen preferències estables; aquests podrien regir-se per una única relació de preferències, definida sobre un espai d'alternatives més ric que el considerat per l'observador del seu comportament, i els canvis en la seva conducta ser només respostes a modificacions de condicions no observables sota les quals es produeixen les eleccions. Així doncs, a més d'una possible hipòtesi alternativa sobre el comportament dels agents, el model de preferències estocàstiques es presenta també com a una fórmula flexible per representar situacions en què, d'entrada, reconeixem dificultats sistemàtiques d'observació de variables rellevants: d'aquí les seves connexions amb certs models econòmics de gran interès, com els PROBIT i LOGIT.

Igualment que en el cas determinista, certes regularitats de les funcions estocàstiques d'elecció podrien acceptar-se com a condicions de racionalitat sense necessitat de ser justificades mitjançant relacions binàries, o distribucions sobre aquestes. Aquesta és la posició adoptada per una àmplia literatura, més lligada a la psicologia que no a l'economia, on es descriuen restriccions diverses, sota el nom genèric de transitivitat estocàstica. Un exemple en

seria la condició que la probabilitat d'escollir x sobre z fos almenys tan gran com la mínima entre les probabilitats d'escollir x sobre y i la d'escollir y sobre z .

7. Consideracions finals

Malgrat el grau de complexitat creixent que hem introduït en la descripció de contextos en què es pot qüestionar la racionalitat individual, o senzillament expressar-ne possibles significants, hem mantingut fins ara una hipòtesi enormement simplificadora. Hem suposat que les accions de l'agent considerat tenen conseqüències ben definides, encara que potser no en termes de certesa, independentment de l'acció de qualsevol altre agent. Si volem reconèixer que, de fet, la majoria de les accions individuals interaccionen amb les d'altres agents, i que les conseqüències d'un conjunt d'accions no poden ser controlades totalment per cap dels qui les conformen parcialment, haurem de revisar les nostres nocions de racionalitat. Els instruments apropiats per a fer-ho són els de la Teoria dels Jocs, que es l'objecte d'un altre treball en aquest volum.¹ Ens limitarem a fer uns comentaris generals sobre les noves dificultats que cal superar en intentar definir què s'entén per racionalitat en un context de jocs no cooperatius.

Existeixen diferents formes de modelar les interaccions estratègiques entre agents per part de la teoria de jocs: la forma normal o estratègica i l'extensiva són les utilitzades en l'anàlisi de comportaments no cooperatius.

En totes dues, per regla general, se suposa que els agents tenen preferències ben comportades, i amb freqüència que les alternatives són loteries i les preferències sobre elles representables mitjançant funcions d'utilitat Von Neuman-Morgenstern. Però si, precisament, gran part del que ha estat tractat en aquest treball es refereix a l'existència i representabilitat de preferències que justifiquen les eleccions dels agents, què queda aleshores per explicar si aquelles es donen per suposades? L'èmfasi es desplaça ara cap a l'altre ingredient ja discutit del comportament racional —la hipòtesi de maximització—, a la vegada que adquireix importància un nou aspecte: la formació de creences i expectatives. No és que es qüestionï l'interès de cada agent per actuar «el millor possible» donades les seves preferències, sinó que es fa difícil, en un context d'interacció estratègica, definir què és «millor». El marge d'actuació de cada agent consisteix a escollir la seva estratègia, però les conseqüències de l'esmentada elecció també dependran de les estratègies adoptades pels altres. La seva decisió dependrà, doncs, de quines cregui que seran les dels altres jugadors. En la teoria de l'elecció sota condicions de risc se suposa que aquelles creences, encertades o no, vénen donades per una distribució de probabilitat sobre els estats de la natura, reflex de condicions aleatòries no controlades, però tampoc modificables per l'acció de l'agent decisor. En una decisió estratègica, en canvi, un agent racional, sigui quina sigui la definició concreta que adoptem per a aquest terme, reconeixerà que les accions dels altres agents dependran de les que ell pugui adoptar, tant com les seves depenen de les dels altres, sota la hipòtesi essencial en teoria de jocs segons la qual cada agent atribueix als altres igual capacitat d'anàlisi que la seva pròpia. Les accions òptimes per a un depenen de les dels altres, però quines vagin a ésser aquelles depèn de la d'aquell, i això ens obliga a reconsiderar què és «el millor». Sorgeix, imponent, un problema de regressió a l'infinit, en el qual «el que jo crec que els altres creuen que jo crec que els altres creuen...» es converteix en la clau, aparentment insoluble, per definir un comportament racional.

1. Vegeu Ricart i Costa en aquest mateix volum.

Una manera de trencar aquesta regressió a l'infinit consisteix a adoptar algun concepte d'equilibri, la qual cosa suposa renunciar a les idees de causalitat en favor de les de simultaneïtat. Dir què és racional, en aquest context, equival a elaborar una teoria del comportament estratègic d'agents no cooperatius, amb la propietat que, si tots ells compartissin aquesta teoria, cada un d'ells consideraria que el seu comportament, i el dels altres, són els millors possibles. No parlarem, doncs, del comportament racional d'un agent, sinó de combinacions d'estratègies que són conjuntament compatibles amb la hipòtesi que tots els jugadors són racionals, tot sabent que els altres ho són també, i comparteixen una mateixa teoria de la racionalitat. Però, fins i tot des d'aquest punt de vista no existeix una forma única de descriure la racionalitat, perquè hi ha diferents teories alternatives que els agents poden comparar. En conseqüència, la teoria de jocs no gaudeix del benefici d'una única noció d'equilibri, sinó de diverses. Darrera de cada una de les quals es troba, implícita, una teoria del comportament racional.

Sense necessitat de penetrar en aquest món més complex, espero que els models analitzats en el cos d'aquest treball hagin servit per il·lustrar tant les dificultats com l'interès que té analitzar críticament els significats del concepte de racionalitat en economia.

Referències

- Aizerman, M. A. (1985): «New Problems in the General Choice Theory», *Social Choice and Welfare*, p. 235-282.
- Allais, M. (1953): «Le Comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine», *Econometrica*, 21, p. 503-546.
- M. i O. Hagen (1979a): *Expected Utility Hypotheses and the Allais Paradox*, Dordrecht, Reidel, p. 27-145.
- Anscombe, F. J. i R. J. Aumann (1963): «A Definition of Subjective Probability», *Annals of Mathematical Statistics*, 34, p. 199-205.
- Arrow, K. J. (1959): «Rational Choice Functions and Orderings», *Econometrica*, 26, p. 121-127.
- i L. Hurwicz (1972): «An Optimality Criterion for Decision-Making Under Ignorance», a C. F. Carter i J. L. Ford (eds.), *Uncertainty and Expectations in Economics*.
- (1965): *Aspects of the Theory of Risk-Bearing*, Yrjö Jahnsson Lectures, Helsinki, Yrjö Jahnssonin Säätiö.
- Barberà, S. i M. Jackson (1988): «Maximin, Leximin and the Protective Criterion», *Journal of Economic Theory*, vol. 46, n.º 1, pp. 34-44.
- i P. Pattanaik (1982): «Extending an Order on a Set to the Power Set: Some Remarks on Kannai and Peleg's approach», *Journal Economic Theory*, vol. 32, n.º 1, pp. 185-191.
- (1986): «Falmagne, and the Rationalizability of Stochastic Choices in Terms of Random Orderings», *Econometrica*, vol. 54, n.º 3, p. 707-715.
- Bewley, T. F. (1986): «Knightian Decision Theory Part I», Cowles Foundation Discussion Paper n.º 807.
- Chipman, John S.; Leonid Hurwicz; Marcel K. Richter i Hugo F. Sonnenschein (1971): *Preferences, Utility, and Demand*, Harcourt Brace Jovanovich, Inc.
- Cohen, M. i J. Y. Jaffray (1980): «Rational Behaviour Under Complete Ignorance», *Econometrica*, 48, p. 1.281-1.299.
- Debreu, Gerard (1954): «Representation of a preference ordering by a numerical function», capítol XI a *Decision Processes*, editat per R. M. Thrall, C. H. Coombs i R. L. Davis, Nova York, John Wiley & Sons, Inc., p. 159-165.
- (1958): «Stochastic Choice and Cardinal Utility», *Econometrica*, 26, p. 440-444.
- Ellsberg, D. (1961): «Risk Ambiguity and the Savage Axioms», *Quarterly Journal of Economics*, 75, p. 643-669.

- Falmagne, J. C. (1978): «A Representation Theorem for Finite Random Scale Systems», *Journal of Math. Psychology*, 18, p. 52-72.
- Fishburn, P. (1970): *Utility Theory for Decision Making*, Nova York, John Wiley i Sons, reimprès per Krieger Press, Huntington, 1979.
- (1988): «Expected Utility: An Anniversary and a New Era», *Journal of Risk and Uncertainty*, 1, p. 267-283.
- Friedman, Milton i L. J. Savage (1948): «The Utility Analysis of Choices Involving Risk», *Journal of Political Economy*, 56, pp.279-304. Reimprès a *Readings in Price Theory*, editat per George J. Stigler i Kenneth E. Boulding, Homewood, Illinois, Richard D. Irwin, Inc., p. 57-96.
- Gilboa, I. (1989): «A Necessary But Insufficient Condition for the Stochastic Binary Choice Problem», Northwestern University, Evanston.
- Goldman, S. M. (1974): «Flexibility and the Demand for Money», *Journal of Economic Theory*, vol. 9, p. 203-222.
- Hammond, P. J. (1976): «Changing Tastes and Coherent Dynamic Choice», *Review of Economic Studies*, 43, p. 159-173.
- (1988b): «Consequentialist Foundations for Expected Utility», *Theorem and Decision*, 25, pp. 25-78.
- Harsanyi, J. C. (1953): «Cardinal Utility in Welfare Economics and in the Theory of Risk Taking», *Journal of Political Economy*, 61, p. 434-435.
- Houthakker, H. S. (1950): «Revealed Preference and the Utility Functions», *Economica*, N.S. 17, p. 159-174.
- Kannai, Y. i B. Peleg (1984): «A Note on the Extension of an Order on a Set to the Power Set», *Journal of Economic Theory*, 32, p. 172-175.
- Katzner, D. W. (1970): *Static Demand Theory*, Nova York, The Macmillan Company.
- Kelsey, D. (1988): «A Survey of Ignorance or Alternatives to Subjective Expected Utility», W. P. Churchill College, Cambridge.
- Kim, K. H. i I. Roush (1980): «Preferences on Subsets», *J. of Math. Psychology*, 21, p. 279-282.
- Knight, Frank H. (1921): *Risk, Uncertainty and Profit*, Boston, Houghton Mifflin.
- Koopmans, T. C. (1964): «On Flexibility of Future Preference», a *Human Judgements and Optimality*, editat per M. W. Shelly i G. L. Bryan, Nova York, John Wiley & Sons, Inc., p. 243-254.
- Krantz, D. H., R. D. Luce, P. Suppes i A. Tversky (1971): *Foundations of Measurement*, vol. 1, Nova York, Academic Press.
- Kreps, D. M. (1979): «A Representation Theorem for Preference for Flexibility», *Econometrica*, vol. 47, n.º 3.
- (1988): *Notes on the Theory of Choice*, Underground Classics in Economics, Westview Press.
- Luce, R. D. i H. Raiffa (1957): «Games and Decisions», Nova York, John Wiley & Sons.
- Machina, Mark J. (1982a): «Expected Utility Analysis without the Independence Axiom», *Econometrica*, 50, p. 277-323.
- (1983): «The Economic Theory of Individual Behaviour Toward Risk: Theory, Evidence and New Directions», Technical Report, n.º 433, Institute for Mathematical Studies in the Social Sciences, Stanford University.
- (1985): «Stochastic Choice Functions Generated from Deterministic Preferences over Lotteries», *The Economic Journal*, 95, p. 575-594.
- (1987): «Choice under Uncertainty: Problems Solved and Unsolved», *Economic Perspectives*, vol. 1, p. 121-154.
- Marschak, J. (1959): «Binary Choice Constraints and Random Utility Indicators», a *Mathematical Methods in the Social Science*, editat per K. J. Arrow, S. Karlin i P. Suppes; Stanford, California, Stanford University Press, p. 312-329.
- Mas-Colell, A. (1977): «On the Continuous Representation of Preorders», *International Economic Review*, vol. 18, n.º 2.
- Milnor, J. (1954): «Games Against Nature in Decision Processes», R. M. Thrall, C. H. Coombs i R. L. Davis (eds.), Nova York, Chapman & Hall.

- Neumann, J. von i O. Morgenstern (1953): «Theory of Games and Economic Behaviour», Princeton, Princeton University Press.
- Nitzan, S. I. i P. K. Pattanaik (1984): «Median Based Extension of an Ordering Over a Set to the Power Set: An Axiomatic Characterization», *Journal of Economic Theory*, 34, p. 252-261.
- Pattanaik, P. i B. Peleg (1984): «An Axiomatic Characterization of the Lexicographic Maximin Extension of an Ordering over a Set to the Power Set», *Social Choice and Welfare*, 1, p. 113-122.
- Plott, C. R. (1973): «Path Independence, Rationality, and Choice», *Econometrica*, 41, p. 1.075-1.091.
- Richter, Marcel K. (1971): «Rational Choice», a *Preferences, Utilities, and Demand*, editat per Chipman, John S., Leonid Hurwicz, Marcel K. Richter i Hugo F. Sonnenschein, 1971, Harcourt Brace Jovanovich, Inc.
- Savege, L. J. (1954): *Foundations of Statistics*, Wiley.
- Sen, A. K. (1977): «Social Choice Theory: A re-examination», *Econometrica*, 45, p. 53-89.
- Sonnenschein, H. F. (1971): «On the Continuity of Demand Functions Which Satisfy the Strong Axiom of Revealed Preference», a *Preferences, Utilities, and Demand*, editat per Chipman, John S., Leonid Hurwicz, Marcel K. Richter i H. F. Sonnenschein, Harcourt Brace Jovanovich, Inc.
- Strotz, R. H. (1956): «Myopia and Inconsistency in Dynamic Utility Maximization», *The Review of Economic Studies*, p. 165-189.
- (1957): «The Empirical Implications of a Utility Tree», *Econometrica*, 25, p. 269-280.
- Suzumura, K. (1976): «Rational Choice and Revealed Preference», *The Review of Economic Studies*, vol. XLIII, p. 149-158.
- Tversky, A. i D. Kahneman (1974): «Judgement under Uncertainty: Heuristics and Biases», *Science*, 185, p. 1.124-1.131, reimprès a Kahneman, Slovic i Tversky, 1982.
- i S. Sattath (1976): «Unite and Conquer: a Multiplicative Inequality for Choice Probabilities», *Econometrica*, vol. 44, n.º 1.