

Commenti sulla relazione di Rustichini

di Marco Dardi

Le linee di ricerca che Rustichini riunisce sotto la formula «teoria delle decisioni post-bayesiana» hanno un punto di partenza comune nel riesame dei presupposti logico-epistemici delle teorie del comportamento in condizioni d'incertezza. Questa frontiera interdisciplinare fra teoria economica e logica epistemica, aperta solo da poco tempo a seguito di sviluppi interni della teoria dei giochi, è ancora relativamente poco esplorata, e forse anche un po' remota rispetto agli immediati interessi di ricerca di molti economisti. In questi commenti darò un'interpretazione del lavoro che vi si svolge così come lo vedo da non specialista, con un interesse rivolto, più che agli aspetti tecnici, alle prospettive di indebolimento dell'ipotesi di razionalità che sta alla base delle versioni più diffuse di teoria dei comportamenti economici.

Da quando l'assiomatizzazione è diventata metodo di lavoro predominante in teoria economica (cioè all'incirca dagli anni '50), su una cosa almeno sembra che sia stata fatta chiarezza: le proposizioni della teoria sono dimostrabilmente valide solo in quanto si riferiscano non al mondo ma a modelli del mondo, cioè a strutture che verificano un insieme di proposizioni specifiche - gli assiomi del caso - più le proposizioni generali della matematica. Sviluppi successivi della teoria, in particolare quelli legati alla teoria dei giochi, ci hanno abituato a includere nei nostri modelli anche rappresentazioni di stati mentali: per parlare di agenti che decidono razionalmente, infatti, non basta attribuirgli preferenze e poteri decisionali, si deve anche specificare di quali conoscenze e capacità di ragionamento a partire dalle conoscenze dispongono al momento di decidere. Inevitabile quindi che la teoria economica sia venuta a trovarsi in un'intersezione con discipline (in senso lato) della mente, in particolare, per quanto qui ci riguarda, l'intelligenza artificiale e la logica epistemica. Ma l'estensione del dominio della teoria agli eventi mentali suscita immediatamente il so-

Ringrazio, esentandoli da ogni corresponsabilità, Vittorio Emanuele Ferrante, Antonio Gay e Salvatore Modica per osservazioni su una stesura precedente.

spetto di una difficoltà fondamentale, derivante dallo stesso metodo assiomatico che ci ha guidato fino a questo punto. Costruire un sistema di assiomi, insistiamo, significa proiettare ipotesi specifiche su oggetti appartenenti alla classe generale degli oggetti matematici, cioè oggetti che per loro costruzione non possono contraddire le leggi della logica. Ora, finché le ipotesi specifiche con cui lavoriamo si riferiscono a aspetti oggettivi del mondo, possiamo discutere quanto vogliamo la loro verosimiglianza, ma non c'è motivo di mettere in dubbio la struttura matematica del modello: per farlo, dovremmo aver ragione di pensare che nel mondo oggettivo le leggi della logica sono contraddette, il che ci lascerebbe del tutto disarmati nell'impresa di interpretazione razionale della realtà. Quando (però) lavoriamo con ipotesi relative a aspetti soggettivi, agli stati mentali degli agenti di un modello, questa fiducia nella validità delle leggi logiche non è più pacifica. Al di là della verosimiglianza o meno dei nostri assiomi su conoscenze e capacità raziocinante, c'è il problema più fondamentale che qualunque modello matematico della mente impone - per costruzione - una coerenza logica che normalmente gli individui, per quanto «intelligenti», non hanno. Che le leggi (se di leggi si tratta) secondo cui la mente associa ed elabora nozioni e credenze non siano riducibili a specificazioni particolari delle leggi della logica, è un fatto su cui c'è generalmente consenso¹.

Così enunciato, il problema è probabilmente troppo vago e «filosofico» per sapere cosa farne. È alla riformulazione dei modelli di comportamento economico in termini di sistemi logico-modali che dobbiamo una decisiva chiarificazione sulla sua natura, e qualche indicazione su come provare a superarlo. L'idea fondamentale sviluppata in questa letteratura è che un modello di un problema decisionale che includa fra l'altro le conoscenze di cui è dotato il decisore ha

¹ Per essere più chiari, non si tratta di un problema risolvibile a buon mercato con un «come se» alla Friedman. È ragionevole che un modello della condotta di gioco di un giocatore di biliardo si basi su un sistema meccanico anche se il giocatore non possiede nessuna nozione di meccanica, perché la sua abilità consiste appunto nella capacità di sfruttare le leggi della meccanica come se le conoscesse. Ma semmai ci interessasse un modello della mente del giocatore di biliardo, non potremmo più ignorare che le sue procedure di ragionamento non sono le stesse di un matematico che imposta e risolve un sistema di equazioni. Anche se i risultati a cui il giocatore arriva sono molto vicini a quelli del matematico, il giocatore ci è arrivato per altre vie, e l'oggetto del modello dovrebbe essere proprio la differenza fra le une e le altre. Semmai una via d'uscita dal problema potrebbe consistere nell'attribuire ai modelli una funzione puramente prescrittiva e non predittiva: ma non tutti gli economisti sembrano disposti ad accettare un'interpretazione così austera delle teorie formalizzate.

dietro di sé un sistema logico costituito da un insieme di enunciati relativi sia ai fatti del modello, sia al fatto che l'agente li conosca o meno (gli enunciati modali), chiuso rispetto al calcolo proposizionale. Il sistema logico-modale determinato da un modello o classe di modelli rende quindi esplicita la struttura sintattica completa di cui i modelli costituiscono interpretazioni. Sono possibili vari sistemi logico-modali a seconda di quali e quanti enunciati relativi ai fatti epistemici vogliamo che siano verificati: i modelli utilizzati dagli economisti, a partire da Savage fino a tempi recentissimi, rientrano tutti nella classe che verifica il sistema logico-modale noto come (S5). In questi modelli, come ricorda Rustichini nel paragrafo 2 della sua relazione, i possibili eventi epistemici sono rappresentati da una funzione $P: \Omega \rightarrow 2^{\Omega}$ che associa a ogni stato di cose nell'insieme degli stati possibili Ω l'evento di Ω che rappresenta l'informazione massima che l'agente ha in quello stato; funzione il cui codominio (l'insieme delle massime informazioni in tutti i possibili stati) costituisce una partizione veridica di Ω stesso, intendendo per veridica che lo stato effettivo verifica sempre l'informazione, $\omega \in P(\omega)$ per qualunque $\omega \in \Omega$. L'aver esplicitato la struttura sintattica di questa classe di modelli ha permesso di portare alla luce un'implicazione che ha fatto molto discutere e che si collega strettamente al problema di adeguatezza dei modelli assiomatici da cui siamo partiti. Fra gli enunciati necessariamente veri in S5 ce n'è infatti uno noto come «onniscienza logica», che equivale ad asserire che l'agente conosce tutti i teoremi del sistema, cioè tutti gli enunciati che all'interno del sistema possono esser dimostrati come certamente veri con i metodi usuali della logica. Quale che sia la plausibilità degli altri enunciati epistemici del sistema, l'inevitabilità dell'onniscienza logica sembra dare un corpo ben definito ai vaghi dubbi che abbiamo enunciato sopra: la mente umana questo dono di conoscere tutte le implicazioni logiche delle proprie conoscenze non ce l'ha - deve lavorare per scoprirle, e la sua capacità di lavoro è fisicamente limitata.

Non mi spingerei fino a sostenere che tutte le ragioni per cui la teoria delle decisioni lascia insoddisfatti si possano ridurre a questa implicazione di onniscienza logica, ma molte certamente sì, anche se non sempre ciò emerge con la chiarezza propria dei modelli in cui le ipotesi epistemiche sono formulate esplicitamente. Quel po' di evidenza empirica che ognuno di noi può mettere insieme per via d'introspezione o di colloquio con altri mostra che molti casi tipici di fallimento delle predizioni della teoria delle decisioni si spiegano non con l'irragionevolezza delle sue ipotesi, ma con l'incapacità o la non volontà del decisore di sviluppare tutte le inferenze del caso. Prendiamo due esempi celebri, ampiamente utilizzati come evidenze contrarie

all'ipotesi dell'utilità attesa: il caso di Allais, di scelta fra due coppie di lotterie, di cui la seconda coppia è ricavabile dalla prima per sostituzione di uno stesso esito in entrambe le lotterie; e l'esperimento Tversky e Kahneman, di scelta fra due programmi sanitari le cui prospettive di successo vengono presentate una volta in termini di speranza di vita, e un'altra in termini di speranza di morte. In entrambi i casi la mia esperienza (in test condotti con studenti) è che coloro che hanno espresso preferenze incompatibili con l'utilità attesa, messi di fronte al fatto, riconoscono di aver sbagliato perché non si sono accorti di certe equivalenze - nel caso di Allais, che c'è una terza coppia di lotterie, stocasticamente equivalente alla seconda, che rende evidente la derivazione dalla prima coppia per sostituzione di un esito; nel caso di Tversky e Kahneman (molto più impressionante dell'altro, perché qui l'equivalenza è talmente trasparente che il non accorgersene fa pensare quasi all'esistenza di un tabù) che dal punto di vista delle distribuzioni vivi/morti attese le due coppie di scelte sono assolutamente identiche. Se si fossero accorti di queste equivalenze, le loro scelte sarebbero state diverse, e probabilmente conformi all'utilità attesa (infatti, tutti i soggetti dell'esperimento concordavano nel ritenere ragionevoli gli assiomi dell'utilità attesa). Se quest'ultima è stata contraddetta, è per l'incapacità di tradurre alternative di scelta formulate verbalmente in punti di un insieme di distribuzioni di probabilità, cioè in definitiva per una visione non chiara dell'insieme di scelta proposto. È allora il caso di chiedersi se esempi come questi, e più in generale tutti quelli di violazione dell'utilità attesa per l'azione di un qualche framing effect, disprovino gli assiomi dell'utilità attesa o non piuttosto l'ipotesi tacita che il soggetto sia dotato della capacità di identificare perfettamente gli oggetti di scelta. I passaggi logici necessari per convertire alternative variamente rappresentate in punti di un dato insieme di supporto delle preferenze possono avere un costo, in termini vuoi di lavoro mentale vuoi di superamento di barriere psicologiche, che i soggetti di esperimento non si sentono di sostenere. Se quindi vogliamo costruire modelli di decisione sotto incertezza più verosimili, non è soltanto sugli assiomi dell'utilità attesa che dobbiamo lavorare, ma anche sulla struttura di conoscenze e capacità logiche che il soggetto riesce ad attivare per capire il problema di decisione che gli sta di fronte. Eccoci quindi di nuovo all'esigenza di capire meglio la struttura di sistemi di conoscenze e credenze incompleti dal punto di vista della chiusura logica, o eventualmente anche logicamente incoerenti.

Si può discutere se per andare avanti in questa direzione non sa-

rebbe più sicuro affidarsi alla guida della psicologia cognitiva piuttosto che della logica epistemica. Poiché però si dovrà comunque passare per modelli derivanti da indebolimenti di S5 o da sistemi logico-modali alternativi, dalla logica possiamo almeno attenderci qualche indicazione sul tipo di strutture utilizzabili. La logica ci mostra, in primo luogo, quanto sia difficile liberarsi del tutto dall'onniscienza logica se non si è disposti a rivoluzionare la nostra pratica modellistica in aspetti fondamentali: è infatti ormai acquisito [Fagin, Halpern, Moses e Vardi 1995, 312] che l'onniscienza logica è insita nel fatto stesso di rappresentare gli eventi epistemici mediante insiemi di stati di cose possibili legati allo stato effettivo dalla parziale ignoranza del soggetto – che è appunto la rappresentazione tuttora canonica fra gli economisti. Sono possibili però attenuazioni significative, e Rustichini qui sintetizza i risultati del lavoro svolto su una di esse insieme con Salvatore Modica e altri. Si tratta di elaborare un sistema modale nel quale possa esservi inconsapevolezza – nel senso non solo di non conoscenza, ma di non sapere di non avere conoscenza – di enunciati (fatti) che fanno parte del sistema stesso, e quindi di tutte le inferenze che se ne possono trarre [cfr. Modica e Rustichini 1993; 1994]. Questa non è naturalmente l'unica strada possibile. Un'alternativa molto esplorata in tempi recenti sono i sistemi modali determinati da modelli epistemici veridici ma non partizionali, di cui Rustichini dà un esempio nel paragrafo 2 della sua relazione². Anche in questi modelli vi sono fatti di cui il soggetto può non avere consapevolezza, ma, a differenza che nel sistema di Modica e Rustichini, l'inconsapevolezza ha qui l'inconveniente di poter essere asimmetrica e di non essere a sua volta (come suggerirebbe l'intuizione) inconsapevole: possono esservi stati di cose in cui il soggetto è, poniamo, inconsapevole di un enunciato a ma consapevole di $non-a$, e in cui inoltre sa di essere inconsapevole di a , cosa che sembra contraddire l'idea stessa di inconsapevolezza. D'altra parte il sistema di Modica e Rustichini conserva la struttura partizionale della conoscenza e l'onniscienza logica proprie della modellistica tradizionale, benché entrambe riferite a sottoinsiemi degli enunciati che compongono il sistema, cioè non agli stati quali effettivamente si danno, ma agli stati quali sono percepiti da un soggetto con percezione inconsapevolmente ridotta. Si direbbe

² Andrebbero ricordati anche i modelli partizionali ma non veridici [cfr., per es., Ferrante 1991].

che il decisore rappresentato dal modello di Modica e Rustichini sia un individuo di potenza logica illimitata ma di limitata capacità immaginativa, un individuo quindi fallibile, ma solo su ciò che proprio non gli viene in mente.

Nonostante le critiche di Rustichini, personalmente continuo a trovare interessante anche la strada dei modelli non partizionali, e vorrei concludere queste note spezzando una lancia in loro favore. Se consideriamo che l'inconsapevolezza, o comunque il non tener conto di un fatto possibile, né in positivo né in negativo, può esser dovuta non soltanto a un limite dell'immaginazione ma anche alla deliberata applicazione di una strategia di semplificazione di un problema decisionale (secondo il principio «una cosa alla volta»), l'asimmetria dell'inconsapevolezza, o il fatto di sapere di essere inconsapevoli, non appaiono poi eventualità così paradossali. Per di più, mi sembra che si possa sostenere che i modelli non partizionali presuppongono l'esistenza di limiti alla capacità inferenziale del decisore in modo forse più sottile e fondamentale di qualunque modello partizionale. Un agente che prende le sue decisioni in base a una struttura informativa non partizionale è infatti – quasi per definizione – un individuo che rinuncia a esplorare tutte le nozioni implicite nella struttura informativa che gli è assegnata, in un senso che cercherò di chiarire con il ragionamento che segue (riferito alle sole strutture informative veridiche).

Se Ω rappresenta l'insieme degli stati di cose possibili, indichiamo con Ω^* l'insieme degli stati esteso fino a includervi anche le possibili strutture informative del decisore. Posto che \mathcal{P} rappresenti l'insieme di tutte le possibili strutture informative veridiche su Ω (cioè l'insieme delle funzioni di Ω in 2^I tali che l'argomento appartiene sempre all'immagine), sarà $\Omega^* = \Omega \times \mathcal{P}$. Un decisore dotato di una struttura informativa $P \in \mathcal{P}$ qualunque, può dedurre – purché ci pensi, e sia capace di ragionare – dall'informazione $\bar{P}(\omega)$ che ha in ogni stato $\omega \in \Omega$, un'informazione più completa, collocata in Ω^* , che indichiamo con $P^*(\omega, P)$ ed è data da

$$P^*(\omega, \bar{P}) = \{(\omega', P') \in \Omega^*: P'(\omega') = \bar{P}(\omega)\}$$

La struttura informativa definitiva di questo individuo «riflessivo» è quindi rappresentata non più da \bar{P} ma da P^* , funzione con dominio nel sottoinsieme di Ω^* formato dall'unione delle informazioni $P^*(\omega, P)$ estesa a tutti gli $\omega \in \Omega$, e le cui immagini, come si può verificare immediatamente, costituiscono una partizione veridica del dominio –

qualunque sia la struttura della \bar{P} di partenza¹. Ecco in che senso un decisore che «si contenta» di una struttura informativa non partizionale si può considerare un individuo che o non riflette o non è capace di ragionare sulla propria informazione: se lo facesse, lascerebbe la propria struttura informativa per un'altra, più completa e inevitabilmente partizionale.

Se Rustichini ha ragione di chiamare «post-bayesiano» il suo modello di consapevolezza (ma non capacità inferenziale) limitata, potremmo forse altrettanto legittimamente chiamare «anti-bayesiani» i modelli non partizionali. Per quanto detto sopra, infatti, un decisore con informazione non partizionale è uno che, al contrario del decisore bayesiano, non usa tutta l'informazione che ha a disposizione. Il decisore di Rustichini invece la usa, e differisce da un vero bayesiano solo perché non sempre sa che cosa non sa. Ci troviamo così di fronte a due possibili vie d'uscita dalla teoria bayesiana – una che procede per piccoli passi, l'altra che tenta subito grandi rotture. Da entrambe possiamo aspettarci novità interessanti, e credo che anche i non specialisti farebbero bene a non perdere di vista quello che si fa su questa frontiera apparentemente così lontana.

Riferimenti bibliografici

- Fagin, R., Halpern, J.Y., Moses, Y. e Vardi, M.Y. (1995), *Reasoning About Knowledge*, Cambridge (Mass.), The MIT Press.
- Ferrante, V. (1991), *Doxastic Information Partitions, and an Anti-No-Trade Theorem*, in «Metroeconomica», 42, pp. 47-61.
- Modica, S. e Rustichini, A. (1993), *Unawareness: A Formal Theory of Unforeseen Contingencies*, Part II, in «CORE Discussion Paper», n. 9404.
- (1994), *Awareness and Partitional Information Structures*, in «Theory and Decision», 37, pp. 107-124.

¹ Questo, si noti, nel caso che il decisore non conosca \bar{P} e ritenga possibili tutte le strutture in \mathcal{P} . Se invece il decisore conosce \bar{P} , Ω rimane l'insieme degli stati rilevanti, e su di esso definiamo, analogamente a come abbiamo fatto nel testo, la funzione $P^*(\omega) = \{\omega' : \bar{P}(\omega') = \bar{P}(\omega)\}$, a rappresentare l'informazione più completa che il decisore può raggiungere decodificando la propria struttura informativa. È facile dimostrare che $P^* = \bar{P}$ se e solo se \bar{P} è partizionale: se \bar{P} non lo è, allora P^* è una struttura partizionale più ricca di \bar{P} , nel senso che in ogni stato ω risulta $P^*(\omega) \subseteq \bar{P}(\omega)$, con inclusione stretta per almeno un $\omega \in \Omega$. In altri termini, un decisore che conosca la propria struttura informativa, se questa non è partizionale, può aumentare la propria informazione semplicemente ragionando su ciò che sa. Solo le strutture partizionali sono definitive.