

Teoria dei giochi ed evoluzione delle norme morali

ROBERTO FESTA

Dipartimento di filosofia

Università di Trieste

festa@units.it

ABSTRACT

Mathematical game theory - developed starting from the publication of *The Theory of Games and Economic Behavior* (1944), by John von Neumann and Oskar Morgenstern - aims to outline an ideal model of behaviour of rational agents involved in some interaction with other rational agents. For this reason, game theory has immediately attracted the attention of philosophers dealing with practical rationality and, since the fifties, has been applied to the analysis of several issues concerning ethics and philosophy of politics. Here we will focus on one of the most interesting applications of game theory to ethical-political inquiry, i.e., with the game theoretic analysis of some problems related to the evolution of moral norms. Firstly, we will provide a short outline of the development of game theory, which has led to the formulation of a plurality of different game theories. It will be shown that such theories can be classified in two main groups: rationalistic game theories -in two different versions: classical and epistemic - and evolutionary game theories. Moreover, some basic elements of classical game theory will be introduced and the key ideas of epistemic and evolutionary game theories will be illustrated. Afterwards, the main approaches developed within the ethical-political applications of game theory will be shortly described. Finally, some results obtained in the last twenty years by the researchers who have analysed the evolution of moral norms by the conceptual tools of evolutionary and epistemic game theories, will be examined.

0. La teoria dei giochi è certamente, fra le teorie matematiche inventate nel corso del Novecento, quella che ha suscitato il maggiore interesse al di fuori del ristretto ambito dei matematici di professione. Le ragioni di questo interesse sono in gran parte riconducibili al fatto che, diversamente dalla maggior parte delle teorie matematiche, la teoria dei giochi non riguarda numeri, figure geometriche o altre entità astratte, ma si propone di delineare un modello ideale del comportamento di agenti razionali coinvolti in qualche interazione con altri agenti razionali. Non stupisce, quindi, che già negli anni immediatamente successivi alla sua scoperta, che si può convenzionalmente far risalire al 1944, anno della pubblicazione di *The Theory of Games and Economic Behavior* ad opera di John von Neumann e Oskar Morgenstern, la teoria dei giochi abbia attratto l'attenzione sia degli scienziati sociali sia dei filosofi che si occupavano del comportamento umano e della razionalità pratica. In particolare, la teoria dei giochi è stata applicata fin dagli anni

cinquanta all'analisi di diverse questioni concernenti l'etica e la filosofia della politica. Ci occuperemo qui di una delle applicazioni più interessanti della teoria dei giochi alla riflessione etico-politica, e cioè dell'analisi giochistica di alcuni problemi relativi all'evoluzione delle norme morali.¹

Nel *primo* paragrafo verranno dati alcuni ragguagli sullo sviluppo impetuoso della teoria dei giochi nei suoi pochi decenni di vita, un sviluppo che, a partire dalla teoria classica dei giochi di von Neumann e Morgenstern, ha condotto alla formulazione di una pluralità di teorie dei giochi molto diverse tra loro. Si mostrerà che tali teorie possono venire classificate in due gruppi principali: teorie razionalistiche, in differenti versioni, classiche ed epistemiche, e teorie evoluzionistiche dei giochi. Successivamente si introdurranno alcuni elementi essenziali della teoria classica dei giochi e si illustreranno le idee di fondo delle teorie epistemiche ed evoluzionistiche. I principali approcci sviluppati nelle applicazioni etico-politiche della teoria dei giochi saranno brevemente ricordati nel *secondo* paragrafo. Infine nel *terzo* paragrafo si esamineranno alcuni risultati ottenuti dagli studiosi che, nel corso dell'ultimo ventennio, hanno analizzato l'evoluzione delle norme morali utilizzando gli strumenti concettuali delle teorie evoluzionistiche ed epistemiche dei giochi.

1. *Teoria dei giochi*

1.1. *Pluralità delle teorie dei giochi*

La teoria dei giochi può venire considerata come una branca della teoria della scelta razionale. Quest'ultima è una teoria di carattere *normativo* che si propone di determinare le scelte che un individuo razionale *dovrebbe* compiere, nelle più diverse situazioni, dati i suoi fini e le sue opinioni.² Possiamo distinguere due tipi fondamentali di scelte: le scelte di agenti *isolati* alle prese con un ambiente di cui non conoscono con certezza tutte le caratteristiche rilevanti e quelle di agenti impegnati in un'*interazione strategica* con altri agenti, cioè in un'interazione in cui devono agire sulla base delle proprie aspettative circa le azioni altrui. L'analisi sistematica di questi due tipi di scelte dà origine a due fondamentali aree della teoria della scelta razionale: la teoria delle decisioni, che si occupa del primo genere di scelte, e la teoria dei giochi, che si occupa del secondo. Il termine *teoria dei giochi* è legato al

¹ Traduciamo con l'aggettivo 'giochistico' il termine inglese *game theoretic*.

² Per un'introduzione alla teoria della scelta razionale si veda, per esempio, Hargreaves Heap *et al.* (1996).

fatto che gli scacchi, i giochi con le carte, la morra cinese e altri giochi da tavolo rientrano nell'ambito della teoria, che comprende però anche molti altri tipi di interazioni strategiche. Sono giochi, infatti, anche un accordo commerciale, una trattativa politica o sindacale, una battaglia o una conferenza internazionale: sono, insomma, giochi tutte le situazioni in cui due individui, o, più in generale, due 'attori', ove questo termine può applicarsi non solo a persone, ma anche a organizzazioni, imprese, governi e così via, sono coinvolti in qualche interazione strategica.

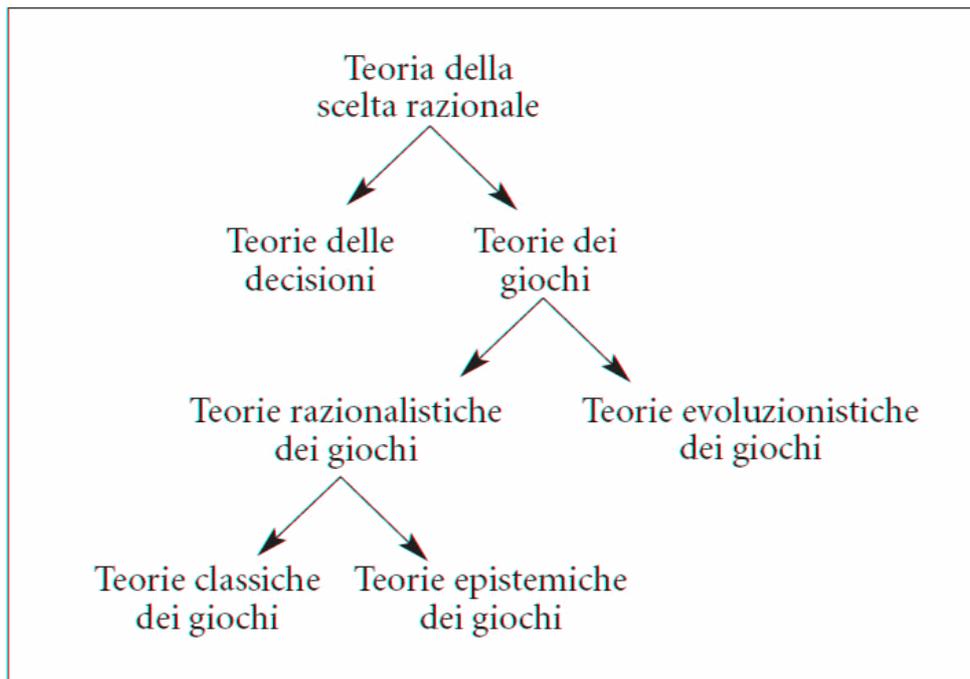
La prima esposizione organica della teoria dei giochi si deve a John von Neumann e Oskar Morgenstern, matematico il primo, economista il secondo, che nel 1944 pubblicarono l'influente libro *The Theory of Games and Economic Behavior*. Lo scopo della loro teoria era quello di individuare la *soluzione* di un gioco, cioè la combinazione delle *strategie ottimali* che dovrebbero essere adottate da ciascuno dei giocatori. La natura generale e astratta della nozione di gioco da loro introdotta ha consentito di applicare la teoria dei giochi nei campi più disparati, dalle scienze sociali e all'economia, dagli studi strategici all'etica e alla filosofia della politica. A questo proposito occorre osservare che la teoria dei giochi di von Neumann e Morgenstern ha trovato, a dispetto del suo carattere normativo, numerose *applicazioni descrittive*, nella spiegazione e nella previsione di svariati comportamenti umani; tali applicazioni si basano sul presupposto che gli agenti impegnati nelle interazioni sociali sono almeno approssimativamente razionali.

Con riferimento alle versioni tradizionali della teoria dei giochi, che presentano il comportamento dei giocatori come il risultato di scelte razionali, potremmo parlare di *teorie razionalistiche* dei giochi. Nell'ambito delle teorie razionalistiche dei giochi conviene distinguere tra teorie classiche e teorie epistemiche. Le *teorie classiche* condividono il presupposto fondamentale della teoria di von Neumann e Morgenstern, secondo il quale la razionalità individuale, assieme alla conoscenza comune delle preferenze e della razionalità altrui, è *sufficiente* a individuare la soluzione di ogni gioco. Ci sono però buoni motivi per dubitare dell'ottimismo razionalistico delle teorie classiche. Uno di questi è legato alla circostanza che, nelle nostre interazioni sociali, non sempre abbiamo a che fare con agenti *perfettamente* razionali. Queste considerazioni hanno ispirato lo sviluppo di alcune versioni della teoria dei giochi in grado di tenere conto delle effettive *condizioni epistemiche* in cui si svolgono le interazioni sociali. Queste teorie, che potremmo denominare *teorie epistemiche dei giochi*, hanno trovato interessanti applicazioni nelle scienze sociali; per esempio la teoria epistemica sviluppata da Cristina Bicchieri (1993 e 2006) è stata applicata con successo nell'analisi delle norme sociali.

A partire dalla fine degli anni settanta le teorie razionalistiche sono state affiancate dalle *teorie evuzionistiche dei giochi*, di carattere squisitamente

descrittivo, sviluppate a partire dalle ricerche del biologo John Maynard Smith (1982). Tali teorie non concepiscono le strategie adottate dai giocatori come il prodotto di scelte razionali, bensì come *schemi di comportamento* acquisiti mediante i meccanismi della trasmissione genetica o culturale. Dopo il 1982 le teorie evoluzionistiche dei giochi sono state ampiamente impiegate nelle scienze sociali, per esempio nell'analisi dell'evoluzione delle norme e delle istituzioni sociali.

Le relazioni tra la teoria della scelta razionale e le diverse versioni della teoria dei giochi sono rappresentate nel seguente grafico.



1.2. Teoria classica dei giochi: concetti e principi fondamentali

GIOCHI A DUE. I giochi più semplici sono i *giochi a due (giocatori)*. Il risultato di un gioco a due è determinato solo dalle *mosse* adottate dai giocatori; in particolare, se i giocatori hanno a disposizione due mosse ciascuno, il risultato del gioco corrisponderà a una delle quattro possibili combinazioni delle loro mosse. Possiamo quindi rappresentare il gioco come una matrice con due righe e due colonne, come nella seguente figura relativa al gioco 1.

		C	
		c_1	c_2
R	r_1	$\begin{matrix} -1 & \\ 1 & \end{matrix}$	$\begin{matrix} 3 & \\ 2 & \end{matrix}$
	r_2	$\begin{matrix} 2 & \\ 2 & \end{matrix}$	$\begin{matrix} 0 & \\ 0 & \end{matrix}$

Gioco 1. Due giocatori, entrambi con due mosse a disposizione

Nel gioco 1 i giocatori sono rappresentati dalle lettere R (per ‘Riga’) e C (per ‘Colonna’); le righe r_1 ed r_2 , della matrice rappresentano le mosse di R , mentre le colonne c_1 e c_2 rappresentano le mosse di C . Ogni cella della matrice corrisponde a un *risultato* del gioco. Se, per esempio, R sceglie r_1 e C sceglie c_2 , il risultato corrisponderà alla cella in alto a destra, che possiamo indicare con (r_1, c_2) .

STRATEGIE PURE E STRATEGIE MISTE. Diciamo che un giocatore adotta una *strategia pura* quando sceglie direttamente una delle mosse a sua disposizione; diciamo invece che adotta una *strategia mista* quando la sua scelta viene attuata facendo ricorso a qualche procedimento di natura probabilistica. Per esempio, una strategia mista può consistere nel lancio di una moneta, seguito dall’attuazione di r_1 nel caso esca testa e di r_2 nel caso esca croce.

LA STRUTTURA DEI PAYOFF. Le *preferenze* di ciascun giocatore tra i possibili risultati del gioco possono venire rappresentate mediante opportune misure di *utilità*. L’utilità che un giocatore attribuisce a un determinato risultato viene normalmente indicata come il *payoff* di quel risultato.³ I payoff vengono talvolta espressi in termini monetari, per esempio in dollari; tuttavia, conviene pensare ai payoff in termini più generali, come a una misura del ‘beneficio percepito’ che i giocatori associano ai possibili risultati di un gioco. Ogni cella include due numeri che rappresentano i *payoff* ottenuti dai giocatori quando il risultato del gioco corrisponde a quella cella: il numero in basso a sinistra indica il payoff di R e quello in alto a destra il payoff di C . Per esempio, nella matrice del gioco 1 il risultato corrispondente alla cella (r_1, c_2) assegna un payoff pari a 2 a R e un payoff pari a 3 a C .

³ Il termine inglese payoff viene talvolta tradotto con ‘ricompensa’, ‘premio’, ‘incentivo’ o ‘vincita’. Tuttavia, dato che i payoff possono essere negativi preferiamo mantenere il termine inglese originale, che può riferirsi sia alle ‘vincite’ sia alle ‘perdite’ dei giocatori.

Vi è un'enorme varietà di tipi di gioco, che possono essere classificati sulla base di diversi criteri. Un importante criterio di classificazione viene formulato con riferimento alla *struttura dei payoff* di un gioco. Tale struttura è determinata dalle *preferenze* dei giocatori circa i possibili risultati del gioco o, in termini più precisi, dal loro *ordinamento di preferenza* sull'insieme dei risultati. In alcuni casi i giocatori avranno lo stesso ordinamento di preferenza; in altri casi avranno ordinamenti parzialmente coincidenti o anche totalmente opposti. Per riferirci ai giochi caratterizzati dal fatto che i giocatori condividono lo stesso ordinamento di preferenza parleremo di *giochi con totale coincidenza di interessi* o, più brevemente, di *giochi puramente cooperativi*. Un esempio di questo genere di giochi è il gioco 2 rappresentato nella seguente matrice.

		C	
		c_1	c_2
R	r_1	20 1	10 0
	r_2	30 2	40 3

Gioco 2. Gioco con totale coincidenza di interessi

Si noti, infatti, che nel gioco 2 il risultato preferito entrambi giocatori è (r_2, c_2) , seguito da (r_2, c_1) , da (r_1, c_1) e, infine, da (r_1, c_2) , che è il risultato peggiore per entrambi.

Il caso opposto a quello dei giochi puramente cooperativi è rappresentato dai *giochi con totale conflitto di interessi* o, più brevemente, dai *giochi di puro conflitto*.⁴ Un esempio di questo genere di giochi è il gioco 3 rappresentato nella seguente matrice.

⁴ Il termine 'giochi di puro conflitto' si deve a Schelling (1960/1980).

		C	
		c_1	c_2
R	r_1	30 1	10 3
	r_2	40 0	20 2

Gioco 3. Gioco con totale conflitto di interessi

Nel gioco 3 i giocatori hanno preferenze esattamente opposte: per esempio il risultato preferito da R corrisponde al risultato peggiore per C , e viceversa.⁵

Tra i casi estremi costituiti dai giochi puramente cooperativi e da quelli di puro conflitto, vi è un ampio spettro di casi intermedi rappresentato dai giochi in cui gli ordinamenti di preferenza dei giocatori sono parzialmente coincidenti; con riferimento a questo genere di giochi si può parlare di *giochi con parziale coincidenza di interessi* o, più brevemente, di *giochi parzialmente cooperativi*. Naturalmente si potrebbe parlare anche, come faremo talvolta nel seguito, di *giochi con parziale conflitto di interessi* o *giochi parzialmente conflittuali*. Un esempio di questo genere di giochi è il gioco 4 rappresentato nella seguente matrice.

		C	
		c_1	c_2
R	r_1	30 0	20 2
	r_2	10 1	30 2

Gioco 4. Parziale coincidenza di interessi

⁵ Nella letteratura, i giochi di puro conflitto vengono spesso chiamati *giochi a somma costante* o anche *giochi a somma zero*, con riferimento alla somma dei payoff dei giocatori in ogni cella. Questa terminologia può tuttavia risultare fuorviante, poiché ciò che davvero importa non è la somma numerica dei payoff bensì il loro ordinamento.

Come si può facilmente vedere dalla matrice, nel gioco 4 le preferenze dei giocatori coincidono rispetto ad alcune coppie di risultati, ma non rispetto ad altre. Per esempio, entrambi i giocatori preferiscono (r_2, c_2) a (r_2, c_1) , ma R preferisce (r_2, c_1) a (r_1, c_1) mentre C preferisce il secondo risultato al primo.

È importante notare i giochi parzialmente cooperativi includono gran parte delle comuni interazioni sociali e, ciò che qui più ci interessa, gran parte delle interazioni considerate nelle analisi giostiche delle norme.

PRINCIPIO DI DOMINANZA. Lo scopo della teoria classica dei giochi è quello di identificare, sulla base di appropriati principi di scelta, la *strategia ottimale* di ciascun giocatore in ogni possibile gioco. La combinazione delle strategie ottimali di tutti i giocatori viene chiamata *soluzione* del gioco. Possiamo quindi dire che l'ambizioso obiettivo della teoria classica è quello di identificare la soluzione di qualunque gioco.

Un esempio molto chiaro di strategia ottimale è la cosiddetta *strategia dominante*. Supponiamo che, in un gioco a due giocatori, x e y siano due fra le strategie a disposizione di R . Diremo allora che x *domina* y se, *qualunque sia* la strategia adottata da C , il risultato ottenuto da R adottando x non è mai peggiore, e, in alcuni casi, è persino migliore, di quello che otterrebbe adottando y . Ciò significa che, adottando x , R ottiene payoff almeno pari e, in qualche caso, superiori ai payoff che avrebbe ottenuto adottando y . Un esempio di strategia dominante è la strategia indicata in grassetto nel gioco 5, cioè la strategia r_2 di R .

		C	
		c_1	c_2
R	r_1	1	0
	r_2	0	3
		0	2

Gioco 5. r_2 è la strategia dominante di R

In questo gioco, la strategia r_2 di R domina r_1 , perché, rispetto a questa, garantisce a R lo stesso payoff se C sceglie c_1 e un payoff maggiore se C sceglie c_2 . Quando, come accade in questo caso con r_2 , una strategia domina *ogni altra* strategia a disposizione di un giocatore, diremo che quella è la

strategia dominante del giocatore. Si noti che nel gioco 5, C non ha, al contrario di R , alcuna strategia dominante: infatti, se R gioca r_1 , C otterrà un payoff maggiore scegliendo c_1 , mentre se R gioca r_2 a C converrà scegliere c_2 .

Scegliere strategie dominate significa garantirsi payoff sicuramente non superiori, e talvolta inferiori, a quelli che si potrebbero ottenere scegliendo un'altra strategia. Per questa ragione, un principio di scelta largamente accettato, noto come *principio di dominanza* (D), afferma che:

- (D) (i) Un giocatore non dovrebbe mai scegliere una strategia dominata da qualche altra sua strategia.
 (ii) Quindi, se un giocatore ha una strategia dominante, questa è la sua strategia ottimale.

In base a (D) (ii), se un giocatore dispone di una strategia dominante allora dovrebbe adottarla, indipendentemente dalle sue opinioni su quello che farà l'altro giocatore.

PRINCIPIO DI NASH. Nel caso, molto frequente, in cui nessun giocatore disponga di una strategia dominante, non è immediatamente chiaro come determinare la soluzione del gioco. Per analizzare questo problema si ricorre al concetto di *equilibrio di Nash*, detto anche, semplicemente, *equilibrio*.⁶ In un gioco a due giocatori, la strategia x di R e la strategia y di C costituiscono un equilibrio di Nash, o, equivalentemente, sono in equilibrio, quando l'una rappresenta la risposta ottimale all'altra. Se x e y sono in equilibrio, nessuno dei giocatori, dopo essere venuto a conoscenza della strategia adottata dall'altro, avrà motivo di pentirsi della propria scelta. In altre parole, se anche i giocatori avessero la possibilità di cambiare *unilateralmente* la propria scelta dopo aver visto quella dell'altro giocatore, nessuno dei due avrebbe interesse a farlo.

Per questo motivo, appare piuttosto naturale accettare il seguente principio, che potremmo chiamare *principio di Nash* (N):

- (N) (i) La soluzione di un gioco deve essere un equilibrio di Nash; in altri termini, la strategia ottimale di ciascuno dei giocatori deve essere la risposta ottimale alla strategia dell'altro.
 (ii) Quindi, se un gioco ha un unico equilibrio, tale equilibrio è la soluzione del gioco, e le strategie che contribuiscono a formare tale equilibrio sono le strategie ottimali dei giocatori.

⁶ Questo concetto è stato introdotto negli anni Cinquanta dal matematico e Nobel per l'economia John Nash (nato nel 1928), uno dei pionieri della teoria dei giochi.

Uno dei principali risultati ottenuti da Nash è la dimostrazione del fatto che qualsiasi gioco, a patto di soddisfare alcune condizioni non troppo restrittive, possiede *almeno* un equilibrio. In generale, tuttavia, un gioco ha una molteplicità di equilibri. Si consideri, per esempio, il gioco 6 nel quale gli equilibri corrispondono alle celle con i payoff in grassetto.

		C	
		c_1	c_2
R	r_1	2 0	3 2
	r_2	1 1	0 1

Gioco 6. Gioco con due equilibri di Nash

Si noti che nel gioco 6 nessun giocatore dispone di una strategia dominante. Vi sono invece due equilibri di Nash, rappresentati dalle celle (r_1, c_2) e (r_2, c_1) . In queste due celle, infatti, la strategia di ogni giocatore è la risposta ottimale a quella dell'altro. In altre parole, se il gioco termina con uno di questi due risultati, entrambi i giocatori hanno ragione di essere soddisfatti della propria mossa, e non avrebbero interesse a cambiarla unilateralmente. L'esistenza di una molteplicità di equilibri di Nash caratterizza, in particolare, i giochi con parziale o totale coincidenza di interessi, che sono il genere di giochi di maggior rilevanza per le scienze sociali.

IL PRINCIPIO DI CONOSCENZA COMUNE. Di solito la 'bontà' della strategia adottata da un giocatore dipende dalle strategie adottate dagli altri; per esempio, in un gioco con due giocatori, *A* e *B*, la bontà della strategia di *A* dipenderà da quella adottata da *B*. Quindi, nella scelta razionale della propria strategia, *A* dovrà in qualche modo basarsi sulle sue attese circa la scelta di *B*, che a sua volta dipenderà dalle attese di *B* circa la scelta di *A*. È naturale chiedersi se la dipendenza delle scelte dei giocatori dalla loro attese circa le attese degli avversari non dia inizio a una sequenza infinita di *attese di attese*, dalla quale non si capisce come un giocatore possa giungere all'identificazione delle sua strategia ottimale. Fortunatamente, in molti tipi di gioco gli agenti possono anticipare con certezza la strategia adottata dagli avversari e, di conseguenza, scegliere la loro strategia senza lasciarsi impigliare nella spirale delle attese di attese. A tale scopo basta accettare il *principio di conoscenza*

comune (CC) secondo il quale le preferenze di ogni giocatore sono note a tutti gli altri e vi è inoltre conoscenza comune dell'altrui razionalità, nel senso che:

(CC) (i) Tutti sanno che tutti i giocatori sono perfettamente razionali.

(ii) Tutti sanno che tutti sanno che tutti sono perfettamente razionali, e così via all'infinito.

Per comprendere in che modo il principio di conoscenza comune (CC) possa condurre all'identificazione delle strategie ottimali considereremo due tipi di gioco, vale a dire quelli in cui almeno uno dei giocatori dispone di una strategia dominante e quelli caratterizzati dall'esistenza di un unico equilibrio di Nash.

GIOCHI IN CUI ALMENO UN GIOCATORE DISPONE DI UNA STRATEGIA DOMINANTE. Si è già visto che il principio di dominanza (D) implica che, se entrambi i giocatori, A e B , dispongono di una strategia dominante, la soluzione del gioco è data dalla combinazione delle loro strategie dominanti. Il principio di conoscenza comune (CC) ci permette di identificare la soluzione del gioco anche nel caso in cui solo uno dei giocatori, poniamo A , dispone di una strategia dominante. In questo caso, infatti, A adotterà, in accordo con (D), la sua strategia dominante; poiché, secondo (CC), B sa che A è razionale e che, di conseguenza, attuerà la propria strategia dominante, la strategia ottimale di B sarà quella che fornisce la risposta ottimale alla strategia dominante di A .

GIOCHI CON UN UNICO EQUILIBRIO. Come si ricorderà, il risultato (x, y) di un gioco con due giocatori A e B , dove x è la strategia di A e y quella di B , è un equilibrio nel caso in cui x e y sono l'una la risposta ottimale all'altra. Supponiamo che il gioco abbia un unico equilibrio (x, y) ; allora, in base al principio di Nash (N)(ii), (x, y) è la soluzione del gioco. Se si accetta (CC), si può affermare che ciascun giocatore sa che (x, y) è la soluzione del gioco e che anche l'altro giocatore ne è consapevole; di conseguenza, ciascun giocatore attuerà la strategia che contribuisce alla formazione (x, y) , del tutto fiducioso che anche l'altro si comporterà nello stesso modo.

GIOCHI CON UNA MOLTEPLICITÀ DI EQUILIBRI. Un'importante classe di giochi caratterizzata dalla presenza di un *unico* equilibrio è costituita dai giochi di puro conflitto. Nel 1951 Nash dimostrò che qualunque gioco, a patto di soddisfare alcune condizioni non troppo restrittive, possiede *almeno* un equilibrio. Sfortunatamente, però, la maggior parte dei giochi puramente o

parzialmente cooperativi è caratterizzata dalla presenza di una molteplicità di equilibri. Supponiamo di accettare l'idea che la soluzione di un gioco debba essere, in accordo con (N)(i), un equilibrio di Nash. Allora, di fronte a una molteplicità di equilibri e al fatto che, in generale, i giocatori preferiscono equilibri diversi, dobbiamo chiederci *quale* di essi sia la soluzione del gioco; in altre parole, dobbiamo chiederci in che modo i giocatori riusciranno a coordinare tacitamente le loro strategie così da giungere alla formazione di un particolare equilibrio.

UN PRESUPPOSTO FONDAMENTALE DELLA TEORIE CLASSICHE DEI GIOCHI. Alcuni teorici ritengono che i giocatori possano ottenere tale coordinazione sulla base di un appropriato *criterio di ottimalità* in grado di determinare, nell'ambito di una molteplicità di equilibri, l'equilibrio ottimale. Infatti, se tutti i giocatori condividessero il criterio di ottimalità, ciascuno di loro attuerebbe la strategia che contribuisce a formare l'equilibrio ottimale dato che, in base alla sua conoscenza dell'altrui razionalità, sarebbe sicuro che anche gli altri si comporteranno nello stesso modo. L'idea intuitiva appena descritta costituisce il nocciolo delle *teorie classiche dei giochi*, cioè delle teorie che conservano il presupposto fondamentale della teoria di von Neumann e Morgenstern, secondo il quale la razionalità individuale, assieme alla conoscenza comune delle preferenze e della razionalità altrui, è sufficiente a individuare la soluzione di ogni gioco.

1.3. *Teorie epistemiche dei giochi*

Alcuni buoni motivi per dubitare dell'ottimismo razionalistico delle teorie classiche sono connessi al fatto che sono stati finora proposti almeno tre dozzine di criteri di ottimalità. Di conseguenza, affinché i giocatori possano coordinarsi su un particolare equilibrio, identificandolo con l'equilibrio ottimale, il principio di conoscenza comune dell'altrui razionalità non basta: occorre anche che *tutti* i giocatori identifichino la razionalità con l'adozione di uno *specifico* criterio di ottimalità. In attesa di un consenso universale sul 'giusto' criterio di ottimalità, consenso che appare tutt'altro che imminente, occorre riconoscere che (CC) è troppo debole per consentire la soluzione di ogni gioco. Sotto altri aspetti, però, tale principio appare troppo forte, poiché richiede di supporre che, nelle nostre interazioni sociali, abbiamo sempre a che fare con agenti *perfettamente* razionali. In realtà, siamo quasi sempre alle prese con persone che condividono le nostre stesse limitazioni cognitive e che non di rado agiscono sulla base di abitudini o schemi fissi. I dubbi su (CC) hanno ispirato lo sviluppo di quelle che abbiamo sopra denominato *teorie epistemiche dei giochi*.

Tali teorie si propongono di analizzare le interazioni strategiche sulla base di un'attenta considerazione delle effettive condizioni epistemiche in cui si svolgono. Per esempio, Bicchieri (1994) dimostra che la razionalità dei giocatori e la loro conoscenza delle altrui preferenze e razionalità non è, in generale, sufficiente a *prescrivere* quello che i giocatori dovrebbero fare e, di conseguenza, non è neppure sufficiente a *prevedere* quello che faranno o a *spiegare* quello che hanno fatto. Per affrontare questi compiti vanno considerate anche le condizioni epistemiche di un gioco, cioè le capacità cognitive dei giocatori, le loro informazioni riguardo alle caratteristiche del gioco e i processi attraverso i quali possono formarsi determinate aspettative sul comportamento degli avversari. Bicchieri dimostra che, in molti casi, è sufficiente indebolire (CC) e supporre che i giocatori abbiano solo una *conoscenza limitata* dell'altrui razionalità, per poter prevedere su quale particolare equilibrio di Nash convergeranno le loro scelte.

L'ipotesi di conoscenza limitata consente, per esempio, una soddisfacente comprensione dei giochi che comportano il ricorso a minacce e promesse.⁷ Consideriamo, in particolare, il caso delle minacce. Una minaccia efficace dovrebbe agire da deterrente nei confronti di un potenziale trasgressore. L'efficacia della minaccia fatta da *A* nei riguardi di *B* dipende dalla sua credibilità, cioè dalla sua capacità di influenzare le aspettative di *B* su come *A* reagirebbe in caso di trasgressione. D'altra parte, colui che proferisce una minaccia ha sempre molte difficoltà a convincere l'avversario che la metterà in atto: infatti, la punizione di un trasgressore comporta *sempre* dei costi per chi la attua, cosicché non vi è alcun motivo razionale per infliggerla. Ne segue che, se *A* minaccia *B* e *B* ritiene che *A* sia perfettamente razionale, allora la minaccia di *A* non sarà credibile e non potrà quindi impedire la trasgressione. Se, invece, *B* ha qualche sia pur piccolo dubbio sulla razionalità di *A*, allora la minaccia di *A* potrebbe riuscire a influenzare le scelte di *B*. Ne segue che *A* avrà tutto l'interesse a convincere *B* di *non* essere perfettamente razionale e di avere invece la tendenza ad attuare le proprie minacce. Per raggiungere questo obiettivo, *A* può fare ricorso alla propria *reputazione* di persona inflessibile, reputazione che di solito dipende dal comportamento passato di *A*, cioè dal fatto che molto spesso, nelle passate interazioni, *A* ha effettivamente attuato le proprie minacce, accettandone i relativi costi. Come si comprende da questa breve descrizione, il meccanismo della reputazione può funzionare solo in un contesto di conoscenza limitata, dove le parti sono incerte riguardo razionalità degli avversari, e devono quindi apprendere, nel corso di ripetute interazioni, che tipi di giocatori hanno di fronte. Il campo di applicazione dell'ipotesi di conoscenza limitata, e dei connessi meccanismi di

⁷ La prima analisi sistematica dei giochi che comportano minacce, promesse e altri tipi di 'mosse strategiche' si deve a Schelling (1960/1980). Su questo argomento si veda anche Festa (2003; 2004).

apprendimento relativi ai tipi di giocatori con cui si interagisce, si estende ben al di là dei giochi che comportano minacce e promesse. Le ricerche effettuate da Bicchieri (1993, cap. 6 e 2006) e altri autori mostrano, infatti, che tale ipotesi svolge un ruolo fondamentale nei processi di formazione delle norme sociali.

1.4. *Teorie evoluzionistiche dei giochi*

Un'assunzione fondamentale della biologia evoluzionistica, a partire dalle prime formulazioni della teoria dell'evoluzione di Darwin, è quella per cui la dinamica dell'evoluzione biologica si basa sulla riproduzione differenziale degli individui più riusciti. La biologia evoluzionistica ipotizza l'esistenza di diversi meccanismi evolutivi che garantiscono la riproduzione differenziale degli individui più adatti. Le principali caratteristiche qualitative della riproduzione differenziale vengono catturate da un modello dinamico, noto come *dinamica dei riproduttori*. Le teorie evoluzionistiche dei giochi, nate dalla sintesi di teoria dei giochi e biologia evoluzionistica, si basano sull'idea di applicare la dinamica dei riproduttori a certe *popolazioni di giocatori* per comprendere l'evoluzione di determinate strategie di gioco. Più precisamente, tali teorie ipotizzano svariati meccanismi B noti come dinamiche evoluzionistiche B che consentono la riproduzione differenziale delle strategie che hanno dato buoni risultati in confronto alla media della popolazione, così da accrescere la proporzione di giocatori che giocano con 'buone' strategie.

Diversamente dalle teorie razionalistiche dei giochi, le teorie evoluzionistiche hanno carattere squisitamente descrittivo. Infatti, le strategie adottate dai giocatori non rappresentano il prodotto di scelte razionali, ma sono *schemi di comportamento* acquisiti mediante trasmissione genetica o culturale. Sulla scorta dei notevoli successi ottenuti dalle teorie evoluzionistiche dei giochi nella spiegazione del comportamento animale, alcuni studiosi ne hanno applicato i modelli formali all'analisi dell'evoluzione culturale. Tale estensione dell'originario ambito di applicazione delle teorie

evoluzionistiche è reso possibile dal carattere assai astratto dei loro modelli formali; in particolare, tali modelli *non* presuppongono che la trasmissione delle strategie di successo si attui necessariamente attraverso la riproduzione biologica. Le strategie di successo potrebbero diffondersi anche attraverso i meccanismi della trasmissione culturale; per esempio, i giocatori potrebbero cambiare le proprie strategie in base all'osservazione dei risultati ottenuti dagli altri giocatori e all'imitazione delle strategie che hanno ottenuto i maggiori successi.

Le più note applicazioni delle teorie evoluzionistiche dei giochi all'analisi delle norme sociali si devono a Robert Axelrod (1984 e 1997) e Brian Skyrms

(1996 e 2004).⁸ In particolare, Axelrod (1984) usa sofisticate simulazioni computerizzate per esplorare i meccanismi attraverso i quali determinate strategie basate sulla cooperazione possono affermarsi in una società di egoisti razionali. Metodi sostanzialmente simili a quelli di Axelrod vengono utilizzati da Skyrms (1996) nelle sue ricerche sui meccanismi che favoriscono l'affermarsi delle strategie di equità nelle interazioni sociali. Le simulazioni computerizzate di Axelrod e Skyrms permettono di scoprire il risultato aggregato di alcuni semplici schemi di comportamento individuale, rappresentati dalle strategie utilizzate da ciascuno dei membri di un determinata popolazione.

2. Applicazioni etico-politiche della teoria dei giochi

Poiché la teoria dei giochi mira a valutare, spiegare e prevedere il comportamento umano nelle interazione strategiche, era del tutto naturale attendersi che i filosofi morali e politici avrebbero ben presto cercato di applicarla al loro campo di ricerca. Infatti il primo tentativo di applicare la teoria dei giochi nell'analisi di questioni etico-politiche si deve al filosofo inglese Richard Braithwaite che, già alla metà degli anni cinquanta dello scorso secolo, in una conferenza intitolata *Theory of Games as a Tool for the Moral Philosopher* (1955), affrontò in termini giochistici il problema della giustizia distributiva.⁹ Le applicazioni etico-politiche della teoria dei giochi hanno dato origine a tre distinti approcci, vale a dire gli approcci contrattualista, funzionalista ed evolucionista, i cui caratteri essenziali verranno ora illustrati.¹⁰

APPROCCIO CONTRATTUALISTA. A partire dalle ricerche di Richard Braithwaite (1955), svariate applicazioni della teoria dei giochi alla riflessione etico-politica sono state ispirate dall'intento di riformulare e fondare su basi giochistiche la tradizionale teoria del contratto sociale.¹¹ Per esempio, i cosiddetti contrattualisti della scelta razionale hanno usato gli strumenti concettuali della teoria della contrattazione, un'importante area della teoria

⁸ Sulle ricerche di Axelrod e Skyrms si veda anche Festa (1999; 2001).

⁹ Per quanto riguarda la storia della filosofia morale e politica, gli studi di David Gauthier (1969), Michael Taylor (1987) e altri hanno mostrato che si possono rintracciare intuizioni giochistiche già negli scritti di Thomas Hobbes, David Hume e Jean-Jacques Rousseau.

¹⁰ Prendiamo a prestito questa classificazione da Verbeek e Morris (2004). Questo scritto offre, fra l'altro, un'eccellente rassegna delle applicazioni etico-politiche della teoria dei giochi.

¹¹ Dopo Braithwaite, la teoria del contratto sociale è stata riformulata in termini giochistici da svariati autori tra i quali John Harsanyi (1955), Brian Barry (1965), John Rawls (1971) e David Gauthier (1967, 1987).

dei giochi sviluppata da John Nash, per giustificare le norme e le istituzioni politiche, viste come il risultato di una contrattazione tra agenti razionali. I sostenitori dell'approccio contrattualista tendono a identificare le norme morali con determinati 'principi di giustizia' posti alla base di un sistema di norme e istituzioni politiche; ciò significa che un sistema politico 'giusto' andrebbe difeso con riferimento a una 'comunità morale' formata da agenti razionali che condividono alcuni fondamentali principi morali.

APPROCCIO FUNZIONALISTA. Le indagini effettuate nell'ambito dell'approccio funzionalista mirano a identificare la funzione della moralità o, equivalentemente, a identificare i problemi che insorgono in assenza della moralità. Un fondamentale contributo a questo approccio si deve a Edna Ullmann-Margalit (*The Emergence of Norms*, 1977) che utilizza l'apparato formale della teoria dei giochi per esplorare la funzione delle norme morali. L'autrice riesce, fra l'altro, a mostrare che l'adesione alle norme morali consente agli attori impegnati in diversi tipi di interazioni strategiche di ottenere risultati molto migliori di quelli che sarebbero consentiti dal loro puro interesse egoistico, perseguito in conformità con i principi di scelta della teoria classica dei giochi.

APPROCCIO EVOLUZIONISTA. L'approccio evoluzionista concepisce la moralità come l'effetto non intenzionale delle interazioni degli agenti. Tenta quindi di offrire B in accordo con i principi fondamentali delle teorie evoluzionistiche dei giochi B spiegazioni evoluzionistiche delle pratiche morali consistenti nel mostrare le ragioni che stanno alla base del loro successo evolutivo. Nelle loro analisi dei processi di formazione e sviluppo delle norme gli evoluzionisti fanno ampio ricorso a *simulazioni computerizzate*, o *società artificiali*, in grado di riprodurre le interazioni tra agenti.¹² Va inoltre osservato che, soprattutto a partire dagli anni novanta, i sostenitori dell'approccio evoluzionista hanno fatto ampio ricorso non solo alle teorie evoluzionistiche dei giochi ma anche a quelle epistemiche (Bicchieri, 1993 e 2006).

Vale la pena mettere in rilievo alcune importanti somiglianze e differenze fra i tre approcci sopra menzionati:

(1) Mentre l'approccio contrattualista ha fini essenzialmente normativi, gli approcci evoluzionista e funzionalista hanno intenti apertamente descrittivi. Per questa ragione, i contrattualisti si avvalgono quasi esclusivamente della

¹² Per un'eccellente rassegna dei risultati ottenuti nell'ambito dell'approccio evoluzionista all'analisi delle norme morali, si veda Harms e Skyrms (2007).

teoria classica dei giochi, mentre i funzionalisti e gli evoluzionisti fanno ampio uso anche delle teorie evoluzionistiche ed epistemiche.

(2) I funzionalisti e gli evoluzionisti tendono a trascurare il problema di valutare lo specifico contenuto delle norme morali e di giustificarne le pretese normative. Sono invece accomunati dall'interesse per la descrizione di quelle norme o pratiche morali che vengono comunemente considerate buone. Si può anzi affermare che la maggior parte delle indagini effettuate nell'ambito degli approcci funzionalista ed evoluzionista si basano, più o meno esplicitamente, sull'identificazione delle norme morali con le cosiddette *norme prosociali*, cioè con le norme che risultano utili al funzionamento di una società e al benessere dei suoi membri. Tale identificazione, che appare piuttosto problematica, e che ci guardiamo bene dal sottoscrivere, spinge gli evoluzionisti a usare quasi sempre in modo interscambiabile i termini 'norme', 'norme morali', 'norme sociali' e 'norme prosociali'.

(3) Va infine osservato che le differenze tra gli approcci funzionalista ed evoluzionista non sono meno importanti delle loro somiglianze. Per esempio, i sostenitori dell'approccio evoluzionista ritengono che identificare la funzione delle norme non basti a spiegarne l'emergere, la persistenza e l'evoluzione. Pensano, inoltre, che l'approccio funzionalista si precluda la possibilità di spiegare *anche* l'emergere e la persistenza di norme disfunzionali, o antisociali.

3. *L'evoluzione delle norme morali nella teoria evoluzionistica dei giochi*

In questo paragrafo considereremo alcuni risultati ottenuti dagli studiosi che hanno indagato l'evoluzione delle norme di cooperazione e di equità dalla prospettiva della teoria evoluzionistica dei giochi. In particolare, ci occuperemo dell'evoluzione delle norme di cooperazione altruistica nel dilemma del prigioniero e nel gioco della caccia al cervo (3.1 e 3.2) e dell'evoluzione delle norme di equità nel gioco dell'ultimatum (3.3).

3.1. *Dilemmi sociali. Il dilemma del prigioniero e l'evoluzione delle norme di cooperazione altruistica*

IL DILEMMA DEL PRIGIONIERO. A dispetto della grande plausibilità intuitiva del principio di dominanza, ci si è presto resi conto che l'incondizionata applicazione di questo principio conduce ad alcune conseguenze tanto interessanti quanto problematiche. La più famosa tra queste è il cosiddetto

dilemma del prigioniero, il quale deve la sua notorietà soprattutto al fatto che sembra mostrare l'insuccesso della razionalità individuale nel raggiungere risultati soddisfacenti per *tutti* i partecipanti a un certo genere, presumibilmente molto diffuso, di interazione strategica.

Il dilemma del prigioniero è un gioco che deve il suo nome alla storiella con cui viene solitamente illustrato. Due malviventi vengono arrestati a bordo di un'auto rubata. La polizia può condannarli entrambi a 1 anno per il furto, ma sospetta che i due abbiano commesso una grossa rapina avvenuta recentemente e preferirebbe incriminarli per questo reato, che costerebbe loro 10 anni di prigione. Per ottenere una confessione, il commissario propone a ognuno dei prigionieri la seguente alternativa: o confessare la rapina e tradire il proprio complice, o rimanergli fedele e tacere. Se un solo prigioniero confessa, permettendo al commissario di risolvere il caso, verrà lasciato libero e il complice sconterà l'intera pena, cioè 10 anni. Se entrambi i prigionieri confessano, otterranno entrambi uno sconto di pena di 5 anni per aver collaborato con la polizia, scontando solo i restanti 5. Se entrambi tacciono, la polizia non avrà prove sufficienti a incriminarli per la rapina, cosicché entrambi sconteranno solo 1 anno per il furto dell'auto.

La situazione dei due prigionieri, che chiameremo P1 e P2, può venir rappresentata mediante la matrice del gioco 7, in cui ciascun giocatore ha due strategie: tacere e non tradire il complice o, viceversa, confessare e tradirlo. Dal punto di vista dei prigionieri, la confessione equivale a *defezionare* nei confronti del compagno, mentre tacere equivale a *cooperare* con lui. Chiameremo quindi *c* (cooperazione) la strategia del silenzio e *d* (defezione) la strategia della confessione. I payoff rappresentano gli anni di galera e sono quindi negativi o, nel migliore dei casi, nulli.

		P2	
		<i>c</i>	d
P1	<i>c</i>	-1 -1	0 -10
	d	-10 0	-5 -5

Gioco 7. Dilemma del prigioniero

La matrice mostra che il dilemma del prigioniero è un gioco parzialmente conflittuale o, se si preferisce, un gioco parzialmente cooperativo. Infatti le preferenze dei due prigionieri coincidono rispetto ad alcune coppie di risultati, ma non rispetto ad altre. Per esempio, entrambi preferiscono (c, c) a (d, d) , ma P1 preferisce (d, c) a ogni altro risultato, mentre il risultato preferito da P2 è (c, d) che invece, dal punto di vista di P1, è il risultato meno gradito.

La matrice mostra anche che entrambi i giocatori hanno una strategia dominante, cioè la defezione. Vediamo, per esempio, che a P1 conviene scegliere d indipendentemente da quel che farà P2. Infatti, se P2 coopera, a P1 conviene defezionare, così da ottenere subito la libertà (payoff 0) invece di scontare 1 anno di galera (payoff -1); d'altra parte, se P2 sceglie d , a P1 conviene ancora scegliere d , in modo da scontare 5 anni di galera (payoff -5) invece che 10 (payoff -10). I payoff di P2 sono simmetrici rispetto a quelli di P1: quindi d è la strategia dominante anche per P2. Poiché, in base al principio di dominanza (D), un giocatore deve adottare la propria strategia dominante, la soluzione del gioco sarà (d, d) , cioè il risultato in cui entrambi i prigionieri defezionano, scontando così 5 anni di carcere.¹³

Si noti che (d, d) è un equilibrio del gioco. Infatti, una volta raggiunto questo risultato, nessun prigioniero accetterebbe di cambiare unilateralmente la propria strategia, passando a c , poiché se lo facesse sconterebbe 10 anni di galera invece di 5. Come si può facilmente controllare dalla matrice, (d, d) è anche l'unico equilibrio del gioco. In base al principio di Nash (N)(ii), il quale afferma che, se un gioco ha un unico equilibrio, tale equilibrio è la soluzione del gioco, arriviamo nuovamente, per altra via, alla conclusione che (d, d) è la soluzione del gioco. Ciò significa che la soluzione (d, d) , può venire determinata non solo sulla base di (D), ma anche di (N). Poiché (D) ed (N) sono due principi di scelta fondamentali della teoria classica dei giochi, e, in particolare, (D) appare estremamente plausibile, non vi sono motivi per dubitare che (d, d) sia la giusta soluzione del gioco.

IL DILEMMA DEL PRIGIONIERO COME DILEMMA SOCIALE. La circostanza che la soluzione del gioco è (d, d) apre la strada al dilemma associato al gioco del prigioniero. Il dilemma nasce, infatti, dal rilievo che il risultato (c, c) , i cui payoff sono indicati in corsivo nella matrice, è nettamente preferibile a (d, d)

¹³ Nella sua forma generale, il dilemma del prigioniero può venire definito come un gioco in cui ciascun giocatore può cooperare (c) o defezionare (d) e i payoff delle quattro possibili coppie di atti sono i seguenti: R è il premio attribuito a entrambi se entrambi cooperano; P è la penalità attribuita a entrambi se entrambi defezionano; mentre B (la 'paga del babbeo') e T (il 'guadagno della tentazione') vengono attribuiti, rispettivamente, al primo e al secondo giocatore, quando il primo coopera e il secondo defeziona. Affinché il gioco sia un dilemma del prigioniero la condizione essenziale è che valga l'ordinamento $T > R > P > B$.

per *entrambi* i giocatori. Se, infatti, i prigionieri cooperassero, cioè se nessuno dei due confessasse, sconterebbero solo 1 anno di galera invece di 5, come accade in (d, d) . In altre parole, benché la defezione domini la cooperazione, e sia quindi la soluzione del gioco, quest'ultima garantisce risultati migliori a entrambi i giocatori. Ciascun prigioniero si trova quindi alle prese con un dilemma: da un lato sa che l'unica strategia razionale è confessare, ma dall'altro comprende che tacere converrebbe a entrambi.

L'immensa fortuna del dilemma del prigioniero tra i filosofi e gli scienziati sociali è dovuta propria alla sua capacità di rappresentare in forma semplice e suggestiva un'ampia gamma di situazioni in cui diversi individui, perseguendo ciascuno il proprio interesse, scelgono di defezionare invece che di cooperare, ottenendo così risultati negativi, o persino rovinosi, per tutti. Con riferimento a situazioni di questo genere si parla di *dilemmi sociali*. Come osserva Bicchieri (2006, p. 140),

[u]n dilemma sociale è, per definizione, una situazione in cui ogni membro del gruppo ottiene un risultato migliore se persegue il proprio interesse personale, ma, allo stesso tempo, ciascuno trae beneficio dal fatto che tutti i membri del gruppo assecondino l'interesse comune.

Occorre ricordare che il dilemma del prigioniero è il più famoso dei dilemmi sociali, ma certamente non l'unico. Possiamo caratterizzare in termini assai generali i dilemmi sociali mediante il concetto di *Pareto-ottimalità*.¹⁴

Consideriamo i possibili risultati di un gioco con due o più giocatori. Diremo che un determinato risultato è *Pareto-dominato* da un altro risultato quando il secondo assegna a tutti i giocatori payoff non inferiori a quelli assegnati dal primo e, per di più, assegna a uno o più giocatori payoff superiori a quelli assegnati dal primo. Diremo poi che un risultato è *Pareto-ottimale* quando non è Pareto-dominato da nessun altro risultato. Per esempio, la matrice del gioco 7 ci mostra che nel dilemma del prigioniero ci sono tre risultati Pareto-ottimali, corrispondenti alle celle (c, c) , (c, d) , e (d, c) . L'unico risultato non Pareto-ottimale è proprio la soluzione del gioco, (d, d) , che è Pareto-dominata da (c, c) : ciò significa che vi è un risultato del gioco che entrambi i giocatori preferiscono alla soluzione del gioco. Quest'ultima circostanza non è soltanto un aspetto saliente del dilemma del prigioniero, ma costituisce il tratto distintivo dei dilemmi sociali: tali dilemmi possono infatti venire definiti, in termini generali, come *giochi la cui soluzione non è Pareto-ottimale*.

¹⁴ Tale concetto venne introdotto dall'economista italiano Vilfredo Pareto (1848-1923) nell'ambito della cosiddetta economia del benessere, e fu successivamente applicato nella teoria dei giochi.

NORME DI COOPERAZIONE ALTRUISTICA NEI DILEMMI SOCIALI. La nozione di dilemma sociale dischiude la possibilità di definire in termini generali anche le nozioni di *cooperazione* e *norma di cooperazione*. Consideriamo, per semplicità, un dilemma sociale costituito da un gioco a due, che chiameremo G . La soluzione di G sarà rappresentata da una determinata cella (r_0, c_0) , dove r_0 e c_0 sono, rispettivamente, le mosse ottimali dei giocatori R e C . Poiché G è un dilemma sociale, (r_0, c_0) non può essere Pareto-ottimale; ciò significa che (r_0, c_0) è Pareto-dominata da qualche altro risultato del gioco. Supponiamo, per semplicità, che (r_0, c_0) sia Pareto-dominata solo da un altro risultato di G , che chiameremo (r^*, c^*) , ove $r^* \neq r_0$ e $c^* \neq c_0$. Diremo allora che R coopera quando sceglie r^* e, analogamente, che C coopera quando sceglie c^* . Possiamo dire, inoltre, che r^* e c^* costituiscono le *mosse di cooperazione*, rispettivamente, di R e C . Diremo, infine, che un giocatore agisce sulla base di una *norma di cooperazione* quando, in conflitto con i principi di scelta della teoria dei giochi, adotta la sua mossa di cooperazione.

Possiamo facilmente applicare la definizione appena introdotta di norma di cooperazione al dilemma del prigioniero; come era lecito attendersi dalla definizione generale di norma di cooperazione segue che un prigioniero agisce sulla base di una norma di cooperazione quando sceglie c . È inoltre interessante rilevare che la cooperazione attuata scegliendo la mossa c costituisce una forma particolarmente forte di cooperazione, che potremmo definire *cooperazione altruistica*. Con questo termine ci riferiamo alla circostanza che, come si può facilmente vedere dalla matrice del gioco 7, la scelta di c da parte di un prigioniero accresce in ogni caso i payoff dell'altro prigioniero. Se, per esempio, P2 coopera allora il suo payoff sarà -10 se P1 sceglie d e -1 se sceglie c ; se, invece, P2 defeziona allora il suo payoff sarà -5 se P1 sceglie d e 0 se sceglie c . Ciò significa che P2, quale che sia la sua mossa, trae un notevole vantaggio dalla mossa c di P1; con riferimento alla mossa c di P1 possiamo quindi a buona ragione parlare di *mossa di cooperazione altruistica*.

La nozione di mossa di cooperazione altruistica può venire opportunamente applicata a qualunque dilemma sociale. Dato un dilemma sociale G diremo, che un giocatore agisce sulla base di una *norma di cooperazione altruistica* quando B in conflitto con i principi di scelta della teoria dei giochi B adotta la sua mossa di cooperazione altruistica.¹⁵

¹⁵ Molte norme prosociali sono, presumibilmente, norme di cooperazione altruistica. Inoltre, non ci appare irragionevole supporre che alcune norme morali abbiano la struttura formale delle norme di cooperazione altruistica. Nella loro introduzione a *Preferences* (1998), Fehige e Wessels, parlano di *preferenzialismo* con riferimento a una concezione della morale secondo la quale una persona è morale se agisce in modo tale da soddisfare *al meglio* le *preferenze di ognuno*. Se si accetta la concezione preferenzialista della morale si potrebbe forse sostenere che le norme di cooperazione altruistica possono essere considerate norme morali. Sulla concezione preferenzialista della morale si veda anche Festa (2000).

L'EVOLUZIONE DELLE NORME DI COOPERAZIONE ALTRUISTICA NEL DILEMMA DEL PRIGIONIERO. Nei dilemmi del prigioniero che si presentano nella vita sociale gli individui sembrano manifestare, a dispetto dei principi della teoria classica dei giochi, che suggeriscono di scegliere in ogni caso la defezione, una spiccata tendenza a cooperare, cioè ad agire in conformità a una norma di cooperazione altruistica. Come possiamo spiegare questa benefica, per quanto irrazionale, tendenza alla cooperazione?

Nel famoso volume *The Evolution of Cooperation*, Robert Axelrod (1984) risponde a questo interrogativo osservando, anzitutto, che nella maggior parte delle situazioni dove la tendenza a cooperare sembra più accentuata, i giocatori in realtà non affrontano un semplice dilemma del prigioniero, bensì un *gioco iterativo del prigioniero*. In tale gioco essi possono incontrarsi per un numero indefinito di volte, senza sapere quale sarà l'ultima occasione di interazione. Il problema affrontato da Axelrod è quello di identificare le condizioni necessarie e sufficienti per l'affermarsi di norme di cooperazione altruistica in un gioco iterativo del prigioniero.

Definire una strategia per tale gioco presuppone che ogni giocatore abbia la capacità di riconoscere un giocatore già affrontato e abbastanza memoria da ricordare la storia di tutte le precedenti interazioni con lui. Sulla base di queste assunzioni, una strategia per il gioco iterativo del prigioniero consiste nel decidere se attuare *c* o *d* al primo incontro e nello scegliere, sulla base della storia dei precedenti incontri con il nostro attuale compagno di gioco, tra *c* e *d* per ogni mossa successiva. Un aspetto peculiare dell'interazione strategica nel gioco iterativo è che la scelta che uno dei giocatori opera in qualunque mossa del gioco potrebbe influenzare le scelte future dell'altro giocatore e, quindi, quelle di entrambi: in tal modo il futuro può sempre gettare la sua ombra sul presente, incidendo sulla situazione strategica attuale (Axelrod, 1984, p. 18).

Supponiamo che un individuo debba affrontare il gioco iterativo con altri membri tratti a caso dalla popolazione. Il problema di stabilire la strategia migliore per questo individuo, cioè quella che, se adottata, gli frutterebbe il 'massimo punteggio', è talmente complesso da lasciare ben poche speranze di risolverlo per via puramente deduttiva. Per questo motivo, Axelrod affrontò il problema con un metodo assai ingegnoso di simulazione al computer. Dopo avere invitato alcuni specialisti della teoria dei giochi a presentare una propria strategia per il gioco iterativo, codificata in un appropriato programma, organizzò un torneo computerizzato, facendo interagire ripetutamente tra loro le quattordici strategie concorrenti più una strategia randomizzata. Con grande sorpresa di Axelrod la più semplice fra tutte le strategie proposte, vale a dire *Tit For Tat* ('colpo su colpo') di Anatol Rapoport, vinse il torneo sbaragliando gli avversari. Alla prima mossa *Tit For Tat* (d'ora in poi, TFT) coopera, e in

ogni mossa successiva sceglie fra *c* e *d* semplicemente riproducendo la scelta effettuata dall'avversario nella mossa immediatamente precedente. Questo significa che TFT è caratterizzata dal massimo grado di *reciprocità* e, quindi, dalla massima disponibilità a cooperare con altri giocatori che si dimostrino inclini alla cooperazione. Un risultato ancora più sorprendente fu ottenuto da Axelrod con un secondo torneo computerizzato al quale erano stati ammesse sessantadue strategie proposte da studiosi a conoscenza dei risultati del primo girone. Nonostante alcune di queste strategie fossero ispirate dallo sforzo consapevole di perfezionare TFT, la versione originale di TFT vinse anche il secondo girone.

Il successo di TFT nel secondo girone, contro una così grande varietà di avversari, suggerisce l'ipotesi che TFT sia una strategia talmente 'robusta' da comportarsi bene nelle più svariate situazioni, e persino nell'ambito di popolazioni nelle quali possa essere rappresentata *qualsiasi* strategia.

Per determinare il grado di robustezza di TFT, e di strategie consimili fondate su criteri di reciprocità e disponibilità a cooperare, Axelrod ha effettuato svariati esperimenti di simulazione basati sui modelli dinamici della teoria evolutiva dei giochi. Questi esperimenti consistono in una lunga serie di gare nel seguito delle quali le strategie di maggior successo sono ammesse a partecipare a un numero maggiore di incontri, mentre viene limitato il grado di partecipazione delle strategie meno fortunate. Ogni gara rappresenta una generazione di individui ciascuno dei quali impiega costantemente una determinata strategia nelle sue interazioni con altri membri della popolazione. Il successo di una strategia in ogni singola gara determinerà la numerosità della sua 'prole', e quindi la misura in cui tale strategia sarà rappresentata nella generazione successiva. A partire da una determinata distribuzione iniziale, e dal valore di alcuni altri parametri, potremo stabilire, attraverso ben congegnati esperimenti di simulazione, come evolveranno le diverse strategie e quale sarà la loro distribuzione finale nella popolazione. In effetti, gli esperimenti condotti da Axelrod hanno permesso di comprendere le ragioni del successo di TFT e delle consimili strategie cooperative nelle più svariate condizioni iniziali, e anche i meccanismi attraverso i quali tali strategie possono penetrare in un 'mondo di cattivi', dove domina la strategia, 'Defeziona sempre!', diffondersi in questo mondo fino a invaderlo, e resistere poi al tentativo di invasione da parte di altre strategie.

Axelrod mostra che il meccanismo fondamentale attraverso il quale le strategie cooperative possono prosperare nel gioco iterativo del prigioniero è quello della correlazione, attraverso il quale un individuo che usa una certa strategia ha una probabilità sufficientemente alta di incontrarne un altro che usa la sua stessa strategia.

LA CACCIA AL CERVO: UN DILEMMA DELLA SFIDUCIA. In quest'ultimo decennio gli studiosi interessati all'analisi giochistica delle norme hanno manifestato un crescente interesse per un gioco noto come *caccia al cervo*. Un'eccellente esempio delle ricerche su questo argomento è dato dal recente *The Stag Hunt and the Evolution of Social Structure* (2004) di Brian Skyrms. Il gioco della caccia al cervo prende nome da una breve storia raccontata da Jean-Jacques Rousseau (1755) nel *Discorso sull'origine e i fondamenti della disuguaglianza*:

Se si dovesse cacciare un cervo, ognuno comprenderebbe che deve restare fermo e fiducioso al suo posto; ma se passasse una lepre nei paraggi di ciascuno dei due cacciatori, non possiamo dubitare che egli lascerebbe il posto per cacciarla, senza alcuno scrupolo.¹⁶

Nella caccia al cervo i giocatori sono due cacciatori, posti di fronte alla scelta se cacciare la lepre oppure il cervo. Si tratta di opzioni sono molto diverse. Mentre un cacciatore può cacciare la lepre da solo, la caccia al cervo può essere attuata solo in collaborazione con il suo socio; d'altra parte, la ricompensa che ciascuno si può attendere dalla decisione solitaria di cacciare la lepre è piuttosto piccola, mentre dalla collaborazione nella caccia al cervo entrambi i cacciatori possono attendersi una ricompensa maggiore.

La situazione dei due cacciatori B che chiameremo C1 e C2 B può venir rappresentata mediante la matrice del gioco 8, in cui ciascun giocatore ha due mosse: cacciare il cervo o cacciare la lepre. Ci riferiremo a queste due mosse semplicemente con *cervo* e *lepre*.

		C2	
		<i>cervo</i>	<i>lepre</i>
C1	<i>cervo</i>	4	3
	<i>lepre</i>	0	3
		4	0
		3	3

Gioco 8. Caccia al cervo

¹⁶ Si veda p. 176 dell'edizione italiana del *Discorso*, citata in bibliografia. La traduzione è stata qui leggermente modificata.

Naturalmente non è del tutto certo che un cacciatore impegnato in una solitaria caccia alla lepre riesca effettivamente a catturarla, e neppure che due cacciatori, collaborando nella caccia al cervo, abbiano successo. Possiamo però supporre che entrambi i cacciatori possano attendersi dalla caccia congiunta al cervo una ricompensa maggiore di quella che si attendono da una solitaria caccia alla lepre: in termini più esatti, diremo che il valore atteso della caccia congiunta al cervo è per entrambi superiore al valore atteso della caccia solitaria alla lepre. Per questo motivo, nella matrice del gioco 8 i giocatori attribuiscono un payoff 4 al risultato (*cervo, cervo*) e un payoff inferiore, pari a 3, ai due risultati del gioco che possono derivare dalla scelta di cacciare la lepre B così, per esempio, C1 attribuirà un payoff 3 ai risultati (*lepre, cervo*) e (*lepre, lepre*). Infine, dato che la caccia solitaria al cervo è destinata al fallimento, ciascun giocatore attribuisce un payoff 0 al risultato in cui si impegna nella caccia al cervo, mentre il suo socio si dedica alla caccia alla lepre, così, per esempio, C1 attribuirà un payoff 0 al risultato (*cervo, lepre*).

La matrice mostra che la caccia al cervo è, alla pari del dilemma del prigioniero, un gioco parzialmente cooperativo. Infatti le preferenze dei due prigionieri coincidono solo rispetto ad alcune coppie di risultati, ma non rispetto ad altre. Per esempio, entrambi preferiscono (*cervo, cervo*) a (*lepre, lepre*) e a ogni altro risultato, ma C1 preferisce (*lepre, cervo*) a (*cervo, lepre*), mentre C2 ha naturalmente la preferenza opposta. Dalla matrice emergono chiaramente anche altre importanti caratteristiche della caccia al cervo:

- (a) nessun giocatore dispone di una strategia dominante;
- (b) vi sono due equilibri di Nash, vale a dire (*cervo, cervo*) e (*lepre, lepre*);
- (c) uno dei due equilibri, cioè (*cervo, cervo*), Pareto-domina l'altro;
- (d) inoltre (*cervo, cervo*) Pareto-domina anche ogni altro risultato del gioco.

Mettendo chiaramente in luce le caratteristiche (a)-(d) della caccia al cervo, la rappresentazione astratta della caccia al cervo fornita nella matrice del gioco 8 ci permette di riconoscere che molte situazioni della vita quotidiana e svariati problemi inventati dai filosofi possono essere considerati come casi di caccia al cervo. Per esempio, non è difficile vedere che il problema del prosciugamento di un campo, affrontato da David Hume (1739) nel *Trattato sulla natura umana*, ha la struttura formale della caccia al cervo:

Due vicini possono mettersi d'accordo per prosciugare un campo di loro comune proprietà perché è facile per entrambi conoscere i loro

reciproci propositi; e ciascuno può rendersi conto che l'immediata conseguenza di non fare la sua parte è l'abbandono dell'intero progetto.¹⁷

Ci si può ora chiedere quale sia la soluzione della caccia al cervo. Occorre anzitutto notare che, data l'assenza di strategie dominanti (vedi (a)) e la pluralità degli equilibri del gioco (vedi (b)), i due principi fondamentali di scelta della teoria dei giochi, vale a dire il principio di dominanza (D) e il principio di Nash (N), non possono essere applicati. Tuttavia, la clausola (i) di (N) B per cui la soluzione di un gioco deve essere un equilibrio di Nash B restringe il campo delle possibili soluzioni ai due equilibri (*cervo, cervo*) e (*lepre, lepre*). In vista di (c), il risultato (*cervo, cervo*) sembra essere l'equilibrio ottimale poiché Pareto-domina non solo l'equilibrio (*lepre, lepre*), ma ogni altro risultato del gioco (si veda (d)). Ciò significa che (*cervo, cervo*) è il risultato che entrambi i giocatori vorrebbero vedere realizzato. Non dovrebbe esservi quindi alcun dubbio nell'affermare che, dal punto di vista della teoria classica dei giochi, (*cervo, cervo*) è la soluzione del gioco o, equivalentemente, che *cervo* è la mossa ottimale di entrambi i cacciatori.¹⁸

Vale la pena osservare che B diversamente dal dilemma del prigioniero B la caccia al cervo *non* è un dilemma sociale: infatti la soluzione del gioco non è Pareto-dominata da nessun altro risultato del gioco, ma al contrario Pareto-domina ogni altro risultato.

Poiché *cervo* è la mossa che, se attuata da entrambi, porta al risultato preferito da entrambi, può venire intesa come la (mossa di) cooperazione; al contrario *lepre* rappresenta la non-cooperazione, cioè la defezione. Di conseguenza, diremo che (*cervo, cervo*) è l'*equilibrio cooperativo* del gioco, mentre (*lepre, lepre*) è l'*equilibrio non cooperativo*.

Come si ricorderà la teoria classica dei giochi si basa sull'accettazione, più o meno esplicita, del principio di conoscenza comune (CC) secondo il quale tutti i giocatori sono perfettamente razionali, sanno che tutti sono perfettamente razionali, sanno che tutti sanno che tutti sono perfettamente razionali, e così via all'infinito. Nel caso particolare della caccia al cervo, da (CC) consegue che ciascun cacciatore sceglierà la propria mossa ottimale, cioè *cervo*, poiché sarà completamente fiducioso che anche il suo socio opererà la stessa scelta. Ci si può chiedere, tuttavia, se nelle reali condizioni epistemiche

¹⁷ Si veda p. 570 dell'edizione italiana del *Trattato*, citata in bibliografia. La traduzione è stata qui leggermente modificata. Hume (*ibidem*) osserva anche che, mentre l'accordo tra due persone per il prosciugamento del campo ha molte probabilità di funzionare, è impossibile che un migliaio di persone si accordino per un'impresa del genere. Questo rilievo è in pieno accordo con le ricerche svolte nell'ambito della teoria dei giochi le quali mostrano che, in generale, le possibilità della cooperazione decrescono mano a mano che aumenta il numero dei giocatori.

¹⁸ Nella terminologia della teoria dei giochi si dice che (*cervo, cervo*) è l'*equilibrio payoff dominant* poiché, fra tutti gli equilibri del gioco, dà a entrambi i giocatori i payoff più alti.

in cui si svolgono le interazioni tra esseri umani, un cacciatore può davvero riporre un'incondizionata fiducia nella perfetta razionalità del suo socio. Sembra infatti del tutto plausibile supporre che un cacciatore possa porsi i seguenti interrogativi. Riuscirà il mio socio a comprendere che *cervo* è la sua mossa ottimale? E avrà i nervi abbastanza saldi per resistere alla tentazione di dare la caccia alla lepre se dovesse vederne una nei dintorni? E avrà abbastanza fiducia nella mia razionalità e nei miei nervi saldi da convincersi che io sceglierò la mia mossa ottimale, cioè *cervo*? Ma anche in questo caso, non temerà che io non abbia abbastanza fiducia nella sua razionalità e nei suoi nervi saldi, al punto che, non fidandomi della sua collaborazione nella caccia al cervo, sarò spinto dalla mia sfiducia a cacciare la lepre? E così via all'infinito.

Gli interrogativi appena formulati suggeriscono che, nelle condizioni epistemiche della vita reale, la nostra fiducia nella razionalità altrui, e anche la nostra fiducia nella fiducia che gli altri hanno nella nostra razionalità e nella nostra fiducia nella loro razionalità, e così via, è tutt'altro che illimitata. Se uno dei cacciatori non si fida completamente dell'altro B e in molti casi potrebbe avere buonissimi motivi per non fidarsi B allora la scelta razionale da parte sua potrebbe essere proprio quella della defezione, a dispetto dei suggerimenti della teoria classica dei giochi. Per questo potremmo riferirci alla caccia al cervo, e a giochi analoghi nei quali la mancanza di una completa fiducia reciproca potrebbe indurre i giocatori alla defezione, con il termine *dilemma della sfiducia*. Nell'analisi della caccia al cervo, e di altri dilemmi della sfiducia, un ruolo fondamentale viene svolto dalla nozione, che ci accingiamo a spiegare, di *risk dominance*.

La possibilità che uno dei partecipanti alla caccia al cervo scelga la defezione è legata alla circostanza che uno dei due equilibri, cioè (*cervo, cervo*), anche se più vantaggioso dell'altro, cioè di (*lepre, lepre*), è di gran lunga più rischioso. Supponiamo, per esempio, che C1 scelga la cooperazione, cioè *cervo*: se anche C2 sceglierà *cervo*, allora C1 otterrà un payoff pari a 4, ma se C1 defezionerà allora C1 resterà a mani vuote, con un payoff pari a 0. Poiché anche C2 può fare lo stesso ragionamento occorre concludere che (*cervo, cervo*) è un equilibrio assai rischioso, nel senso che il payoff di ciascuno dipende in grande misura dalla mossa dell'altro. Al contrario, l'equilibrio (*lepre, lepre*) è del tutto immune da rischi. Supponiamo, per esempio, che C1 scelga la defezione, cioè *lepre*: in tal caso, quale che sia la mossa di C2 B C1 otterrà un payoff pari a 3. Supponiamo ora che C1 abbia una limitata fiducia nella razionalità di C2 e attribuisca un'uguale probabilità all'eventualità che C2 scelga *cervo* o *lepre*: in tal caso C1 farebbe bene a cacciare la lepre, poiché è meglio avere la sicurezza di ottenere un payoff pari a 3 che avere il 50% di probabilità di ottenere un payoff pari a 4 e il 50% di ottenerne uno pari a 0. Se le condizioni epistemiche della caccia al

cervo sono quelle appena descritte, diremo allora che l'equilibrio non-cooperativo (*lepre, lepre*) è *risk dominant* rispetto all'equilibrio cooperativo (*cervo, cervo*).

Un'adeguata teoria epistemica dei giochi consiglia la scelta dell'equilibrio *risk dominant*, cioè, in questo caso, dell'equilibrio non-cooperativo. B ogniqualvolta a fiducia nella possibilità che il nostro socio cooperi non supera una determinata soglia. Ciò significa che possiamo prevedere che il risultato della caccia al cervo sarà l'equilibrio cooperativo (*cervo, cervo*) solo nel caso in cui giocatori hanno un'elevata fiducia reciproca.

È interessante osservare che, mentre nel dilemma del prigioniero le vie della cooperazione sono ostruite dalla razionalità dei giocatori, e dalla loro fiducia nella razionalità dei loro soci, nella caccia al cervo la tentazione spesso irresistibile alla defezione viene dalla sfiducia dei giocatori nella razionalità dei loro soci. Le ricerche empiriche finora condotte su diverse interazioni sociali che esibiscono la struttura formale della caccia al cervo confermano che, a dispetto dei dettami della teoria classica dei giochi, non sempre i giocatori raggiungono un equilibrio cooperativo, e che la fiducia reciproca svolge un ruolo cruciale nella realizzazione di tale equilibrio.

L'EVOLUZIONE DELLE NORME DI COOPERAZIONE ALTRUISTICA NELLA CACCIA AL CERVO. La teoria epistemica dei giochi consente, come si è appena visto, di comprendere i meccanismi che stanno alla base dei frequenti fallimenti della cooperazione nella caccia al cervo e in altri dilemmi della sfiducia. È però consolante osservare che gli accordi che sorreggono la caccia al cervo e molte altre attività cooperative della vita sociale vengono molto spesso rispettati. Negli ultimi dieci anni i meccanismi della cooperazione e della defezione nella caccia al cervo sono stati approfonditamente analizzati nell'ambito della teoria evoluzionistica dei giochi. Rinviando il lettore interessato a una conoscenza approfondita dell'argomento al volume di Skyrms (2004), ci limitiamo qui a notare che tali ricerche, condotte sulla base di procedimenti di simulazione analoghi a quelli sopra descritti con riferimento al dilemma del prigioniero, hanno mostrato che anche nella caccia al cervo, in un'ampia varietà di condizioni, l'evoluzione di una popolazione di giocatori che si incontrano per un numero indefinito di volte conduce all'emergere e alla propagazione di norme di cooperazione altruistica.

3.3. *Problemi di equità. Il gioco dell'ultimatum e l'evoluzione delle norme di giustizia*

IL GIOCO DELL'ULTIMATUM. PROBLEMI DI EQUITÀ. Nella discussione filosofica, i termini 'giustizia distributiva' ed 'equità' vengono spesso usati con riferimento al problema di come distribuire i beni comuni tra i membri di una comunità. I principi della giustizia distributiva sono illustrati dai giochi di contrattazione, nei quali occorre suddividere tra due o più persone qualche bene posseduto in comune; il payoff che un giocatore attribuisce ai possibili risultati del gioco è semplicemente la quantità di bene ricevuto.

Nel volume *The Evolution of Social Contract* (1996), Brian Skyrms esamina l'evoluzione della giustizia distributiva nell'intento di spiegare come potrebbero essersi sviluppati determinati comportamenti improntati a *norme di giustizia*, cioè a *criteri di equità distributiva*. A tale scopo, analizza una serie di semplici giochi di contrattazione tra i quali il *gioco dell'ultimatum*. Una vasta letteratura psicologica indica chiaramente che i soggetti sperimentali che partecipano a questo gioco rivelano una stabile inclinazione per un comportamento che può venire descritto come giusto, cooperativo, o altruistico, e che si trova in conflitto con le norme basate sulla teoria classica dei giochi. Un obiettivo fondamentale di Skyrms è quello di spiegare l'alta frequenza di comportamenti di questo tipo in base alla teoria evoluzionistica dei giochi.

Nel gioco dell'ultimatum occorre suddividere un bene, per esempio una somma di dieci dollari, tra due giocatori. Uno dei giocatori, che chiameremo *Proponente*, deve dare l'ultimatum, cioè fare una proposta all'altro giocatore, che chiameremo *Controparte*. Quest'ultimo può solo accettare o rifiutare la proposta: se l'accetta il Proponente ottiene quello che chiede e la Controparte quello che resta, se la rifiuta nessun giocatore ottiene nulla.

Svariati esperimenti hanno mostrato che molti soggetti affrontano il gioco sulla base di principi di equità che includono non solo norme per effettuare proposte eque quando i soggetti svolgono il ruolo di Proponente, ma anche norme per punire offerte inique quando svolgono il ruolo di Controparte. Ciò significa che molti soggetti sono disposti a rinunciare a un dollaro o due per punire un Proponente ingordo che chiede di avere otto o nove dollari su dieci. È evidente che i soggetti che agiscono sulla base di norme di questo tipo violano l'imperativo di massimizzare il guadagno monetario atteso, che caratterizza la razionalità economica, codificata dalla teoria della scelta razionale.

Skyrms applica la teoria evoluzionistica dei giochi a una versione semplificata del gioco dell'ultimatum in cui occorre dividere dieci dollari e ciascun giocatore ha soltanto due opzioni: il Proponente può chiedere per sé cinque dollari oppure nove, e la Controparte può accettare o rifiutare la

proposta. Si noti, tuttavia, che mentre il Proponente ha solo due strategie a propria disposizione, ‘Chiedi 9!’, ‘Chiedi 5!’, la Controparte ne ha quattro: ‘Accetta tutto!’, ‘Rifiuta tutto!’, ‘Accetta se il Proponente chiede 5, e rifiuta se chiede 9!’, ‘Accetta se il Proponente chiede 9, e rifiuta se chiede 5!’.

Immaginiamo una popolazione in cui gli individui giocano ripetutamente tra loro, talvolta nel ruolo di Proponente e talvolta in quello di Controparte, e supponiamo che ciascun individuo adotti una strategia fissa, che gli dice cosa fare in ciascun ruolo. Vi saranno otto strategie da considerare, che caratterizzano otto tipi di individui. Ad alcuni di questi, particolarmente interessanti nel contesto della nostra discussione, conviene attribuire un nome: lo Spregiudicato chiede 9 nel ruolo di Proponente e accetta tutto nel ruolo di Controparte; l’Imparziale chiede 5 e accetta la richiesta di 5, ma rifiuta quella di 9; il Cane Arrabbiato chiede 9 e accetta la richiesta di 9, ma rifiuta quella di 5; il Disinteressato chiede 5 e accetta tutto.

In uno dei modelli dinamici per il gioco dell’ultimatum esplorati da Skyrms si assume che in ogni turno di gioco le coppie siano formate in modo casuale, e che sia determinato nello stesso modo anche il ruolo assunto da ogni individuo. Inoltre, si suppone che i dollari guadagnati nel gioco rappresentino qualche vantaggio evolutivo, in grado di influenzare la numerosità della prole, a cui viene trasmessa la stessa strategia del genitore. Sulla base di queste assunzioni, che corrispondono alla dinamica dei riproduttori della teoria evuzionistica dei giochi, si può programmare un computer allo scopo di simulare questa dinamica e osservare come evolveranno popolazioni in cui sono rappresentate porzioni variabili delle otto strategie.

Simulazioni di questo genere conducono a conclusioni illuminanti, e spesso inattese. Per esempio, se nella popolazione iniziale sono rappresentate proporzioni uguali di ciascuna strategia, si trova, come è naturale aspettarsi, che il processo dinamico conduce, dopo un certo punto, all’estinzione degli Imparziali. Tuttavia, piuttosto sorprendentemente, si scopre anche che gli Spregiudicati non giungeranno a dominare l’intera popolazione, che evolverà invece verso uno stato polimorfico con circa l’87% di Spregiudicati e il 13% di Cani Arrabbiati. Ciò che sorprende è la persistenza della stravagante strategia dei Cani Arrabbiati che rifiutano le offerte eque e accettano quelle inique.

In base alla teoria classica dei giochi, la strategia degli Cani Arrabbiati è irragionevole poiché è dominata da altre strategie. Un aspetto interessante della teoria evuzionistica dei giochi, basata sulla dinamica dei riproduttori, consiste proprio nel fatto di consentire, in determinate condizioni, la sopravvivenza di strategie anomale, cioè di strategie dominate.

Ci possiamo ora chiedere se esistono condizioni in cui un’altra strategia anomala, vale a dire quella degli Imparziali B che si basa su norme di giustizia B possa sopravvivere. A questo proposito, Skyrms dimostra che, per evitare l’estinzione degli Imparziali, è sufficiente che si attribuiscono valori

appropriati alle proporzioni con cui le diverse strategie sono rappresentate nella popolazione iniziale. Per esempio, se si parte con una popolazione composta per il 30% da Imparziali, e il resto equamente suddiviso tra le altre sette strategie, allora la dinamica evuzionistica condurrà la popolazione ad uno stato polimorfico con circa il 64% di Imparziali e il 34% di Disinteressati. Ciò significa che la strategia anomala degli Imparziali può non solo sopravvivere, ma anche prosperare.

Un'interessante estensione del modello dinamico sopra illustrato si basa sull'assunzione che gli accoppiamenti non siano effettuati in modo completamente casuale, bensì in base a qualche *meccanismo di correlazione*, tale che, come sembra accadere in natura, individui che usano la stessa strategia tendano a giocare tra di loro più spesso che con individui che usano una strategia diversa. Utilizzando questo modello dinamico 'correlato', Skyrms riesce a dimostrare che, in condizione favorevoli di correlazione, possono ottenere un straordinario successo non solo i comportamenti imparziali, ma anche quelli altruistici, in cui un soggetto accresce il guadagno degli altri giocatori a spese del proprio.

4. *Natura ed evoluzione delle norme nelle teorie epistemiche dei giochi*

In questi ultimi quindici anni, le indagini sui problemi relativi alla natura e all'evoluzione delle norme sono stati caratterizzati dalla crescente integrazione tra la prospettiva delle teorie evuzionistiche e quella delle teorie epistemiche dei giochi. Un buon esempio di queste recenti direzioni di ricerca è costituito dai lavori di Cristina Bicchieri (1993 e 2006).

NATURA DELLE NORME. Riprendendo una tesi formulata da David Lewis (1969) ed Edna Ullmann-Margalit (1977), Bicchieri sostiene che le norme sociali dipendono dalle credenze e dalle preferenze di coloro che vi si conformano. Questa idea viene formalmente espressa nella seguente definizione:

NORMA SOCIALE. Data una regolarità N del comportamento individuale in una determinata popolazione, diremo che N è una *norma sociale* nel caso in cui:

- (1) Quasi tutti *preferiscono* conformarsi a N a condizione (e solo a condizione) che quasi tutti si conformino a loro volta.
- (2) Quasi tutti *credono* che quasi tutti si conformino a N .

La condizione (1) equivale ad affermare che una norma sociale N è un equilibrio di Nash, cioè una combinazione di strategie in cui la strategia di

ognuno è la risposta ottimale alle strategie altrui. La condizione (2) esprime la credenza di quasi tutti i membri della popolazione che, fra gli equilibri possibili, quasi tutti sceglieranno N . Se si accetta questa definizione di norma sociale occorre spiegare in che modo possa emergere una norma N , cioè in che modo quasi tutti i membri di una popolazione possano convergere nell'aspettativa che quasi tutti si conformeranno a N .

GENESI ED EVOLUZIONE DELLE NORME. L'elaborazione di adeguati modelli formali per la nascita delle norme sociali è un problema distinto da quello dell'evoluzione di tali norme, una volta che esse siano, in qualche modo, già emerse in una popolazione. Inoltre, sfortunatamente, come osserva Bicchieri (1993, p. 242), la comprensione dei motivi per cui le norme sociali persistono nel tempo, "non getta alcuna luce sul processo di formazione delle norme"; infatti, "il problema di *come* emergono le norme è diverso dal problema del perché esse emergono e si stabilizzano" (ibidem).

Secondo Bicchieri il processo di formazione delle norme sociali si svolge attraverso due stadi: nel primo, una norma emerge all'interno di piccoli gruppi, nel contesto di ripetute interazioni strategiche governate dalla razionalità individuale e da specifici processi di apprendimento; nel secondo, la norma può estendersi all'intera popolazione attraverso determinati meccanismi evolutivi. Il processo di apprendimento attraverso il quale una regola di comportamento emerge all'interno di un piccolo gruppo è qualcosa di molto diverso dalla dinamica evolutiva attraverso la quale, in condizioni favorevoli, essa può estendersi all'intera popolazione, fino a diventare una norma sociale. L'analisi del processo di formazione delle norme richiede, quindi, l'applicazione sia delle teorie epistemiche dei giochi sia della teorie evoluzionistiche: le prime verranno impiegate per studiare l'emergere delle norme all'interno di piccoli gruppi, le seconde per analizzare la loro propagazione ed evoluzione nell'intera popolazione.

L'idea fondamentale su cui si basa la spiegazione, suggerita da Bicchieri, della *genes* delle norme all'interno di piccoli gruppi, è quella che gli individui possono venire rappresentati come agenti in grado di apprendere, attraverso interazioni ripetute, con che tipo di avversario hanno a che fare e, quindi, ad affrontarlo in modo adeguato. A partire da questo genere di apprendimento, derivante dall'interazione all'interno di piccoli gruppi, si possono creare regolarità di comportamento suscettibili di diffondersi, attraverso particolari dinamiche evolutive, in tutta la popolazione, così da divenire norme sociali. Si noti che rappresentare "le norme sociali sono il risultato di un processo di apprendimento in un contesto di interazione strategica" significa accettare la tesi che esse "dipendono dalle scelte e, in ultima analisi, dalle preferenze e credenze individuali" (ibid. p. 242).

Bibliografia

- Axelrod, R. (1984), *The Evolution of Cooperation*, New York, Basic Books.
Trad. it. *Giochi di reciprocità. L'insorgenza della cooperazione*, Feltrinelli, Milano, 1985.
- Axelrod, R. (1997), *The Complexity of Cooperation. Agent-Based Models of Competition and Collaboration*, Princeton University Press, Princeton.
- Barry, B. (1965), *Political Argument, International library of philosophy and scientific method*. Routledge & Kegan Paul, London.
- Bicchieri, C. (1993), *Rationality and Coordination*, Cambridge, Cambridge University Press. Trad. it. *Razionalità e azione collettiva*, Feltrinelli, Milano, 1998.
- Bicchieri, C. (2006), *The Grammar of Society. The Nature and Dynamics of Social Norms*, Cambridge University Press, Cambridge.
- Braithwaite, R. B. (1955), *Theory of games as a tool for the moral philosopher*, Cambridge University Press, Cambridge.
- Fehige C. e Wessels U. (a cura di) (1998), *Preferences*, Walter de Gruyter, Berlino.
- Festa, R. (1999), 'Modelli evolutivisti del contratto sociale', in *Etica & Politica / Ethics & Politics*, 1999, 1, <http://www.units.it/etica/1999>.
- Festa, R. (2000), "Preferenze, razionalità, etica", *Rivista di filosofia*, pp. 333-343.
- Festa, R. (2001), "Come evolvono le norme sociali: la prospettiva della teoria dei giochi", *Biblioteca della libertà*, pp. 75-98.
- Festa, R. (2003), "Minacce e promesse. Logica e metodologia della deterrenza", in G. Manganaro Favaretto (a cura di), *La guerra. Una riflessione interdisciplinare*, Edizioni Università di Trieste, Trieste, pp. 247-298.
- Festa, R. (2004), "Teoria dei giochi e strategie della deterrenza", in *L&PS - Logic and Philosophy of Science*, 2004,
- http://www.univ.trieste.it/episteme/L&PS_Vol2No1/festa_L&PS_Vol2No1.pdf
- Festa, R. (2006), "Giochi di società. La complessità sociale nelle teorie ABM e nelle teorie dei giochi", *élites* 3, pp. 15-30.
- Gauthier, D. (1967), "Morality and Advantage", *Philosophical Review*, pp. 460-475.
- Gauthier, D. (1969), *The Logic of 'Leviathan': the Moral and Political Theory of Thomas Hobbes*, Clarendon Press, Oxford.
- Gauthier, D. (1987), *Morals by Agreement*, Clarendon Press, Oxford.
- Harms W. e Skyrms, B. (2007), "Evolution of Moral Norms", in M. Ruse (a cura di), *Oxford Handbook on the Philosophy of Biology*, Oxford University

- Press, Oxford. Anche sul web:
<http://www.lps.uci.edu/home/fac-staff/faculty/skyrms/Skyrmspapers.html>
- Hargreaves Heap, S., Hollis, M., Lyons, B., Sugden, R. E Weale, A. (1996). *The Theory of Rational Choice. A Critical Guide*, Blackwell Publishers, Cambridge (Mass.), 1992. Trad. it. *La teoria della scelta razionale. Una guida critica*, Laterza, Roma-Bari.
- Harsanyi, J.C. (1955), “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility”, *Journal of Political Economy*, pp. 309-321.
- Hume, D. (1739), *A Treatise of Human Nature*. Edizione recente (a cura L. A. Selby-Bigge e rivista da P. H. Nidditch): Clarendon Press, Oxford. Trad. it. (a cura di E. Lecaldano) *Opere filosofiche, vol. I, Trattato sulla natura umana*, Laterza, Bari, 2004.
- Lewis, D. (1969), *Convention. A Philosophical Study*, Harvard University Press, Harvard. Trad. it. *Convenzione. Uno studio filosofico*, Bompiani, Milano, 1974.
- Maynard Smith, J. (1982), *Evolution and the Theory of Games*, Cambridge University Press, Cambridge.
- Rawls, J. (1971), *A Theory of Justice*, Harvard University Press, Cambridge. Trad. it. *Una teoria della giustizia*, Feltrinelli, Milano, 1986.
- Rousseau, J.-J. (1755), *Discours sur l'origine et les fondaments de l'inégalité parmi le hommes*. Trad. it. (a cura di M. Garin), in Jean-Jacques Rousseau, *Scritti politici, vol. I, Discorso sulle scienze e sulle arti - Discorso sull'origine e i fondamenti della disuguaglianza - Discorso sull'economia politica*, Laterza, Bari, 2005.
- Schelling, T. C. (1960/1980), *The Strategy of Conflict*, Harvard University Press, Cambridge, Mass. Seconda edizione, con una nuova prefazione: 1980. Trad. it. *La strategia del conflitto*, Bruno Mondadori, Milano, 2006.
- Skyrms, B. (1996), *The Evolution of Social Contract*, Cambridge University Press, Cambridge.
- Skyrms, B. (2004), *The Stag Hunt and the Evolution of Social Structure*, Cambridge University Press, Cambridge.
- Taylor, M. (1987), *The Possibility of Cooperation*, Cambridge University Press, Cambridge.
- Ullmann-Margalit, E. (1977), *The Emergence of Norms*, Clarendon Press, Oxford.
- Verbeek, B. e Morris, C. (2004), “Game Theory and Ethics”, *The Stanford Encyclopedia of Philosophy* (Winter 2004 Edition), a cura di E.N. Zalta, <http://plato.stanford.edu/archives/win2004/entries/game-ethics/>
- von Neumann, J. e Morgenstern, O. (1944), *Theory of Games and Economic Behaviour*, Princeton University Press, Princeton, N. J. Seconda edizione: John Wiley and Sons, New York, 1953.