

## Il problema della cooperazione spontanea, ovvero: perché l'altruista è buono?

Guido Ortona<sup>1</sup>, Università del Piemonte Orientale, 20 aprile 2005

Versione provvisoria, si prega di non citare senza il permesso dell'autore

### Introduzione e riassunto.

Dawkins (1980), un biologo evoluzionista noto fra i suoi simili per il suo carattere provocatorio, si pone la seguente domanda: perché lo zucchero è dolce? Non è affatto una domanda stupida; infatti la risposta è: "siccome gli zuccheri sono alimenti preziosi, siamo programmati a provare un grande piacere dal mangiarli, onde essere fortemente motivati a cercare di procurarceli". Lo zucchero è dolce (cioè "buono") perché è utile che lo sia. Qui noi ci poniamo un problema analogo: perché l'altruista è buono? In altri termini, perché in tutte le culture è giudicato buono chi si prodiga a favore degli altri? E poiché i buoni esistono, possiamo porre la questione anche in questi termini: perché la gente mediamente coopera spontaneamente più di quanto apparentemente previsto dalla teoria, cioè quando non sarebbe suo interesse immediato farlo? La risposta che darebbe (e anzi, che ha dato) Dawkins è che essere buoni aiuta la società, e quindi anche l'individuo; e quindi siamo stati programmati dall'evoluzione a essere -a volte- buoni. Ma rispetto al caso dello zucchero qui c'è un problema: e cioè che se gli altri sono buoni, a me conviene fare -di solito- il free-rider. La soluzione di questo problema è l'argomento di questa lezione. Come vedremo, la soluzione non è ancora consolidata, ma è in natura di una ragionevole congettura<sup>2</sup>; e richiede il contributo di almeno tre discipline, la biologia evolutiva, la teoria dei giochi e le neuroscienze, e anche di un po' di filosofia. Siamo sulla frontiera della conoscenza, anzi forse un po' oltre; il discorso sarà quindi relativamente poco rigoroso, ma auspicabilmente piuttosto suggestivo.

**Nota.** Parleremo moltissimo di dilemma del prigioniero; do per scontato che il lettore sappia cos'è, quindi mi limito a riassumerlo schematicamente come segue: un DP a due giocatori e due strategie è un gioco come quello che segue (C = cooperare, N = non cooperare):

		II	
		C	N
I	C	cc, cc	cn, nc
	N	nc, cn	nn, nn

con

a)  $nc > cc > nn > cn$

b)  $cn + nc < 2cc$

---

<sup>1</sup> [guido.ortona@sp.unipmn.it](mailto:guido.ortona@sp.unipmn.it)

<sup>2</sup> Una congettura è un po' meno di una teoria: è una spiegazione che appare plausibile dato lo stato delle conoscenze, ma per la quale non esiste ancora una evidenza empirica sufficiente.

**1. Il problema, parte prima: la gente coopera?** O meglio, è vero che la gente coopera più di quanto la teoria economica, e in particolare la teoria dei giochi, suggeriscono che dovrebbe fare? La risposta è sì, purché sia data una importante condizione, di cui diremo. Questa risposta è sperimentale e conclusiva, e quindi se la teoria non lo prevede, è lei che deve modificarsi. Cito solo due pezzi di letteratura, entrambi di grande interesse.

Il primo è il famoso torneo informatico di Axelrod (Axelrod, 1984). Axelrod chiese a varie persone, scelte sostanzialmente a caso, di inviargli delle strategie di gioco per un DP ripetuto, con valori di payoff indicati, da far giocare (una contro una) ciascuna contro tutte le altre. Vinse (nel senso che ottenne il maggior guadagno) la strategia T (per *tit-for-tat*, "occhio per occhio"), cioè *cooperare alla prima mossa e poi fare ciò che l'altro ha fatto la mossa precedente*. E' chiaro che contro un avversario che non coopera mai questa strategia dà un guadagno molto basso; ergo, se ha vinto è perché *nella realtà* la gente coopera spesso. La strategia tit-for-tat ha un'ovvia e interessante caratteristica: implica una punizione per chi smette di cooperare, anche se molto drastica, e cioè la fine della cooperazione continua. Tale punizione è costosa per chi la applica; infatti se l'avversario giocasse T e si accettasse di non vendicarsi si resterebbe sul sentiero della cooperazione, e si guadagnerebbe di più, come vediamo nello schema 1.

Schema 1. Nel primo caso I e II adottano T; nel secondo, II adotta T e I coopera sempre. In entrambi i casi, II a un certo punto non coopera, per "trembling hand"<sup>3</sup>. "c" sta per cooperare in un DP, "n" per non cooperare.

I	II	I	II
c	c	c	c
c	c	c	c
c	n	c	n
n	c	c	c
c	n	c	c
n	c	c	c

eccetera.

Il secondo pezzo di letteratura sono gli studi sperimentali sulla fornitura privata di beni pubblici. L'esperimento tipo è appunto un gioco *ripetuto* di fornitura privata di beni pubblici. Come è noto, tale gioco è assimilabile a un dilemma del prigioniero: a ciascun soggetto conviene fare il free-rider, ma se tutti producono almeno una data quantità di bene pubblico si ha un esito efficiente rispetto a quello che si ha se tutti si comportano da free-rider. Il risultato è consolidato, e riassumibile come segue:

- a) Inizialmente la fornitura è significativa;
- b) Man mano che il gioco si sviluppa, le strategie tendono all'equilibrio di Nash (non fornire nulla), a meno che
- c) Sia possibile per i partecipanti punire (a un costo) i non cooperatori; nel qual caso la cooperazione iniziale, o comunque un elevato livello di cooperazione, vengono mantenuti, nel senso che non è presente un trend di riduzione<sup>4</sup>.

<sup>3</sup> Col termine "trembling hand" si indica la possibilità di una deviazione per errore dalla strategia adottata.

<sup>4</sup> La letteratura è enorme; se vi interessa approfondire l'argomento potete cominciare da Decker e altri, 2002,

Avrete intuito che in entrambi i casi la condizione cruciale di cui parlavamo prima è la possibilità di punire i non cooperatori. Questa punizione in linea di principio può sostituire l'intervento esogeno (per esempio dello stato), che come sappiamo è condizione necessaria e sufficiente perché i giocatori cooperino in un DP. Se i non cooperatori si aspettano di essere puniti dagli altri, coopereranno. In un certo senso, che non sarebbe piaciuto a Luigi XIV e nemmeno a Hobbes, ma sarebbe però piaciuto a Hume e a Rousseau, ogni cittadino può dire "l'état c'est moi".

**2. Il problema, parte seconda: perché si coopera, parte prima.** Ma qui si apre un problema teorico fondamentale: perché un soggetto dovrebbe darsi la briga di punire un non cooperatore, incorrendo per ciò in un costo? Sappiamo che gli converrebbe fare il free-rider, lasciando che siano gli altri a punire il non cooperatore. La spiegazione che avrebbe suggerito Kant è "perché c'è in noi una norma morale innata, che ci spinge a comportarci correttamente, e quindi fa sì che siamo disposti a pagare qualcosa perché la giustizia si affermi". Questa spiegazione è sostanzialmente giusta, come vedremo. Ma la sua qualificazione ci darà dei contenuti piuttosto interessanti.

Il principale problema con la spiegazione kantiana è che non è una spiegazione, ma un'assunzione. In altri termini, quando Kant dice "la legge morale è in noi", è del tutto sensato porgli la domanda che sicuramente gli porrebbe (e probabilmente gli ha posto) un bambino di tre anni, e cioè "già, ma perché la legge morale è in noi?". Una spiegazione possibile è che l'origine di ciò sia divina, ma Kant stesso ci ha insegnato che questa spiegazione non è accettabile sul piano scientifico. E su questo piano sono state fornite inizialmente due spiegazioni; ma come vedremo entrambe insoddisfacenti.

La prima è stata proposta da Gauthier, in un volume dal significativo titolo "Morals by Agreement" (1986). Il ragionamento è il seguente: si può essere o "massimizzatori puri" (P) o "massimizzatori vincolati" (V). Un P non coopera nel dilemma del prigioniero; un V coopera con i V e non con i P. Cosa capita allora? Supponiamo che il gioco sia (con guadagni positivi)

	c	n
c	3,3	1,4
n	4,1	2,2

inoltre, supponiamo inizialmente che sia subito evidente il tipo di giocatore. Un P guadagna sempre 2, ma un V guadagnerà 3 quando incontra un V e 2 quando incontra un P. Se la frequenza relativa di V è positiva, converrà essere V.

Supponiamo ora che il tipo di giocatore non sia sempre distinguibile. Per non complicare troppo i calcoli, ammettiamo che quando due P si incontrano si riconoscono sempre, e solo i V possono sbagliare. Sia allora:

p = probabilità che un V riconosca un altro V quando lo incontra;  
q = probabilità che un V scambi un P per un V quando lo incontra;  
r = probabilità di incontrare un V.

Avremo allora le seguenti utilità attese:

---

e dalla letteratura ivi citata; oppure da qualsiasi manuale di economia sperimentale.

$$U(V) = r[3pp + 4p(1-p) + 1p(1-p) + 2(1-p)(1-p)] + 1(1-r)q + 2(1-r)(1-q) = \\ = rp - q(1-r) + 2$$

$$U(P) = 2r(1-q) + 4rq + 2(1-r) = \\ = 2(1+rq)$$

E quindi converrà essere V se (potete fare i passaggi per esercizio)

$$p/q > 1 + 1/r$$

cioè -plausibilmente- tanto più facilmente quanto più è facile riconoscere un V (p è alto), è difficile confondere un P con un V (q è basso), ed è difficile incontrare un P (r è alto).

Questa soluzione è utile perché apre la strada alla spiegazione del perché in certe situazioni prevale l'altruismo e in altre no. Tuttavia, rimane aperto il problema del perché un V non defezioni sapendo che l'altro è V: siamo tornati al punto di partenza! Di fatto, Gauthier ammette che essere V o P è qualcosa di non modificabile; il che porta direttamente alla spiegazione successiva.

**3. Il problema, parte seconda: perché si coopera, parte seconda.** La seconda spiegazione è sociobiologica. L'ipotesi di fondo della sociobiologia è che la specie umana ha evoluto per selezione naturale non solo i tratti fisici e comportamentali individuali più idonei, ma anche le istituzioni sociali e i comportamenti collettivi più idonei. Per "più idonei" si intende naturalmente più idonei alla massimizzazione dell'utilità individuale, data la nostra forma tipica di organizzazione sociale, quella di cacciatori-raccoglitori uniti in piccole orde fortemente imparentate fra di loro; infatti, abbiamo vissuto in questo modo per circa il 99% della nostra storia biologica<sup>5</sup>. Stando così le cose, dobbiamo avere sviluppato un qualche istinto per la prosocialità, o almeno, come suggerisce Gallino (1980), all'apprendimento della prosocialità, per il semplice motivo che in assenza di esso le comunità di cacciatori-raccoglitori sarebbero degenerate in gruppi troppo piccoli per sopravvivere, e si sarebbero estinte. Questa ipotesi si presta a tre considerazioni molto interessanti.

La prima è che questa è una formulazione laica<sup>6</sup> del principio kantiano: la norma morale è effettivamente in noi, anche se per motivi puramente materialistici. La seconda è che questa spiegazione, come vedremo, a quanto pare è giusta, nel senso che esiste un'evidenza empirica piuttosto probante che la suffraga. Ma la terza, ahimè, è che anch'essa presta il fianco alla stessa critica che abbiamo<sup>7</sup> rivolto a Gauthier, come ora vedremo.

Il problema è che non basta dire "conviene all'orda che i suoi membri cooperino in modo altruistico, e quindi ciò si verificherà". Se infatti a ogni membro dell'orda conviene comportarsi da free-rider, ci aspettiamo che lo faccia. Il risultato sarà naturalmente inefficiente, e l'intera orda ne sarà danneggiata: ma la teoria dei giochi, e in particolare il

<sup>5</sup> Chi fosse interessato ad approfondire questo discorso può leggere Wilson (1979 o 1980).

<sup>6</sup> Ma non necessariamente tale: Dio può avere indotto il nostro istinto prosociale; a dimostrazione che la spiegazione religiosa, come suggerito da Kant, non è falsificabile.

<sup>7</sup> In realtà la critica a Gauthier è ripresa da Hargreaves Heap e Varoufakis (1995).

dilemma del prigioniero, ci insegnano che questo esito negativo è perfettamente plausibile. In altri termini, perché la teoria Kant-Wilson, chiamiamola così, sia accettabile bisogna dimostrare (o almeno *congetturare*; in effetti arriveremo solo fino a questo punto) che essa è compatibile con il comportamento di soggetti razionali in un dilemma del prigioniero ripetuto.

Infatti, abbiamo visto che il comportamento che occorre spiegare è la cooperazione *altruistica*, cioè che implica una perdita per chi la attua, ma al tempo stesso un guadagno di efficienza se *tutti* la praticano. Quindi siamo nel contesto di un dilemma del prigioniero. Se il DP non fosse ripetuto, la strategia ottimale per un soggetto razionale sarebbe univocamente non cooperare, e quindi il nostro problema non avrebbe soluzione. Ma non è questo il contesto rilevante: i comportamenti sociali sono tipicamente ripetuti, e quindi il contesto giusto è quello di un gioco ripetuto. Il problema, che sarà l'oggetto della nostra discussione, è allora duplice:

a) stabilire se esistono EN in un DP ripetuto che implicano cooperazione;

b) stabilire sotto che condizioni ci aspettiamo che essi vengano effettivamente selezionati.

Se la risposta alla prima domanda è "sì", e se le condizioni individuate dalla seconda sono coerenti con quelle supposte dalla teoria Kant-Wilson, avremo trovato dei solidi argomenti a favore di questa teoria.

Per fare un paragone, la teoria cosmologica dello stato stazionario in una certa fase è sembrata più plausibile di quella del big bang; ma la sua incapacità di spiegare la radiazione fossile di fondo l'ha resa inaccettabile. Analogamente, la teoria Kant-Wilson sembra molto plausibile; ma perché sia accettabile è necessario che possa spiegare la cooperazione spontanea in un DP ripetuto.

**4. L'apprendimento dell'equilibrio.** Per procedere ci occorrono alcune ipotesi sulla scelta delle strategie in un gioco ripetuto. Un ruolo cruciale è naturalmente svolto dall'apprendimento: i giocatori possono modificare la scelta della strategia sulla base della storia del gioco. Questo processo possiamo supporre che avvenga come segue.

a) I soggetti procedono per *tâtonnement* sulla base delle scelte proprie ed altrui, ed elaborano attese su cosa faranno gli altri. Ma quando si gioca una n-upla di strategie che costituiscono un EN, possiamo aspettarci che tali strategie vengano successivamente mantenute.

b) Un EN di strategia semplice funziona quindi da *attrattore* - se c'è. E se ci sono più EN? Allora quello che alla fine verrà selezionato dipenderà dalle regole del *tâtonnement*, e da come si comincia.

c) Se non ci sono EN di strategia semplice, allora verrà presumibilmente scelto l'EN di strategia mista. L'EN di strategia mista può fungere da attrattore anche *se ci sono* EN di strategia semplice.

In altri termini, se il gioco è ripetuto, e se l'apprendimento è possibile, alla lunga una comunità di giocatori adotterà una n-upla di strategie che costituisce un equilibrio di Nash.

In effetti questa condizione è troppo permissiva; la comunità adotterà una n-upla di strategie che costituisce un EN *non invadibile*, ovvero una *convenzione*, ovvero ancora una *strategia evolutivamente stabile* (ESS). Un insieme di strategie forma una convenzione (o una ESS) se non solo nessuno dei giocatori può fare meglio adottando singolarmente un'altra strategia (il che qualifica una ESS come un EN); ma non potrà fare meglio neanche

un giocatore che adotti una strategia *nuova*, cioè che cerchi di "invadere" il gioco precedente. Per fare un esempio, se "mettersi tutti insieme per cacciare il mammut usando delle trappole tradizionali" è una ESS, allora otterrà un esito peggiore non solo chi si metta a cacciare per conto suo, ma anche chi cercherà di usare trappole di nuova concezione. Se invece le nuove trappole avranno successo, la strategia *era* invadibile.

Il motivo per cui esigiamo questo requisito più forte è che l'evoluzione opera su tempi lunghi; quindi una strategia invadibile sarà prima o poi necessariamente invasa. I comportamenti cooperativi sopravvissuti devono quindi essere non invadibili.

Il concetto di *Evolutionary Stable Strategy* è stato introdotto dal biologo Maynard Smith (1982) con riferimento alla biologia. La sua definizione formale è identica a quella di *convenzione*, come definita da Sugden (1986) nell'ambito delle scienze sociali; secondo il quale una convenzione è una regola (non invadibile) che ha due requisiti: a) a nessuno conviene comportarsi in modo diverso, se anche gli altri non lo fanno e b) è una fra più regole che soddisfano questo requisito<sup>8</sup>.

Il prossimo passo è allora dimostrare che esistono insiemi di strategie<sup>9</sup> in un DP ripetuto che implicano cooperazione e che sono EN non invadibili, ovvero convenzioni, ovvero ESS.

**5. Strategie evolutivamente stabili (o convenzioni): definizione.** Indicando con  $E(X,Y)$  il guadagno atteso di chi adopera la strategia X contro la strategia Y, abbiamo che una strategia I è una ESS quando valgono le condizioni seguenti:

- a)  $E(I,I) \geq E(J,I)$
- b) se  $E(I,I) = E(J,I)$ , allora  $E(I,J) > E(J,J)$

per ogni strategia J diversa da I.

La spiegazione è più semplice di quanto non sembri. Intanto, notate che se la condizione a) è rispettata con segno forte è del tutto ovvio che I non è invadibile: se qualcuno prova un'altra strategia J otterrà un esito peggiore, e quindi J non potrà diffondersi. Il problema si pone allora solo se la disuguaglianza è debole. Supponiamo quindi una comunità di individui che giocano I. Improvvisamente arriva un individuo che gioca J. (Questo può succedere per mutazione genetica, per immigrazione, o per un errore casuale). Finché il giocatore di J è solo, potrà continuare a esistere, dato che ottiene lo stesso guadagno degli altri. Ma se ne arriva un altro, nasce la possibilità che si incontrino i due giocatori di J, con un guadagno minore. Il guadagno atteso se si gioca J diventa quindi minore che se si gioca I, e quindi la tendenza a giocare J non può diffondersi.

Fermiamoci un attimo per vedere cosa possiamo congetturare fin qui. Essenzialmente questo:

- a) una società composta da un piccolo gruppo di individui viventi in un ambiente sostanzialmente stabile e con una tecnologia sostanzialmente data tenderà ad assumere

---

<sup>8</sup> Il che fa sì che questa definizione corrisponda al senso comune: mi conviene tenere la destra, se tutti lo fanno, ma "tenere la sinistra" è una regola altrettanto valida.

<sup>9</sup> Poiché considereremo solo due alternative, "cooperare" e "non cooperare", e poiché perché si abbia cooperazione occorre che tutti cooperino, parleremo d'ora in poi di "strategie" e non di "insiemi di strategie", indicando con questo termine la strategia che *ciascun* giocatore adotterà.

norme di comportamento che sono ESS (o convenzioni).

b) Le condizioni precedenti sono quelle proprie dell'umanità, tranne che per un brevissimo periodo.

c) Se ammettiamo che le norme di comportamento ereditate dalla nostra fase preistorica sono divenute istintuali<sup>10</sup>, oppure che esse vengono ancora insegnate di generazione in generazione per inerzia, dato il piccolo numero di generazioni trascorso da quando hanno cessato di essere valide, tali norme sono valide ancora oggi. Per ora diamolo per buono, torneremo su questo punto più avanti.

E' in questo contesto che dobbiamo risolvere i due problemi del paragrafo 3, che ripeto qui:

I) stabilire se esistono ESS in un DP ripetuto che implicino cooperazione;

II) stabilire sotto che condizioni ci aspettiamo che essi vengano effettivamente selezionati.

Infatti se la risposta a I) è "sì" e le condizioni richieste da II sono plausibili, il punto c) ci autorizzerà a ritenere che sia presente in noi una tendenza, innata o appresa di generazione in generazione, alla cooperazione spontanea.

**6. Strategia T, equilibri di Nash, convenzioni.** Cominciamo dal punto I. Si dimostra (troverete la dimostrazione nell'appendice 1) che la strategia T è un EN. Però è invadibile. Infatti, contro un soggetto che adotti T un soggetto che adotti la strategia "cooperare sempre" (chiamiamola B, per "bene") farà meglio, come abbiamo visto nello schema 1, purché sia possibile un errore di trembling hand - come è sensato ammettere. Poco male, direte voi, anzi, meglio: si coopera ancora di più. Il guaio è che la strategia B è invadibile dalla strategia "non cooperare mai" (M, per male): infatti in un DP ripetuto fra un giocatore M e un giocatore B il primo otterrà sempre il massimo payoff, e il secondo sempre il minimo.

Tuttavia esistono altre strategie che producono cooperazione continua, e che *sono* ESS. La più semplice è la strategia T': "cooperare la prima volta, e poi se per qualsiasi motivo una volta non coopero accetto di cooperare mentre l'altro non coopera". La successione di mosse che si produce è quella dello schema 2.

Schema 2. I giocatori adottano T'. A un certo punto il giocatore II gioca N.

I	II
C	C
C	C
C	N
N	C
C	C
C	C

eccetera

La dimostrazione che T' è una ESS è molto semplice. Supponiamo che I adotti T' e II un'altra strategia. Poiché T prevede "cooperare sempre" (in assenza di errore) questa altra

---

<sup>10</sup> Come è capitato per lo zucchero: la nostra elevata motivazione a mangiarlo era funzionale quando trovarlo era difficile, adesso è facile e quindi ne mangiamo tendenzialmente troppo, a dimostrazione che la nostra regola comportamentale non si è (ancora) adattata.

strategia per essere appunto un'altra deve o prevedere di cooperare anche se l'altro non coopera la volta prima (come la strategia B) oppure prevedere a un certo punto di non cooperare. Nel primo caso II ottiene cc anziché nc, quindi non gli conviene. Nel secondo caso la volta dopo I non coopera, e quindi se II riprende a cooperare subito dopo (come nello schema 2) ottiene cn+cn anziché 2cc, e non gli conviene; se invece continua a non cooperare ancora una volta produce una successione siffatta:

I	II
C	C
C	N
N	N
N	C
C	C

ed è ancora peggio, perché aggiunge un guadagno di nn al posto di cc; e questo ogni volta che continua a non cooperare. Quindi una strategia che preveda esiti diversi da T' farà comunque peggio. E' importante (come vedremo) che anche le strategie T'' (se uno sgarra, poi coopera due volte contro un non cooperatore), T''' (tre volte) ecc. sono ESS (Sugden 1986, p.115).

Si noti il significato "reale" della strategia T': se un giocatore sgarra, non solo l'altro lo punisce (giocando N la volta dopo), ma il giocatore che ha sgarrato accetta la punizione.

E si noti anche che T' e derivati non sono naturalmente le uniche ESS disponibili. E' facile constatare che -per esempio- anche "non cooperare mai" non è invadibile.

**7. Quale ESS?** Abbiamo visto che una comunità sufficientemente stabile<sup>11</sup> adotterà necessariamente una ESS; e che esistono ESS che implicano cooperazione e altre che non la implicano. Il problema è quando verrà adottata una ESS del primo tipo e quando una del secondo. Qui dobbiamo abbandonare il rigore proprio della teoria dei giochi e procedere in modo più "umanistico".

Possiamo ragionevolmente aspettarci che un ESS di cooperazione sia tanto più facilmente adottata quanto più una comunità è appunto stabile; e diciamo che una comunità è tanto più stabile quanto più è piccola, i payoffs non mutano, e le ripetizioni sono frequenti. E' abbastanza ovvio che queste condizioni erano (e sono) effettivamente proprie delle comunità di cacciatori-raccoglitori in cui si sono evolute le nostre norme sociali<sup>12</sup>. Ci sono comunque altri due argomenti forti in favore dell'affermarsi di ESS di cooperazione nella preistoria.

Il primo è sociobiologico. Le orde di cacciatori-raccoglitori sono tipicamente composte di individui fortemente imparentati fra di loro. In queste condizioni, l'altruismo è naturale, in quanto i nostri "geni egoisti" (Dawkins, 1981) ci impongono di aiutare i nostri consanguinei.

Il secondo consiste in questo. Le orde di cacciatori-raccoglitori vivono tipicamente in

<sup>11</sup> Vedremo fra poco cosa ciò significa.

<sup>12</sup> Chi volesse approfondire il discorso della trasformazione di abitudini cooperative in convenzioni e in norme etiche può guardare Sugden, 1986, che l'autore stesso considera una rilettura moderna di Hume (1740).

condizioni di conflitto, e sovente di *warfare*, fra di loro (si veda per es. Harris, 1980). Ora, le comunità cooperative sono più efficienti di quelle non cooperative. E' quindi sensato supporre che queste ultime, se mai sono esistite, hanno avuto la peggio nella competizione con le prime, e si siano estinte. Un caso interessante, e non bellicoso, e raccontato nell'appendice 2.

A questo punto sembra che siamo arrivati alla soluzione del nostro problema: una comunità adatterà un insieme di ESS, ESS di cooperazione esistono ed è plausibile che siano state adottate in epoca preistorica, esse implicano la punizione dei non cooperatori, e questo ci ha lasciato in eredità una propensione a cooperare e a punire i non cooperatori. In realtà tuttavia ci sono ancora due problemi da risolvere; saranno oggetto dei prossimi due paragrafi.

**8. Il mancato costo della punizione.** Il primo problema è che nelle ESS di cooperazione che abbiamo visto punire *non è costoso*. Il soggetto che non coopera con chi in precedenza non ha cooperato ottiene il massimo guadagno possibile, *nc*.

Questo però dipende probabilmente solo dal fatto che abbiamo, per semplicità ma un po' impropriamente, utilizzato dei giochi 2x2. Ma in realtà non solo la fornitura privata dei beni pubblici, ma anche i "giochi preistorici" che possiamo realisticamente immaginare, come la caccia al mammut, sono tipicamente giochi a *n* soggetti; e in questi casi è perfettamente possibile che la punizione senza costo non sia praticabile. Nella caccia al mammut, un tipico cacciatore aveva evidentemente tutto l'interesse a fare il free-rider, ma se *tutti* l'avessero fatto si sarebbe rimasti senza mammut. Questo qualifica il gioco come un DP. Tuttavia, possiamo sensatamente supporre che il free-rider difficilmente avrebbe potuto compensare gli altri per la sua mancata partecipazione la volta precedente: certo non poteva cacciare il mammut da solo. In assenza di possibilità di compensazione, la punizione è molto probabilmente costosa.

Supponiamo allora una strategia di questo genere: se uno non coopera, punirlo anche se ciò ha un costo; e chi non punisce diventa a sua volta un non cooperatore. Se tutti adottano questa strategia, chi si comporti in modo diverso dovrà necessariamente o non cooperare o non punire; in entrambi i casi verrà punito, e quindi non gli converrà. E' quindi sensato supporre che una strategia siffatta sia una ESS, anche se non sono a conoscenza di dimostrazioni formali<sup>13</sup>.

**9. L'eredità biologica.** Il secondo problema è che abbiamo *supposto* ma non *dimostrato* che abbiamo ereditato dei comportamenti e dei valori che erano ottimizzanti nella nostra preistoria, o quantomeno la predisposizione biologica a "fare la cosa giusta", e quindi a trasformare convinzioni morali in scelte. Anche se questo passaggio sembra plausibile (dopotutto, in molti altri ambiti *abbiamo effettivamente* ereditato dei comportamenti o degli atteggiamenti non più ottimali e quindi dannosi, come il gusto per lo zucchero), sarebbe utile avere in mano qualcosa di più.

In nostro soccorso giungono le neuroscienze; in effetti, è in fase di codificazione una nuova branca dell'economia, la cosiddetta *neuroeconomics*. Essa consiste essenzialmente nello studio della attivazione cerebrale che avviene in soggetti sottoposti ad esperimenti di economia; vorrete scusare una possibile imprecisione terminologica. E per quanto

---

<sup>13</sup> In effetti questo è quanto sostiene (ma non dimostra) Axelrod (1986).

riguarda il nostro discorso due risultati, recentissimi, sono di grande importanza. Il primo (Rilling e altri, 2002) è la scoperta che la scelta della strategia cooperativa nel dilemma del prigioniero procura sensazioni piacevoli *di per sé*, indipendentemente dal payoff che consegue alla scelta. Il secondo, ancora più interessante e recente (De Quervain e altri, 2004) è la scoperta che analoghe sensazioni piacevoli si ottengono *quando ci si vendica di un non cooperatore*, di nuovo indipendentemente dal payoff<sup>14</sup>: "Our findings support the hypothesis that people derive satisfaction from punishing norm violations" (dall'abstract). Si tratta, ripeto di risultati recentissimi, in effetti ancora in attesa di ulteriori conferme. Ma se diverranno consolidati, saranno evidentemente un formidabile argomento a favore delle tesi sostenute in questa lezione.

**10. Conclusione. Luci ed ombre della cooperazione.** Abbiamo già scritto le conclusioni due paragrafi più in su, ma le riporto di nuovo qui per chiarezza: una comunità adotterà un insieme di ESS, ESS di cooperazione esistono ed è plausibile che siano state adottate in epoca preistorica, esse implicano la punizione dei non cooperatori, e questo ci ha lasciato in eredità una propensione a cooperare e a punire i non cooperatori. Questa doppia propensione fa sì che giudichiamo negativamente chi non coopera (anzi, che l'altruismo sia pressoché universalmente il valore principale su cui si misura lo *standing* morale di un individuo) e ci consente di costruire delle convenzioni di cooperazione. Coerentemente con quanto sopra, ci aspettiamo che la cooperazione spontanea sia particolarmente diffusa nelle comunità piccole e poco esposte al cambiamento. Approfondire questo punto sarebbe molto interessante, ma chiaramente fuori tema; ci basta osservare che l'evidenza empirica è chiaramente in linea con questa previsione.

Possiamo dedurre tre ulteriori conclusioni.

La prima è che *esiste un livello efficiente di punizione*. Abbiamo visto che ci sono diverse ESS di cooperazione, che implicano diversi tipi di punizione. Quale sia stata o sarà adottata dipende evidentemente da circostanze specifiche. Ma in ogni caso una punizione è necessaria. La convivenza umana richiede insomma non *massima bontà* ma *massima giustizia*.

La seconda è che, *a contrario* rispetto a quanto sopra a proposito della cooperazione in comunità piccole e stabili, ci aspettiamo che essa sia bassa in comunità grandi e sottoposte a forte mutamento; e che in situazioni del genere ci sia un elevato *stress* causato dal conflitto fra i valori etici che spingono all'altruismo e la massimizzazione dell'utilità che spinge al free-riding. Per usare una terminologia vetusta ma corretta, che si abbia cioè un conflitto fra *struttura* e *sovrastuttura* etica. Che le cose stiano normalmente così non pare dubbio.

L'ultima conclusione è ancora meno incoraggiante. Sappiamo che una ESS non è invadibile; in realtà ci sono due casi in cui una ESS lo è. Il primo è quello in cui l'invasione avvenga da parte di un *gruppo* di individui, per i quali le strategie disponibili comportano payoff diversi, e che sono quindi abituati a scelte diverse. Questi individui avranno probabilmente difficoltà ad adattarsi al nuovo ambiente; se il gruppo è numeroso, e/o l'inerzia comportamentale è elevata, è possibile che siano gli "indigeni" a doversi adattare<sup>15</sup>. Il secondo è quello in cui mutano i payoffs associati alle diverse strategie.

Vi sarete accorti di dove vogliamo andate a parare. In entrambi i casi, il mantenimento della cooperazione (e quindi dell'efficienza) implica che la comunità si difenda contro i

---

<sup>14</sup> La problematica della c.d. *punizione altruistica* è ben studiata in Ottone, 2004.

<sup>15</sup> Questo caso è stato studiato da Bissey e Ortona, 2002.

mutamenti. Possiamo quindi aspettarci che i membri di una comunità siano tanto più conservatori e xenofobi quanto più è importante la cooperazione spontanea. Purtroppo l'evidenza empirica conferma anche questo risultato; ci sono moltissimi esempi. I *comunitaristi* americani, che praticano un altruismo avanzatissimo nelle loro comunità, non riconoscono alcun diritto di ingerenza al governo, a causa delle sue pretese universalistiche. La Carinzia è probabilmente la regione più civile d'Europa per quanto riguarda la convivenza dei suoi cittadini; e ha espresso un appoggio ampiamente maggioritario ad Haider e alla sua opposizione spietata a qualsiasi mutamento, e in primis all'immigrazione<sup>16</sup>. In altri termini, la disponibilità alla cooperazione spontanea e la sua implementazione non implicano una disponibilità *erga omnes*. Siamo propensi a punire chi si comporta in modo non giusto, e il comportamento non giusto è quello egoista; ma solo nell'ambito di un gruppo di riferimento. Come ci racconta Ulisse (nel canto nono dell'Odissea), "I venti mi portavano sul mare, e mi spinsero al paese dei Ciconi, a Ismaro. E là io distrussi la città e menai strage fra gli abitanti. E dalla città prendevamo donne e ricchezze in abbondanza, e ce le dividevamo: nessuno andava via senza la sua parte giusta di preda" [la sottolineatura è aggiunta].

Dal punto di vista teorico, la situazione non è cambiata rispetto a tremila anni fa. L'universalismo non può essere la generalizzazione della nostra disponibilità a cooperare con i nostri simili. Richiede altri fondamenti. Quali? E' un problema di estremo interesse, ma non è il tema di questa lezione.

**Appendice 1. Dimostrazione che la scelta di T è un EN<sup>17</sup>.** La dimostrazione vale per il caso di gioco indefinito; ma notate che questo è un caso sostanzialmente realistico, a differenza di quello di ripetizione infinita, ed è il caso rilevante per le interazioni sociali significative. Essa inoltre è svolta per il caso a due giocatori, ma è generalizzabile. Consideriamo il dilemma del prigioniero tipico

	II	
	c	n
c	cc, cc	cn, nc
I		
n	nc, cn	nn, nn

dove c sta per *cooperare* e n per *non cooperare* o *defezionare*, e i guadagni sono in unità di utilità.

Il significato dei simboli dovrebbe essere chiaro: per esempio, nc sta a indicare il guadagno che ottiene un giocatore che non coopera contro uno che coopera.

Dimostriamo che la coppia (o n-upla) di strategie [T,T] (ovvero [T,T,T...]) è un equilibrio di Nash, purché siano date tre condizioni di cui ora diremo. Come sappiamo, ciò significa che contro un giocatore (o n-1 giocatori) che adotti (o adottino) T, l'altro giocatore non può ottenere un risultato migliore di quello che può ottenere giocando a sua volta T.

Le tre condizioni richieste sono:

<sup>16</sup> Chi è interessato ad approfondire questo argomento può partire da Ortona, 2001.

<sup>17</sup> Ripresa da Sugden, 1986.

- a) che il gioco sia indefinito, cioè che all'inizio di ogni ripetizione i giocatori non possano sapere se è l'ultima;
- b) che la probabilità, che chiamiamo  $p$  e che supponiamo costante, che il gioco continui sia sufficientemente alta (vedremo fra poco cosa vuol dire questo "sufficientemente");
- c) che il dilemma del prigioniero sia un DP *in senso stretto*, col che vogliamo dire che i payoffs sono tali per cui  $2cc > nc+cn$ . Questa condizione esclude il caso in cui sia preferibile alternare cooperazione e non cooperazione alla continua cooperazione. Essa è tanto più facilmente realizzata quanto più, *ceteris paribus*, il guadagno della reciproca cooperazione è alto e quello della defezione se l'altro coopera è basso.

Ecco la dimostrazione. Per cominciare, osserviamo che se i due giocatori adottano T cooperano sempre. Quindi, nessuna strategia che implichi che si cooperi sempre può dare un risultato migliore di T contro T, e insidiarne perciò lo *status* di equilibrio di Nash. Ne segue che dobbiamo preoccuparci solo di quelle che implichino almeno una volta non cooperazione.

Osserviamo allora che se la probabilità che il gioco continui la volta successiva è 0, e I adotta T e quindi coopera la prima volta, a II conviene non cooperare, quindi la BRS a T è n e non T, e quindi [T,T] non è un EN. Se la probabilità è invece 1, quando un giocatore decide di non cooperare instaura una successione di mosse in cui non si coopera fino a quando lui non decide di tornare a cooperare. Avremo allora per esempio:

	a	b	c
I II	I II	I II	I II
C C	C C	C C	C C
C C	C N	C N	C N
C C	N C	N C	N C
C C	C C	C C	C C
C C	C C	C C	C C

In (a) entrambi adottano T, in (b) II non collabora per una volta, e in (c) per due. Poiché il gioco è infinito (la probabilità che ogni volta continui è 1), il payoff totale è la somma dei payoffs. In (b) abbiamo, come unica differenza rispetto ad (a), che a un certo punto II guadagna  $nc+cn$  anziché  $cc+cc$ , e quindi ci rimette, in base alla terza ipotesi. In (c) è ancora peggio: guadagna a un certo punto  $nc+nn+cn$  contro  $cc+cc+cc$ . Quindi l'esito per II sarà *a fortiori* sempre peggiore al crescere del numero delle volte in cui non coopera. Cooperare sempre, e quindi T, è allora la strategia migliore, e [T,T] è un EN.

Ma se con  $p=0$  [T,T] non è un EN e con  $p=1$  sì, ci sarà un valore di  $p$  oltre il quale lo diviene.

Possiamo trovare  $p$  complicando un po' il discorso, come segue. Ammettiamo che I adotti sempre T. Se II per un po' non coopera, produce una serie di giocate siffatte:

- a) alla volta  $x$  lui non coopera e I si;
- b) poiché I adotta T, alla volta  $x+1$ esima I non coopererà: quindi si avrà N,N per un certo numero di volte;
- c) a un certo punto, II potrà tornare a cooperare: allora in quella volta I non coopererà e

II sì;

d) di lì in poi si torna a cooperare tutti e due.

Si danno allora due casi. Nel primo, a un certo punto II smette di cooperare per sempre. Allora il suo guadagno di lì in poi è:

$$nc + nn(p + p^2 + p^3 + \dots)$$

cioè, per un teorema sui limiti,

$$nc + nn[p/(1-p)]$$

mentre se non smette di cooperare avrà ovviamente

$$cc + cc(p + p^2 + p^3 + \dots)$$

che vale, per lo stesso teorema,  $cc/(1-p)$ . Allora conviene T anziché smettere di cooperare per sempre se

$$cc/(1-p) > nc + nnp/(1-p)$$

cioè se

$$cc > nc(1-p) + nnp$$

$$cc > nc - ncp + nnp$$

$$cc > nc + p(nn-nc)$$

$$cc - nc > p(nn - nc)$$

$$nc - cc < p(nc - nn)$$

ovvero (ricordate che moltiplicando per -1 si inverte il senso della disequaglianza)

$$p > (nc - cc)/(nc - nn) \quad [1]$$

(nel nostro esempio, la condizione è rispettata se  $p > 5/13$ ).

Se invece II a un certo punto torna a cooperare, avremo successioni come

	a	b	c	d
	I II	I II	I II	I II
1	C C	C C	CC	C C
2	C N	C N	CN	C N
3	N C	N N	NN	N N
4	C C	N C	NN	N N
5	C C	C C	NC	N N
6	C C	C C	CC	N C

7 C C C C C C

ecc. Ora, si noti che il caso (d) e il caso (c) sono identici fino alla quarta volta compresa, e l'unica differenza è nella quinta e nella sesta volta, dove si ha (N,N; N,C) in (d) e (N,C; C,C) in (c). Lo stesso vale per (b) e (c), solo che la differenza si ha alla quarta e alla quinta volta. Quindi possiamo impostare così il problema:

Se (N,C; C,C) conviene rispetto a (N,N; N,C) allora conviene smettere di non cooperare alla quinta volta anziché alla sesta, ma anche alla quarta anziché alla quinta, ecc. Quindi il numero più vantaggioso di volte in cui si ha (N,N) è 1. E' allora sufficiente individuare le condizioni sotto le quali T è preferibile per II rispetto a non cooperare una volta.

(N,C; C,C) è preferibile per II rispetto a (N,N; N,C) quando

$$\begin{aligned} cn + cc^*p &> nn + cn^*p \\ (cc - cn)^*p &> nn - cn \\ p &> (nn - cn)/(cc - cn) \quad [2] \end{aligned}$$

Se questa condizione è rispettata, dobbiamo allora confrontare la successione prodotta da [T,T], cioè (C,C; C,C) solo con quella del caso (a) (II non coopera una volta) e con quella del caso (b) (II non coopera due volte).

Rispetto al caso (a), al giocatore II converrà T se per lui (C,C; C,C) è preferibile a (C,N; N,C); e rispetto al caso (b) se (C,C; C,C; C,C) è preferibile a (C,N; N,N; N,C).

La condizione riguardo a (a) è allora che a II conviene T se

$$\begin{aligned} cc + ccp &> nc + cnp \\ (cc-cn)p &> nc-cc \\ p &> (nc-cc)/(cc-cn) \quad [3] \end{aligned}$$

e quella rispetto a (b) è che a II conviene T se

$$cc + ccp + ccp^2 > nc + nnp + cnp^2$$

poiché  $ccp^2 > cnp^2$ , la condizione è rispettata *a fortiori* se

$$\begin{aligned} cc + ccp &> nc + nnp \\ (cc-nn)p &> nc - cc \\ p &> (nc - cc)/(cc - nn) \quad [4] \end{aligned}$$

la [4] è identica alla [3]. E poiché il secondo membro della [3], o della [4], è superiore a quello della [1], se la [3] è rispettata lo è anche la [1]. Ergo, tenendo conto anche della condizione [2], II non può in nessun caso fare meglio di T se

$$p > (nc-cc)/(cc-cn) \quad \text{oppure se} \quad p > (nn-cn)/(cc-cn)$$

a seconda di quale è la probabilità più alta. Si noti tuttavia che tanto più il dilemma è "forte" (nc alto rispetto a cc, nn basso) tanto più è facile che la condizione che conta sia la prima, come nel nostro esempio numerico.

**Appendice 2. Un caso recente di conflitto fra convenzioni.** Mentre scrivevo questo testo, mi sono imbattuto per caso in un articolo (Paciotti e altri, 2005) che racconta un caso di (presumibile) competizione fra convenzioni cooperative e non cooperative particolarmente adatto ad esemplificare alcuni dei suoi contenuti; mi permetto quindi di riassumerne alcuni punti (e di consigliarvi di leggerlo).

I Sukuma della Tanzania sono un popolo dedito all'agricoltura ma soprattutto alla pastorizia, esercitata con grandi mandrie collettive. Probabilmente a causa di ciò hanno tradizionalmente praticato un livello di cooperazione più elevato dei vicini popoli più agricoli. Nel 1979 la guerra fra la Tanzania e l'Uganda ha creato un clima di anarchia, particolarmente propizio all'abigeato. La reazione dei Sukuma è stata quella di creare un sistema di giustizia autogestito, universalmente riconosciuto e in grado di funzionare anche a centinaia di chilometri di distanza, il cosiddetto *Sungusungu*. Esso si basa sull'accettazione del giudizio di magistrati eletti, e sull'applicazione diretta delle loro decisioni da parte dei cittadini (uso a proposito questo termine improprio). L'efficacia di questo strumento è stata tale da garantire ai Sukuma uno sviluppo economico e demografico assai superiore a quello delle etnie circostanti, meno cooperative al loro interno. Almeno una di esse ha provato a imitare il *Sungusungu*, ma senza successo, a causa del conflitto con la tradizione di comportamenti meno partecipativi. Gli autori concludono osservando che "la maggior parte degli individui ha una predisposizione prosociale. Quando sono influenzate da istituzioni sociali su vasta scala che enfatizzano la cooperazione e la condivisione, le persone collaborano per il bene del gruppo".

## **Bibliografia**

Axelrod, R. (1984), *The evolution of Cooperation*, Basic Books.

Axelrod, R. (1986), "An evolutionary approach to norms", *Am. Political Science Review*, 80.

Bissey, M.E. e G. Ortona (2002), *The integration of defectors in a cooperative setting*, "Journal of Artificial Societies and Social Simulation".

Dawkins, R. (1981), *Il gene egoista*, Zanichelli.

Decker, T., A. Stiehler e M. Strobel, *A comparison of punishment rules in repeated public goods games, an experimental study*, Masstricht Economic Research Institute on Innovation and Technology (MERIT), WP 018-2002.

De Quervain, D.J.F, U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck e E. Fehr (2004), "The neural basis of altruistic punishment", *Science*.

Gallino, L. (1980) *Oltre il gene egoista*, in K. Boulding (a cura di), *Sociobiologia e natura umana*, Einaudi.

Gauthier, D. (1986), *Morals by agreement*, Clarendon.

Hargreaves Heap, S.P. e Y. Varoufakis (1995), *Game theory: a critical introduction*, Routledge.

Harris, M. (1980), *Cannibali e re*, Feltrinelli.

Hume, D. (1740), *Trattato sulla natura umana*.

Maynard Smith, J. (1982), *Evolution and the theory of games*, Cambridge University Press.

Ortona, G. (2001), *Economia del comportamento xenofobo*, UTET Libreria.

Ottone, S. (2004), *Transfers and altruistic punishments in third party punishment game experiments*, Università del Piemonte Orientale, dip. POLIS, Working Paper n.46.

Paciotti, B., C. Hadley, C. Holmes e M.B. Mulder (2005), "L'armata del veleno", *Le Scienze*.

Rilling, J.K., D.A. Gutman, T.R. Zeh, G. Pagnoni, G.S. Berns e C.D. Kilts (2002), *A neural basis for social cooperation*, "Neuron".

Sugden, R. (1986) *The economics of rights, cooperation and welfare*, Basil Blackwell.

Wilson, E. O. (1979), *Sociobiologia: la nuova sintesi*, Zanichelli.

Wilson, E.O. (1980), *Sulla natura umana*, Zanichelli.