IL LIMITE PERMEABILE. LA COSTRUZIONE DELLO SPAZIO INTERSOGGETTIVO CONDIVISO NELLE RELAZIONI STRATEGICHE

This article focuses on the relational aspects of strategic interactions. First, we highlight how some of the limitations of the classical theory of games can hinder a deeper understanding of two fundamental dimensions of interpersonal relations, which are essential to our social epistemology: the mentalizing and empathizing processes. Secondly, we present the results of a series of experiments that stress the role of these two elements in the realm of strategic interactions. Finally, we argue that, by conceptualizing a hierarchy of higher order beliefs, psychological game theory seems to constitute a promising step forward towards the introduction of relational elements in the motivational structure of social agents and in the understanding of the intersubjective space we build and share when interacting with others.

di
VITTORIO PELLIGRA

Sophia III (2011-2) 195-211

L'incontro avviene tra stranieri,
altrimenti sarebbe parentela
(Emmanuel Lévinas)

1. Introduzione

La teoria dei giochi, nel suo disegno originario, rappresenta un tentativo di fondare la comprensione delle dinamiche sociali, e quindi anche interpersonali, su basi puramente razionali. Lo schema di analisi della teoria offre delle strutture matematiche descrittive e dei concetti di soluzione che si applicano a tutte le situazioni strategiche; quelle situazioni, cioè, nelle quali il risultato dell'azione di un individuo dipende dalla combinazione delle scelte di tale individuo e di quelle di tutti gli altri soggetti con i quali egli interagisce. Queste situazioni vengono definite in termini di "giochi" nei quali interagiscono soggetti che sono "individui razionali". Da queste due assunzioni, l'"individualità" e la "razionalità", e da una serie di altre ipotesi addizionali, è possibile far derivare delle procedure, degli algoritmi, che portano all'individuazione delle soluzioni dei giochi, che possiamo intendere come indicazioni relative alle scelte ottimali da compiere alla luce delle scelte ottimali degli altri giocatori. Dato questo schema di analisi, è facile intuire come all'interno di un gioco, la più importante abilità che viene messa in campo dai giocatori è la capacità di prevedere accuratamente il comportamento degli altri giocatori. Sia che si voglia cooperare oppure competere con gli altri, è necessario formarsi delle congetture circa le ragioni, le intenzioni e i desideri che spingono gli altri verso una scelta piuttosto che un'altra; questo per poter coordinare le nostre azioni con le loro, in caso di cooperazione, oppure per poter ottimizzare le nostre risposte tenendo conto di quelle degli altri, in caso di competizione.
Dovrebbe quindi apparire naturale che la comprensione del funzionamento di questo processo di anticipazione delle scelte altrui, sia messa al centro della teoria delle scelte strategiche. Eppure, risulta piuttosto sorprendente constatare come gli ultimi cinquant'anni di sviluppi teorici abbiano portato, in questo senso, a pochi progressi; e che piuttosto che spiegare il funzionamento del processo di previsione, i teorici abbiano preferito limitarsi ad assumere che esso esista e che funzioni in maniera perfetta. È anche interessante notare come le limitazioni descrittive della teoria, la sua incapacità a prevedere correttamente il comportamento in situazioni sperimentali che sempre più abbondanti si sono accumulate negli ultimi anni (cf. Camerer, 2003), siano state spiegate molto più frequentemente attraverso la messa in discussione del concetto di "razionalità classica", piuttosto che di quello di "individualità". Ma iniziano ora a sentirsi, a tale proposito, voci fortemente critiche, come quella per esempio di Herbert Gintis, che recentemente ha fatto notare come:

> «*the most fundamental failure of game theory is its lack of a theory of when and how rational agents share mental constructs. The assumption that humans are rational is an excellent first approximation. But, the Bayesian rational actors favored by contemporary game theory live in a universe of subjectivity and instead of con-*

> *structing a truly social epistemology, game theorists have developed a variety of subterfuges that make it appear that rational agents may enjoy a commonality of belief* […]*, but all are failures. Humans have a social epistemology, meaning that we have reasoning processes that afford us forms of knowledge and understanding, especially the understanding and sharing of the content of other minds, that are unavailable to merely "rational" creatures. This social epistemology characterizes our species. The bounds of reason are thus not the irrational, but the social*» (Gintis, 2009, p. XVI).

Questa posizione e altre simili (cf. Pelligra, 2011c) mettono in luce l'esigenza di studiare le interazioni strategiche, e le relazioni interpersonali che all'interno di esse si sviluppano, attraverso un'analisi più attenta della categoria di intersoggettività, attraverso cioè l'indagine approfondita dei processi tramite i quali l'interazione tra "persone" produce uno spazio intersoggettivo condiviso, fatto di reciproche percezioni, ascrizioni e interpretazioni degli stati mentali altrui. La direzione verso la quale si tende ad andare, quindi, è quella dell'abbattimento epistemologico del limite individuale e dell'ingresso nel territorio dello spazio intersoggettivo condiviso.
In questo saggio vorremmo discutere di questo e di altri limiti, confini e barriere, che una scienza che aspira alla comprensione genuina delle relazioni interpersonali non può non proporsi di travalicare. Lo faremo attraverso il ricorso ad alcuni contributi delle neuroscienze sociali e dell'economia sperimentale, mettendo in luce come concetti eminentemente relazionali, quali quello di "intenzionalità" e di "empatia", possano aiutarci a comprendere meglio alcune tra le numerose sfumature del comportamento sociale e strategico.

2. L'io e l'altro della teoria dei giochi

In questa sezione descriveremo brevemente lo sfondo teorico su cui si inseriscono i concetti di "agente" e di "interazione" cui tradizionalmente fa riferimento la teoria dei giochi classica. Questo è importante per poter comprendere meglio il modo in cui la teoria descrive e modellizza l'idea di "alterità". Il punto di partenza di questa descrizione è naturalmente la teoria di Von Neumann e Morgenstern e il concetto di *minimax*, dal quale facilmente si evince l'idea di individuo razionale che tale teoria descrive. Nella prospettiva di Von Neumann e Morgenstern, un dato corso d'azione è definito razionale se porta a minimizzare (*min*) le perdite massime (*max*) che un giocatore può ottenere, "indipendentemente" da quello che gli altri giocatori decideranno di fare. Questa condizione di indipendenza, che sembra strano trovare proprio nel cuore di una teoria dell'interdipendenza strategica, deriva dal tentativo, che accomunava i due autori, di eliminare ogni riferimento di natura psicologica dalle assunzioni della loro teoria, e in particolare, come bene nota Nicola Giocoli, di liberare i giocatori «dalla necessità di formarsi aspettative circa le azioni e i pensieri dei rivali» (2003, p. 282). Questo aspetto è stato fortemente criticato già negli anni Sessanta da un altro padre fondatore della teoria dei giochi, Thomas Schelling, il quale, descrivendo la teoria di Von Neumann e Morgenstern, metteva in luce come, nella loro prospettiva, un giocatore

> «non ha bisogno di comunicare con il suo avversario, non deve neanche sapere chi sia il suo avversario, ma neanche se un avversario realmente esista. […] Con il criterio del minimax – continua Schelling – un gioco è ridotto ad una situazione completamente unilaterale» (1960, p. 105).

Una critica simile viene portata al concetto di strategia mista, che, sempre secondo Schelling, non è altro che: «un mezzo per distruggere deliberatamente ogni possibilità di comunicazione, specialmente di comunicazione delle *intenzioni*» (*ibid*., corsivo aggiunto).

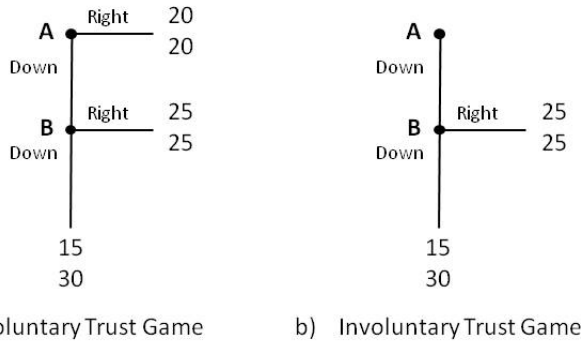
Il progetto di John Nash (1950, 1996), che si pone per molti versi in alternativa a quello di Von Neumann e Morgenstern, si fonda su un concetto di soluzione, l'equilibrio di Nash appunto, decisamente differente dal *minimax*, anche se ancora chiuso al recepimento di una vera "alterità". Per Nash, infatti, un comportamento è razionale quando l'azione di un soggetto è una risposta ottimale ad ogni possibile risposta ottimale degli altri giocatori. Per poter pervenire alla scelta di tale comportamento, un giocatore deve preventivamente farsi un'idea circa quello che sarà il comportamento ottimale degli altri giocatori; deve quindi farsi un'idea circa i desideri e gli obiettivi che gli altri si prefiggono di raggiungere. La complessità di questo processo viene risolta da Nash attraverso un'operazione fortemente riduzionistica, in virtù della quale si assume che le intenzioni degli altri giocatori si limitino al raggiungimento della massima utilità ottenibile. Per rendersi conto delle restrizioni imposte da questa operazione, basti pensare che una delle sue implicazioni è che le credenze che due agenti possono formarsi circa il comportamento di un terzo agente devono necessariamente essere uguali. L'essenza della teoria non-cooperativa dei giochi che Nash sviluppa è, nelle sue parole, l'assunzione secondo cui «ogni partecipante agisce indipendentemente senza nessuna forma di collaborazione o comunicazione con nessuno degli altri giocatori» (1996, p. 22). Queste caratteristiche hanno fatto parlare del solipsismo della prospettiva di Nash, che qualcuno ha anche messo in relazione ai suoi problemi di schizofrenia (cf. Mirowski, 2002). Anche gli ulteriori sviluppi che si sono avuti in questa linea, soprattutto la teoria dei giochi bayesiana di John Harsanyi, non hanno prodotto cambiamenti rilevanti circa l'analisi degli aspetti prettamente e genuinamente relazionali del comportamento strategico (cf. Pelligra, 2011c).

Questa breve e necessariamente superficiale carrellata dovrebbe nondimeno essere sufficiente a far intuire come la visione delle relazioni sociali che si è via via affermata nell'ambito della teoria dei giochi classica sia estremamente semplicistica e basata su assunzioni piuttosto problematiche. Questo è particolarmente vero con riferimento al fatto che, in questo quadro, le intenzioni degli altri giocatori possono essere inferite esclusivamente con riferimento alle loro preferenze circa le conseguenze delle azioni. Tale implicazione determina, come vedremo di seguito, una serie di forti limitazioni nella capacità che la teoria ha di prevedere e descrivere il comportamento reale dei soggetti, come innumerevoli esperimenti hanno ormai definitivamente dimostrato.

3. Intenzioni e conseguenze: alcuni risultati sperimentali

Una delle aree di indagine dove più marcatamente tali limitazioni sono emerse è quella relativa ai comportamenti cooperativi, i quali difficilmente possono essere compresi facendo esclusivo riferimento ai concetti della teoria standard. Tale ambito di analisi, dunque, rappresenta un interessante terreno nel quale è possibile discutere quelle limitazioni della teoria classica che non sono tanto imputabili a un deficit di razionalità, quanto piuttosto ad uno di socialità. In questo senso l'introduzione di categorie marcatamente relazionali, come quelle di fiducia e di reciprocità, ha portato a notevoli progressi verso la creazione di una teoria più adeguata descrittivamente, e maggiormente capace di produrre una visione di agente genuinamente "sociale". Nel seguito considereremo solo alcuni esempi di situazioni "anomale" da un punto di vista standard, che hanno sollecitato interessanti sviluppi sia teorici che empirici.

Il primo esempio cui mi riferisco è un esperimento di McCabe, Smith and Rigdon (2003), nel quale gli autori prendono in considerazione il comportamento dei soggetti in due versioni del *trust-game* leggermente differenti tra di loro. Nella prima versione, il "*trust-game* volontario" (fig. 1a), il giocatore A può ottenere $20 per sé e per il giocatore B scegliendo "destra", oppure può scegliere "giù" e trasferire la possibilità di azione al giocatore B. In questo caso il giocatore B potrà scegliere a sua volta "destra", determinando un guadagno di $25 per entrambi i giocatori, oppure "giù", nel qual caso egli vincerà $30 e farà vincere $15 al giocatore A. La teoria classica prevede che il giocatore A, attraverso un ragionamento retrospettivo (*backward induction*), anticipi la mossa opportunistica di B ("giù"), e per evitare la perdita ad essa connessa preferirà giocare "destra" al primo nodo e far terminare immediatamente il gioco. La versione "involontaria" del *trust-game* (fig. 1b) è simile in tutto e per tutto alla versione "volontaria", tranne per il fatto che, in questo caso, il giocatore A non ha la possibilità di far terminare il gioco al primo nodo. Anche in questo caso la teoria prevede che il giocatore B, per massimizzare il suo guadagno, sceglierà la mossa "giù". Se consideriamo alcune teorie che complicano la struttura motivazionale degli agenti, per esempio inserendo elementi di altruismo (Margolis, 1982) o di equità (Fehr e Schmidt, 1999), le previsioni della teoria cambiano radicalmente. Sia l'altruismo che l'equità porterebbero, infatti, una certa percentuale di giocatori B a scegliere l'opzione cooperativa "destra".



a) Voluntary Trust Game b) Involuntary Trust Game

Figura 1. "*Trust-game* volontario" (a) e "*Trust-game* involontario" (b)

Anticipando questa mossa, poi, un certo numero di giocatori A deciderebbe allora di fidarsi giocando "giù". La questione importante da notare è che sia altruismo che equità prevedono approssimativamente la stessa percentuale di mosse cooperative in entrambi i giochi. Questo perché dal punto di vista dei giocatori B, le conseguenze delle loro scelte nei due giochi sono identiche.

Il risultato dell'esperimento di McCabe e colleghi, invece, mostra chiaramente come la percentuale di mosse cooperative risulti molto maggiore nel *trust-game* "volontario" (94.11%), piuttosto che in quello "involontario" (40.74%). Questo risultato implica che né la versione standard della teoria dei giochi, né i modelli basati sull'idea di altruismo o di equità, che prevedevano percentuali approssimativamente uguali di cooperazione, sono in grado di cogliere appieno le motivazioni che spingono i giocatori reali a fare le loro scelte in situazioni come quelle descritte più sopra. Ciò che questi modelli non riescono a cogliere, in particolare, è, a detta degli autori dello studio, il ruolo cruciale che gioca nell'ambito delle relazioni interpersonali, la capacità di ascrivere intenzioni ai soggetti con i quali interagiamo. Nonostante, infatti, nei due *trust-games* le conseguenze delle azioni del giocatore B siano identiche, nel *trust-game* volontario appare evidente che il giocatore B acquista la possibilità di fare la sua scelta come conseguenza di una scelta libera del giocatore A. Il giocatore B, avendo visto che il giocatore A avrebbe anche potuto scegliere di terminare il gioco senza consentirgli la scelta, è in grado di ascrivere all'azione di A delle intenzioni cooperative. Sono queste intenzioni che il giocatore B riesce a "leggere" e che lo spingono alla cooperazione, con maggiore frequenza nel gioco "volontario" piuttosto che in quello "involontario", nel quale invece, il processo di lettura delle intenzioni è precluso dall'assenza di opzioni alternative per il giocatore A.

Il secondo esempio, sul quale volevamo portare l'attenzione del lettore, è tratto da uno studio di Falk, Fehr e Fischbacher (2003). Nel loro esperimento, gli autori usano quattro differenti versioni del *mini ultimatum game*. In questo gioco il giocatore A fa un'offerta su come dividere una certa somma di denaro al giocatore B, il quale può accettare o rifiutare l'offerta. In caso di accettazione, la proposta viene implementata e i giocatori vincono le somme proposte; in caso di rifiuto, nessuno dei giocatori vince niente. I quattro giochi usati nell'esperimento hanno tutti un'opzione in comune che prevede una divisione (8,2), 8 euro per il giocatore A e 2 per il giocatore B; mentre l'altra opzione cambia di volta in volta, assumendo i seguenti valori: (5,5), (2,8), (8,2) e (10,0) (cf. figura 2a,b,c,d).

Le previsioni della teoria classica suggeriscono che l'offerta (8,2) verrà accettata con la stessa frequenza in tutti e quattro i giochi indifferentemente. Un giocatore B autointeressato, infatti, preferirà sempre una vincita di 2 piuttosto che 0, ciò che otterrebbe nel caso di un rifiuto. I modelli di altruismo e di equità, analogamente a quanto succedeva per i giochi dell'esperimento precedente, prevedono invece che qualche giocatore B rifiuterà l'offerta di 2, ma anche in questo caso la percentuale di rifiuti sarà approssimativamente uguale in tutti i giochi.

I risultati dell'esperimento di Falk e colleghi, invece, mostrano innanzitutto un numero di rifiuti positivo in tutti i giochi; ma, cosa ancor più sorprendente, questi rifiuti diminuiscono sistematicamente quando si passa dal gioco *a* al *b* al *c* fino al *d*. La percentuale di chi rifiuta la stessa offerta (8,2) passa in questi giochi dal 44.4% al 26.7%, al 18%, fino all'8.9%.

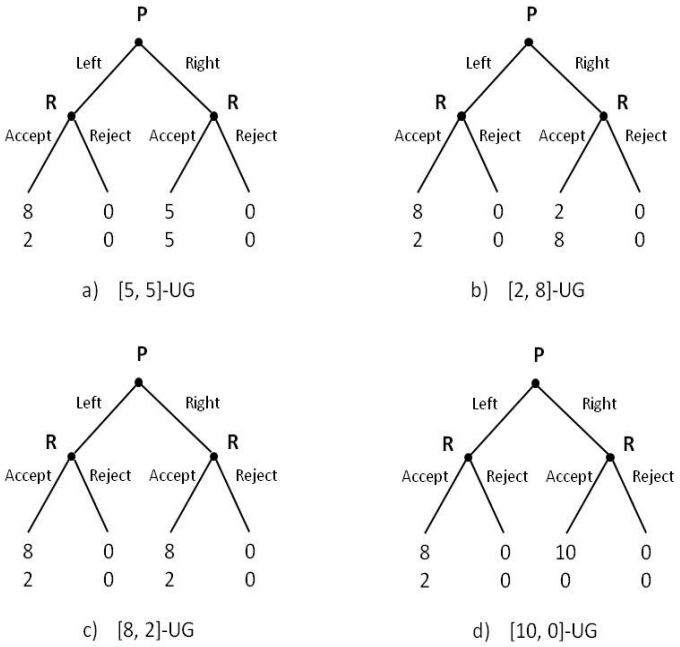


Figura 2. "Le quattro varianti del *mini-ultimatum game*"

Gli autori dello studio interpretano questo risultato come un'indicazione del fatto che i soggetti non sono puramente consequenzialisti: non considerano cioè la preferibilità di un'azione solo in base alle sue conseguenze, ma piuttosto valutano un'azione anche in relazione alle intenzioni che essa incorpora. In questo senso, allora, offrire 2 quando si può anche offrire 5, come nel gioco *a*, è molto diverso dall'offrire 2 quando l'alternativa è 0, come nel gioco *d*. La prima azione è vista come scorretta; mentre la seconda, invece, come gentile. I risultati dell'esperimento, dunque, mettono in luce come le intenzioni, oltre che le conseguenze, siano fondamentali sia nel nostro processo decisionale che nel processo di valutazione delle azioni altrui. Le intenzioni possono essere associate ad un'azione solamente tenendo conto di ciò che l'agente avrebbe potuto fare e non ha fatto, attraverso un processo proiettivo (*forward looking*). Tale processo, però, non rientra nelle possibilità né della teoria classica e neanche delle teoria basate sull'altruismo o l'equità, perché anche queste sono di natura prettamente consequenzialista. Ciò spiega la difficoltà di questo insieme di teorie a dar conto dell'evidenza sperimentale che abbiamo brevemente discusso. La tesi che sosteniamo è che tali limitazioni derivino dalla mancanza di una teoria che riesca a formalizzare il processo di ragionamento che ci consente, per citare Herbert Gintis, di «comprendere e condividere il contenuto delle altre menti» (2009, p. XVI). In altre parole, alla teoria dei giochi manca quella che gli psicologi e i neuroscienziati definiscono "teoria della mente" (TOM).

4. Il limite permeabile I: pensare i pensieri altrui

Attribuire intenzioni significa comprendere le finalità delle azioni altrui. Per poter comprendere tali finalità, dobbiamo essere capaci di attribuire stati mentali a soggetti altri da noi; dobbiamo, in altre parole, essere in grado di pensare i pensieri altrui. Questo è ciò che fa la TOM, un'abilità tipica dei primati superiori ed in particolare degli esseri umani (Rizzolatti *et al.*, 2001), sviluppata probabilmente come conseguenza delle nostre abilità linguistiche (Tommasello, 2000). Tale abilità appare piuttosto precocemente nelle fasi dello sviluppo cognitivo. Già all'età di quattro anni, infatti, i bambini sono in grado di inferire ciò che gli altri sanno, pensano o credono, sulla base di ciò che essi fanno (McCabe *et al.*, 2000:4404). Sui dettagli del funzionamento della teoria della mente si è sviluppato negli ultimi decenni un dibattito serrato e profondo, che ha portato al consolidarsi di due posizioni principali: quella di coloro che considerano la TOM una teoria, appunto, che funziona attraverso l'applicazione di leggi teoriche, e quella di coloro che credono che la nostra comprensione delle menti altrui non sia una questione teorica e cognitiva, ma piuttosto automatica, motoria e simulativa. I primi si definiscono "teorici della teoria", gli altri "teorici della simulazione". I primi ritengono che il processo di lettura della mente avvenga attraverso l'applicazione delle leggi causali della *folk psychology* (Carruthers e Smith, 1996), attraverso la quale diventa possibile associare gli stati mentali non osservabili alle azioni osservabili dei soggetti, per interpretare e prevedere il loro comportamento. I "teorici della simulazione" (Davis e Stone, 1995), invece, ipotizzano che il processo attraverso il quale noi riusciamo ad ascrivere stati mentali agli altri soggetti sia di natura "mimetica". Procediamo, in questo senso, a replicare con i nostri neuroni le configurazioni neuronali che determinano gli stati mentali degli altri soggetti; e in questo modo, attraverso la simulazione appunto, riusciamo a comprendere il contenuto della loro esperienza. Nella versione più spinta, che considera come primario il ruolo della corteccia motoria rispetto a quella prefrontale, questo processo di simulazione è una vera e propria riproduzione neuronale degli stati fisici che contraddistinguono il soggetto con cui stiamo interagendo. I circuiti neuronali coinvolti in questo processo, i cosiddetti neuroni specchio, hanno infatti la capacità di attivarsi sia quando siamo noi in prima persona a compiere una data azione, sia quando osserviamo questa azione in terza persona, compiuta da un altro soggetto (cf. Gallese e Goldman, 1998). La differenza principale, che emerge tra le teorie della teoria e le teorie simulative, è che mentre le prime considerano il processo di *mind-reading* come un fatto oggettivo, distaccato e a contenuto cognitivo, le seconde lo considerano, invece, un fatto legato all'effettiva replicazione delle stesse attività mentali in corso nell'organismo osservato; un processo che avviene in maniera automatica. La comprensione dell'altro assume dunque, in questa seconda prospettiva, la forma di una "intuizione esperienziale" o di una "simulazione incarnata". Nelle parole di Vittorio Gallese:

> «percepire un'azione e comprenderne il significato, equivale a simularla internamente. Ciò consente all'osservatore di utilizzare le proprie risorse per *penetrare il mondo dell'altro* mediante un processo di modellizzazione che ha i connotati di un meccanismo non con-

> scio, automatico, e pre-linguistico di simulazione motoria» (2006, p. 304).

Tale processo sta alla base degli elementi di costruzione della realtà intersoggettiva che si produce non solamente attraverso meccanismi cognitivi, ma anche grazie al ruolo di meccanismo di condivisione emotiva. In questo senso, un ruolo fondamentale è giocato dalla capacità che gli esseri umani posseggono di ascrivere, comprendere e condividere stati emotivi, in una parola di empatizzare.

5. Il limite permeabile II: condividere le emozioni

La capacità di comprendere gli stati mentali dei soggetti con i quali interagiamo sta anche alla base della nostra abilità di empatizzare, vale a dire di ascrivere e condividere gli stati emotivi degli altri soggetti. Come abbiamo detto in apertura, il comportamento strategico si basa fondamentalmente sull'assunzione secondo cui gli agenti sono capaci di prevedere le azioni degli altri interagenti. Questa abilità non si fonda solamente sui meccanismi cognitivi della teoria della mente (mentalizzare), ma anche su processi che attengono alla dimensione emotiva (empatizzare). Se accettiamo infatti che le azioni umane siano, almeno parzialmente, motivate dal desiderio di provare o di evitare certe classi di emozioni, la possibilità di anticipare l'insorgenza e di condividere tali emozioni rappresenta un fattore cruciale di questo meccanismo predittivo più generale. L'empatia, o *perspective-taking*, viene generalmente definita come la nostra abilità a comprendere i sentimenti (*feelings*) degli altri (Preston e de Waal, 2002; Gallese, 2003). Una definizione più specifica è fornita da de Vignemont e Singer (2006), i quali sostengono che, quando empatizziamo, accade: (a) di provare una stato emotivo, (b) che è isomorfo a quello di un altro soggetto, (c) che è stato indotto dall'osservazione o dall'immaginazione dello stato emotivo di tale soggetto, e (d) sappiamo che lo stato emotivo dell'altro soggetto è la fonte del nostro stato emotivo. L'empatia è differente, quindi, sia dalla mentalizzazione, perché produce uno stato emotivo non presente nella mera mentalizzazione, sia dal semplice contagio emotivo, in quanto nell'esperienza empatica noi proviamo ciò che gli altri provano, ma siamo allo stesso tempo in grado di attribuire queste esperienze agli altri e non a se stessi. Preston e de Waal propongono un modello di empatia che spiega in che modo avvenga la comprensione di ciò che qualcun altro prova quando sperimenta semplici emozioni come rabbia, paura, tristezza, gioia o dolore, o anche emozioni più strutturate come disappunto, colpa, orgoglio o vergogna. I due autori suggeriscono che la mera osservazione o immaginazione dello stato emotivo di un altro soggetto attivi automaticamente una rappresentazione di questo stato nell'osservatore. Ulteriori studi (Singer *et al*., 2004) mostrano, attraverso l'uso di strumenti di *imaging* funzionale (fMRI), come tale processo sia effettivamente inconscio e automatico. Gli stessi circuiti neuronali che vengono attivati da un'esperienza diretta di dolore, si attivano quando osserviamo o immaginiamo un'altra persona provare un simile dolore.
La capacità di empatia è fondamentale in ambito strategico, perché ci consente di anticipare le reazioni emotive degli altri rispetto alle nostre potenziali azioni, e quindi di considerare anche le possibili emozioni degli altri come conseguenze

positive o negative delle stesse. Le emozioni, quindi, esercitano una forte influenza, attraverso il processo di empatia, sulle nostre scelte. Possiamo, infatti, voler suscitare emozioni positive negli altri ed evitare di scatenare emozioni negative, anche perché queste, attraverso il processo di rispecchiamento, avrebbero un effetto indiretto su noi stessi.

6. La teoria dei giochi psicologici

Come abbiamo detto, la teoria dei giochi classica non consente la descrizione di processi *forward looking*, che sarebbero necessari per incorporare il processo di attribuzione delle intenzioni che chiamiamo TOM. In questo senso, però, una serie di proposte teoriche promettenti sono quelle che si inseriscono nell'ambito della cosiddetta "teoria dei giochi psicologici" (TGP). La TGP differisce rispetto alla teoria classica, principalmente perché è in grado di considerare l'interazione tra soggetti ad un livello più profondo. Nel processo decisionale, infatti, non si prendono in considerazione solamente le azioni da compiere e le credenze circa le azioni che compiranno gli altri, come nella teoria dei giochi classica, ma anche le credenze di ordine superiore al primo, vale a dire ciò che ogni giocatore crede che gli altri si aspettino da lui, ciò che questi credono che egli creda che i primi si aspettino da lui, e così via (Geneakoplos, Pearce e Stacchetti,1989; Battigalli e Dufwenberg, 2009). Questo espediente epistemico consente di descrivere e formalizzare tutte quelle ragioni per l'azione che sono prettamente emotive e relazionali. Una ben definita classe di emozioni, infatti, hanno radice relazionale: sono quelle che dipendono dalle attese degli altri circa il nostro comportamento e dalle credenze individuali rispetto a tali aspettative. La gioia per essere riusciti a sorprendere piacevolmente un amico deriva dalla credenza che l'amico si aspettava un comportamento differente rispetto a quello che abbiamo messo in essere. Allo stesso modo, ci sentiamo in colpa quando sappiamo che qualcun altro conta su di noi, e noi coscientemente tradiamo tali aspettative. Emozioni come l'orgoglio, il risentimento, la colpa e la gratitudine, hanno tutte la stessa forza motivazionale e la stessa natura epistemica, che può essere descritta attraverso la TGP. Mentre la teoria dei giochi classica e i modelli basati sull'altruismo e l'equità si basano su una visione ipersemplificata dell'intenzionalità, principalmente a causa dell'assunzione di consequenzialismo, la TGP e i modelli che sfruttano la sua struttura epistemica permettono di descrivere agenti che sono capaci di ascrivere intenzioni alle azioni osservate o immaginate degli altri giocatori, e di formalizzare questo genere di motivazioni relazionali. Il processo che viene descritto assomiglia in maniera rilevante a quello ipotizzato dalla teoria della simulazione, anche se è importante notare che queste considerano un processo che è inconscio, automatico e non-cognitivo, mentre la TGP lo considera volizionale e deliberativo. Questa differenza, comunque, non impedisce che la TGP possa rappresentare, con le dovute estensioni, un *framework* teorico adatto alla descrizione dei processi mentali implicati nei processi di empatia e mentalizzazione.

7. Intenzioni, empatia e comportamento pro-sociale

In questa sezione riportiamo alcuni risultati sperimentali che, prendendo come punto di partenza alcuni modelli di TGP, mostrano "direttamente" il ruolo della mentalizzazione e dell'empatia nell'ambito dei comportamenti pro-sociali.

Il primo studio analizza direttamente il ruolo della capacità di attribuzione di intenzioni nello spiegare i comportamenti volti a punire scelte inique. Pelligra *et al*. (2010) hanno replicato l'esperimento di Falk e colleghi (2003), considerando esplicitamente la capacità di mentalizzazione come variabile sperimentale. Per far questo hanno considerato due gruppi distinti di soggetti, posti di fronte alle quattro varianti del *mini-ultimatum game* rappresentate in figura 2. Il primo gruppo è formato da soggetti a sviluppo tipico (*normal developing subjects*, ND), mentre il secondo gruppo è costituito da soggetti affetti da sindrome di Asperger (ASD). I soggetti Asperger hanno, tra le altre caratteristiche, un forte deficit nella capacità di ascrivere stati mentali ad altri soggetti, oltre che livelli di empatia molto inferiori rispetto alla media. L'ipotesi di lavoro, in questo caso, riguardava la diversità nei tassi di rifiuto delle offerte (8,2) da parte dei soggetti appartenenti ai due gruppi. Se lo schema osservato da Falk e colleghi, effettivamente, è dovuto al ruolo cruciale delle intenzioni, si sarebbe dovuto osservare lo stesso schema di rifiuto trovato nello studio originale da parte dei soggetti con alte capacità di mentalizzazione (*high-TOM*), ma uno schema di rifiuto costante da parte dei soggetti provvisti di basso livello di mentalizzazione (*low-TOM*).
I risultati dell'esperimento sono riassunti nella figura 3. I dati dell'esperimento mostrano che, nonostante entrambi i membri dei due gruppi siano tutti sensibili a considerazioni di equità distributiva, i membri del primo gruppo evidenziano delle differenze statisticamente significative nei tassi di rifiuto, così come in Falk *et al*.

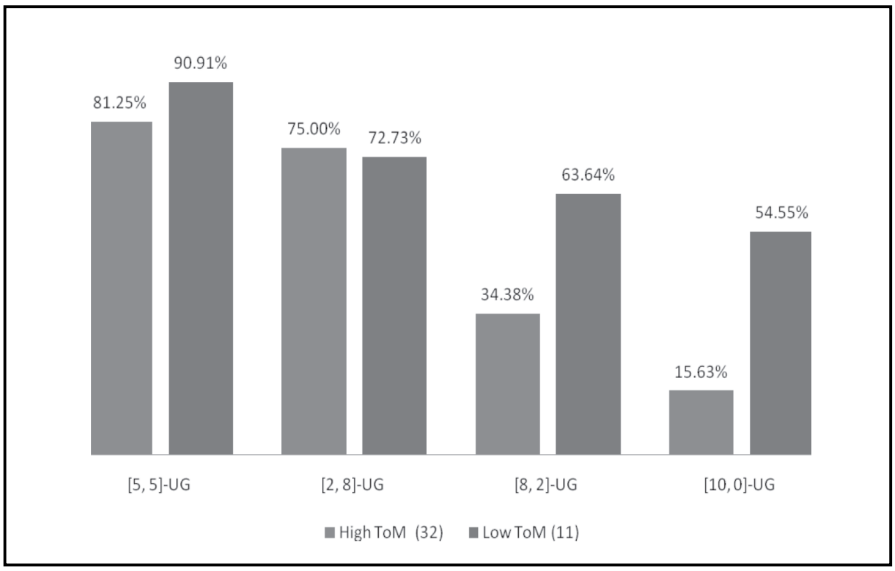


Figura 3. "Tassi di rifiuto degli UG da parte dei soggetti *High-TOM* e *Low-TOM*"
(fonte: Pelligra *et al.* 2010)

Mentre i membri del secondo gruppo segnalano un comportamento molto più omogeneo, con differenze non significative nei tassi di rifiuto. Alla luce di tali risultati è possibile concludere che, da un'analisi "diretta" del ruolo delle abilità di mentalizzazione nei comportamenti strategici, emerge un coinvolgimento rilevante nella spiegazione del comportamento degli agenti, in particolare in risposta alla percezione di intenzioni negative.

Un altro risultato interessante, per quanto riguarda il rapporto tra comportamento pro-sociale e capacità di empatizzazione, è quello che emerge in Becchetti e Pelligra (2011). In questo studio vengono analizzate le scelte dei soggetti in una forma modificata del *dictator-game*, nella quale il soggetto ricopre il ruolo del *dictator*, mentre il ruolo del *recipient* è assegnato ad una serie di nove associazioni di volontariato. Il soggetto ha una dotazione monetaria pari a 10 euro e può selezionare un'associazione di volontariato d'importanza nazionale alla quale donare parte della sua dotazione monetaria. I risultati mettono in luce una disponibilità elevata alla donazione, ed in particolare il fatto che questa aumenti congiuntamente con alcune variabili quali il numero dei componenti il nucleo familiare (1 euro in più, in media, per ogni familiare), il numero di amici nel profilo di Facebook (il 4 percento in più per ogni amico) e, più importante per il tema di questo saggio, il quoziente di empatia. Questo comportamento potrebbe far venire in mente un modello di preferenze altruistiche, nelle quali il benessere degli altri entra proporzionalmente al livello di empatia; alla capacità individuale, cioè, di percepirlo e di condividerlo. Tale spiegazione ci pare però semplicistica, in quanto non tiene conto del ruolo delle intenzioni. Un modello generale di azione pro-sociale, quindi, deve essere più complesso di quello ipotizzato nei modelli distribuzionali (altruismo ed equità), e capace cioè di incorporare il processo di ascrizione delle intenzioni.

In questo senso, uno dei modelli più interessanti che sfrutta la TGP per studiare il comportamento pro-sociale è quello sviluppato da Battigalli e Dufwenberg (2007), che si fonda sul concetto di avversione al senso di colpa (*guilt-aversion*). In questo modello, in un gioco come il *trust-game* volontario descritto più sopra, il giocatore B che osserva il giocatore A scegliere "giù", cioè di fidarsi di lui, sviluppa necessariamente una credenza circa il giocatore A, che potrà scegliere "giù", se razionale, solo nel caso in cui egli si aspetti da B una risposa cooperativa, che produrrà un guadagno mutuo. Solo in questo caso, infatti, ha senso per A scegliere "giù". Questo, B, lo sa, e tale consapevolezza fa scattare il senso di colpa nel caso in cui egli, conoscendo le aspettative di A, decida comunque di frustrarle. L'assunzione di avversione al senso di colpa suggerisce che tale colpa costituisca un costo psicologico per chi la prova, che spingerà, quindi, il soggetto a cercare di evitarla. Se la sensibilità individuale al senso di colpa è sufficientemente elevata, potrà essere sufficiente a indurre il giocatore B a rinunciare al vantaggio materiale che otterrebbe tradendo la fiducia di A, pur di evitare il senso di colpa associato a tale scelta. Maggiore sarà la sensibilità al senso di colpa, quindi, maggiore dovrà essere la probabilità di una scelta affidabile.

In un esperimento progettato per testare proprio tale ipotesi (Pelligra, 2011a), si assume che l'empatia sia il meccanismo che modula la capacità di provare senso di

colpa, proprio in virtù della maggiore o minore capacità di condividere ed anticipare le emozioni altrui, per esempio il disappunto per essere stato tradito. Studiando sperimentalmente differenti versioni del *trust-game*, troviamo che la probabilità di comportamento affidabile appare non essere correlata con il livello di beneficio materiale che l'azione fiduciosa può accordare a B, nel caso di una sua scelta affidabile, mentre risulta significativamente correlata con la misura dell'empatia individuale rilevata attraverso il *Cambridge Empathy Quotient* (Baron-Cohen *et al.,* 2004). Maggiore è la capacità individuale di anticipare e condividere il disappunto dell'altro giocatore, maggiore sarà la sensibilità al senso di colpa derivante da un eventuale tradimento delle aspettative, e maggiore sarà, quindi, il costo psicologico associato a tale scelta. Questo, *ceteris paribus*, porta ad una maggiore disponibilità alla scelta cooperativa.

In un altro studio (Pelligra, 2011b) viene analizzato il comportamento dei soggetti in una forma più complessa di *trust-game*, noto come *investment-game*. In questo gioco, il giocatore A è dotato di una certa somma di denaro. Può decidere quanta parte di tale somma inviare al giocatore B. La somma inviata viene triplicata e il giocatore B, a questo punto, può decidere quanta parte mandarne indietro. L'equilibrio teorico del gioco prevede che, anticipando la scelta autointeressata del giocatore B di non restituire niente, il giocatore A non manderà niente. I risultati dell'esperimento in questione mostrano invece che i giocatori A mandano in media 59.76 della loro dotazione di 100 euro, e che solo 5 su 53 giocatori B decidono di non restituire niente, indipendentemente dalla somma ricevuta da A (figura 4).
Se mettiamo in relazione tali scelte con il livello individuale di empatia, non emerge nessuna correlazione significativa, né dal punto di vista dei giocatori A né da quello dei giocatori B. L'effetto dell'empatia, invece, emerge chiaramente se analizziamo non tanto la somma restituita da ogni giocatore B, quanto piuttosto la variazione della percentuale restituita in relazione alla somma ricevuta (figura 5). Considerando tale variabile, infatti, possiamo distinguere piuttosto chiaramente due tipologie di soggetti: quelli che restituiscono una somma positiva ma in percentuale costante rispetto a quanto ricevuto, e quelli che, invece, restituiscono percentuali crescenti all'aumentare della somma ricevuta. I primi rispondono alla norma di "reciprocità bilanciata", mentre i secondi adottano la norma della "reciprocità condizionale" (Greig e Bohnet, 2008).
Se misuriamo il livello di empatia di coloro che seguono le due norme, notiamo che i reciprocatori "condizionali" hanno un quoziente di empatia medio pari a 47.55, mentre i reciprocatori "bilanciati" totalizzano in media solo 42.15.
Questi risultati sembrano mostrare che, nell'*investment-game*, la scelta di ripagare la fiducia non sia, per molti soggetti, tanto legata ad un complesso calcolo costi-benefici materiali e psicologici, quanto piuttosto al rispetto di un imperativo categorico della forma – "se qualcuno si fida di me, non posso tradire la sua fiducia". La "reciprocità bilanciata" spinge quindi il giocatore B a massimizzare il suo guadagno, sotto il vincolo di non far perdere ricchezza al giocatore A rispetto alla sua situazione di partenza.
L'empatia, però, produce un effetto ulteriore, quello che spinge i giocatori B a fare qualcosa di più rispetto al mero non far perdere denaro a chi si è fidato di loro. I soggetti con alto quoziente di empatia tendono a restituire più di quanto

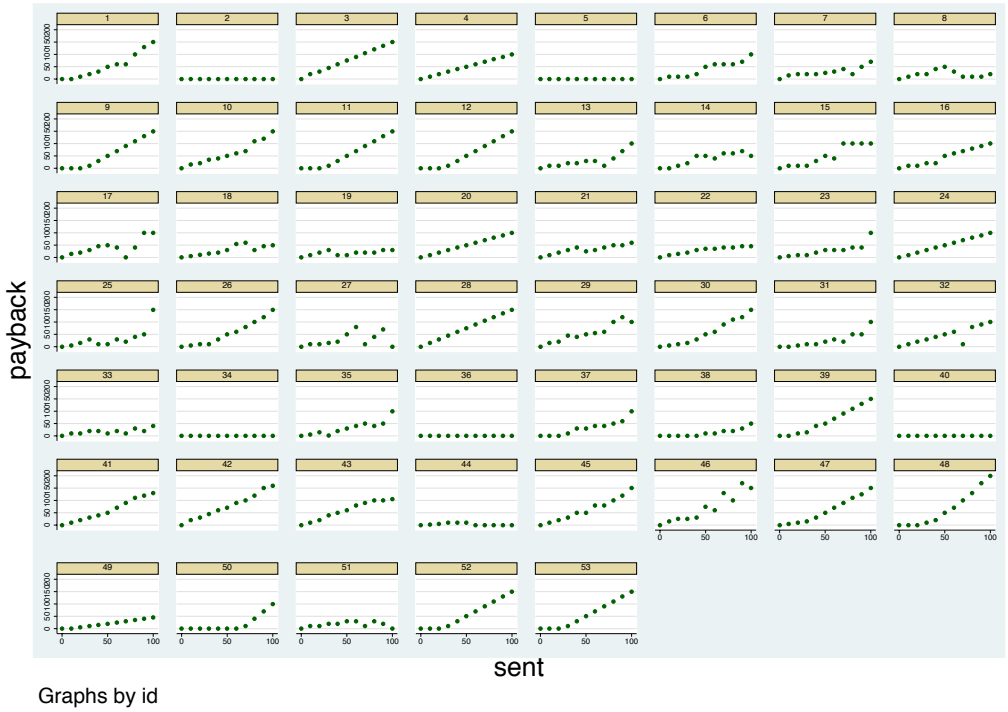


Figura 4. "Investimenti e restituzioni" (fonte: Pelligra 2011b)

sarebbe necessario per lasciare l'altro giocatore in una situazione di non-perdita, e in questo modo ripagano la sua fiducia distribuendo in modo più egualitario i guadagni che si producono nell'interazione. È interessante notare che la "reciprocità bilanciata", e altre forme di norme sociali *no-loss*, tendono a essere prevalenti in quei contesti, in ritardo di sviluppo, basati su sistemi economici nei quali i contratti sono fatti rispettare attraverso un *enforcing* informale fondato proprio su norme di scambio di equivalenti (cf. Platteau, 1997; Thomas and Worrall, 2002). Il limite, legato all'utilizzo di tali norme, riguarda il fatto che il giocatore B si appropria di tutto il surplus generato nell'interazione; e, benché il giocatore A non perda niente, questo per lui può non essere un incentivo sufficiente a spingerlo ad attivare la relazione fiduciaria. Si perde in questo modo l'opportunità di un guadagno per entrambi i giocatori.

8. Conclusione: verso uno spazio intersoggettivo condiviso

Nelle pagine precedenti abbiamo cercato di mettere in luce alcuni dei principali limiti della teoria dei giochi classica, che possono ostacolare la comprensione di dimensioni fondamentali per le relazioni interpersonali. In primo luogo la capacità di concettualizzare gli altri soggetti come agenti dotati di identità, e quindi di intenzioni e finalità eterogenee; "non parenti, ma stranieri", per riprendere la

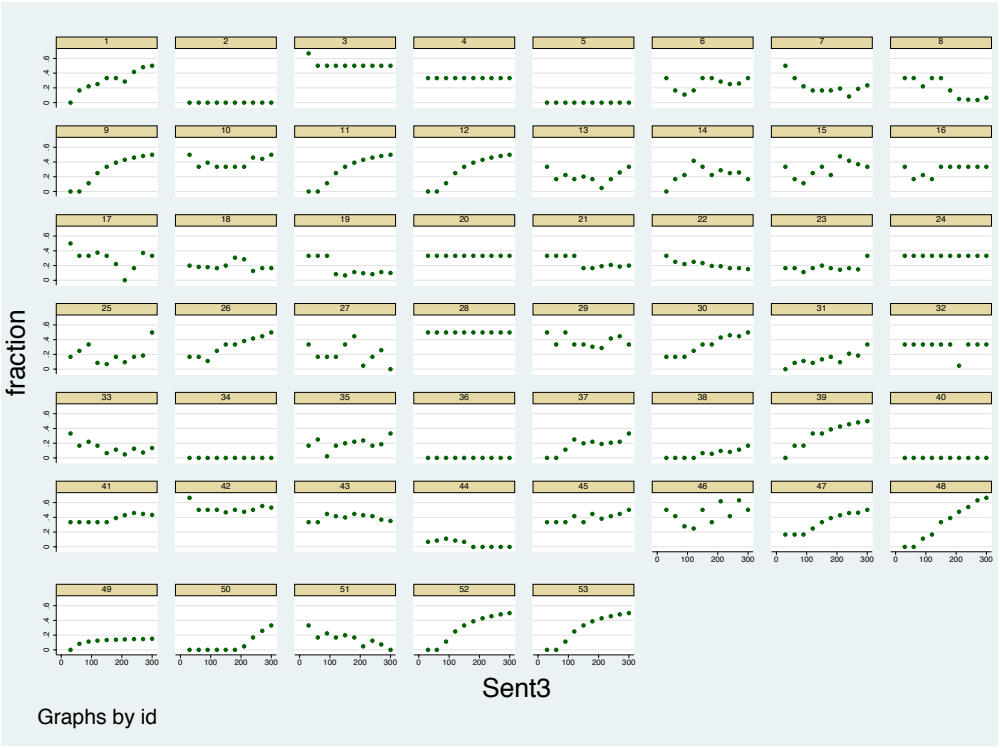


Figura 5. "Investimenti e percentuali di restituzioni" (fonte: Pelligra 2011b)

citazione di apertura di Levinas. E, in secondo luogo, la capacità di attivare una comprensione ed una comunicazione a livello emotivo e non solamente cognitivo. Questi due elementi costituiscono, crediamo, tasselli importanti di quella epistemologia sociale, di quella comprensione del mondo dell'altro, di cui ogni teoria del comportamento strategico, come ci ricorda Gintis (2009), non può fare a meno. Abbiamo analizzato teoricamente il ruolo dei processi di empatia e mentalizzazione, e abbiamo discusso alcuni risultati sperimentali che sottolineano il loro ruolo nelle relazioni strategiche. Abbiamo anche individuato nella TGP un promettente sviluppo che ci pare vada nella direzione, auspicabile, dell'introduzione di elementi maggiormente relazionali nella definizione della struttura di motivazione degli agenti sociali.

Naturalmente questi elementi non sono che primissimi indizi, o suggestioni, che possiamo proporre all'attenzione del lettore. Elementi, però, che, per quanto preliminari, indicano in maniera piuttosto chiara la direzione da intraprendere. La convergenza della teoria dei giochi, della filosofia dell'azione e delle neuroscienze sociali verso una visione di soggetto-relazione, conforta la nostra analisi. Sembra che da più parti, più o meno indipendentemente, stia maturando la consapevolezza che lo studio dei fenomeni interpersonali e sociali non possa più essere basato esclusivamente sull'analisi di soggetti individuali che interagiscono tra loro in modo "freddo". Ciò che emerge, invece, è una visione nella quale l'unità di analisi, sempre più, si sposta dall'io alla diade, dall'individuo alla "relazione", perché si inizia a

comprendere sempre meglio come l'agente "si fa" nella relazione. Un agente che entra in relazione con un altro agente viene modificato da questa relazione, grazie proprio all'entrata in contatto con l'altro. Nella relazione, i soggetti contribuiscono a costruire e a popolare uno spazio intersoggettivo condiviso, nel quale il soggetto non si esaurisce – mantiene quindi la sua individualità – ma muta, cambia, diventa "persona", per usare il linguaggio della filosofia. La mentalizzazione e la capacità di empatizzare sono due dei principali meccanismi attraverso cui tale spazio viene costruito. Il meccanismo dei neuroni specchio, che sottende a tali processi, produce quella che possiamo chiamare una vera e propria "simulazione incarnata", che ci dà accesso al mondo sub-personale dell'altro. In questo modo, la sintonizzazione con le realtà mentali dell'altro non solo ci consente di comprendere l'altro, ma anche, assumendo la sua prospettiva, di abbandonare la nostra posizione egocentrica e di guardare a noi da una prospettiva differente; e di cogliere, in questo modo, aspetti di noi stessi che altrimenti ci sarebbero stati preclusi.

In questo saggio abbiamo incontrato e cercato di oltrepassare differenti "limiti": il limite disciplinare che ci spinge a travalicare le separazioni tra linguaggi e discipline, il limite che si pone tra percezione e oggettività della realtà percepita; e abbiamo cercato di sottolineare quanto questa realtà oggettiva sia costruita proprio dalla percezione della realtà dell'altro. Ed infine il limite della soggettività. L'identità personale non si fonda sulla separatezza degli agenti, ma sulla costruzione di uno spazio intersoggettivo condiviso, che esiste solo in virtù della permeabilità dei confini della nostra propria soggettività.

Bibliografia

S. Baron-Cohen – S. Wheelwright, *The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences*, in «Journal of Autism and Developmental Disorders», 34(2), 2004, pp. 164–175.

P. Battigalli – M. Dufwenberg, *Guilt in Games*, in «American Economic Review, Papers & Proceedings», 97, 2007, pp. 170-76.

P. Battigalli – M. Dufwenberg, *Dynamic psychological games*, in «Journal of Economic Theory», 144, 2009, pp. 1–35.

L. Becchetti – V. Pelligra, *Don't be ashamed to say you didn't get much: redistributive and aggregate effects of information disclosure in donations,* in "AICCON", Working Paper Number 2011_88.

C. Camerer, *Behavioral Game Theory. Experiments in Strategic Interaction*, Princeton University Press, Princeton 2003.

P. Carruthers – P. Smith (eds), *Theories of Theories of Mind*, Basil Blackwell, Oxford 1996.

M. Davis – T. Stone (eds.), *Mental Simulation*, Oxford: Basil Blackwell, Oxford 1995.

F. de Vignemont – T. Singer, *The empathic brain: how, when and why?,* in «Trends in Cognitive Science», 10, 2006, pp. 435-441.

E. Fehr – K.M. Schmidt, *A Theory of Fairness, Competition and Cooperation*, in «Quarterly Journal of Economics», 114, 1999, pp. 817-868.

V. Gallese, *Corpo Vivo, Simulazione Incarnata e Intersoggettività,* in M. Cappuccio (ed.), *Neurofenomenologia*, Bruno Mondadori, Milano 2006.

V. Gallese – A. Goldman, *Mirror neurons and the simulation theory of mind-reading,* in «*Trends in Cognitive Sciences*» 2(12), 1998, pp. 493-501.

J. Geanakoplos – D. Pearce – E. Stacchetti, *Psychological Games and Sequential Rationality*, in «Games and Economic Behavior», 1, 1989, pp. 60–79.

H. Gintis,*The Bounds of Reason, Game Theory and the Unification of the Behavioral Sciences*, Princeton University Press, Princeton 2009.

N. Giocoli, *Modelling Rational Agents*, Cheltenham: Edward Elgar, 2003.
F. Greig – I. Bohnet, *Is There Reciprocity In A Reciprocal-Exchange Economy? Evidence Of Gendered Norms From A Slum In Nairobi, Kenya*, in «Economic Inquiry», 46(1), 2008, pp. 77-83.
K.A. McCabe – M.L. Rigdon – V.L. Smith, *Positive Reciprocity and Intentions in Trust Games,* in «Journal of Economic Behavior and Organization», 52, 2003, pp. 267-275.
K. McCabe – V. Smith – M. Lepore, *Intentionality and "mindreading": Why does the game form matter?*, in «Proceedings of the National Academy of Science», 97(8), 2000, pp. 4404-4409.
H. Margolis, *Selfishness, Altruism, and Rationality. A Theory of Social Choice*, Chicago University Press, Chicago 1982.
P. Mirowsk, *Machine dreams*, Cambridge: Cambridge University Press, Cambridge 2002.
J. Nash, *Non-Cooperative Games*, in «Annals of Mathematics» 54, 1951, pp. 286-295.
J. Nash, *Essays in Game Theory*, Cheltenham: Edward Elgar, 1996.
V. Pelligra, *Reciprocating Kindness. An Experimental Investigation.* Mimeo, Università di Cagliari, 2011a.
V. Pelligra, *Empathy, Guilt-Aversion and Patterns of Reciprocity*, in «Journal of Neuroscience, Psychology and Economics» 4(3), 2011b, pp. 161-173.
V. Pelligra, *Intentions, Trust and Frames: A note on Sociality and the Theory of Games*, in «Review of Social Economy», 2011c, 69(2), pp. 163-188.
V. Pelligra – A. Isoni – R. Fadda – I. Doneddu, *Social Preferences and Perceived Intentions. An experiment with Normally Developing and Autistic Spectrum Disorders Subjects*, in «CRENoS», Working Paper Number 2010_10.
J. Platteau, *Mutual Insurance as an Elusive Concept in Traditional Rural Communities*, in «Journal of Development Studies», 33, 1997, pp. 764-796.
S. Preston – F. de Waal, *Empathy: Its ultimate and proximate bases*, in «Behavioral and Brain Sciences», 25(1), 2002, pp. 1-72.
G. Rizzolati – L. Fogassi – V. Gallese, Neurophysiological *mechanisms underlying the understanding and imitation of action*, in «Nature Reviews of Neuroscience», 2, 2001, pp. 661–70.
T. Schelling, *The Strategy of Conflict*, Cambridge: Boston MA, Harvard University, 1960.
T. Singer – B. Seymour – J.P. O'Doherty – H. Kaube – R.J. Dolan – C.D. Frith, *Empathy for Pain Involves the Affective but not Sensory Components of Pain*, in «Science», 303(5661), 2004, pp. 1157-62.
J.P. Thomas – T. Worrall, *Gift-giving, Quasi-credit and Reciprocity*, in «Rationality and Society», 14, 2002, pp. 308-352.
M. Tommasello, *The Cultural Origins of Human Cognition*, Cambridge, MA: Harvard University Press, Cambridge 2000.

VITTORIO PELLIGRA
Ricercatore di Economia Politica presso il Dipartimento di Economia dell'Università di Cagliari e professore incaricato presso l'Istituto Universitario Sophia.
pelligra@unica.it