

La mente che minaccia: razionalità, patrimonio emotivo, integrità deliberativa

[Pierpaolo Marrone](#)

[Dipartimento di Filosofia](#)
[Università di Trieste](#)

ABSTRACT

The paper is devoted to the concept of deterrence. After a brief review of the conceptual framework of the so-called paradox of deterrence, it focuses on the two concepts of ‘integrity in deliberation’ and ‘emotional assets’ in order to show that they are intrinsic part of the description of an agent as rational and thus not in contradiction with the intention to retaliate.

1. *Coltelli e bombe*

Se tu sei l’unico, in un gruppo abbastanza vasto in cui i singoli membri siano costretti a continue relazioni reciproche, a possedere un coltello, qualcuno, prima o poi, comincerà a coltivare strani pensieri. Magari immaginerà, ad esempio, che lo potresti usare a scopi aggressivi, per acquisire le risorse di altri membri importanti del gruppo e non soltanto per difendere i tuoi amici. Presto o tardi, insomma, comincerà a chiedersi perché non dovrebbe possederlo anche lui. Ammettiamo che nel tuo gruppo il coltello costituisca la migliore tecnologia aggressivo-difensiva disponibile e non esistano tecnologie di difesa efficienti, perché il gruppo dove vivi è costituito da emofiliaci, per i quali anche una ferita poco profonda potrebbe costituire una possibilità molto concreta di morte. Potresti volere, tu l’unico proprietario, possederne allora un altro e dotarne i tuoi amici più stretti; ma anche un tuo potenziale avversario potrebbe cominciare a pensare la stessa cosa e in effetti adoperarsi lungo la stessa via, magari per acquisire la *leadership* di una porzione importante del vostro gruppo. Arriva il momento in cui non esiste più un unico coltello, ma un arsenale relativamente vasto di armi da taglio diviso, più o meno equamente, fra due gruppi in reciproca competizione. La tentazione di un assalto di sorpresa al gruppo rivale è, in effetti, forte, ma, poiché non esistono tecnologie di difesa dalle armi da taglio apprezzabilmente sicure, nessuno dei due gruppi è sicuro di sterminare o disarmare permanentemente il gruppo avversario, ossia non è affatto sicuro che un attacco a sorpresa sia o risolutivo o addirittura attuabile. Un duello fra i due

gruppi potrebbe avere esiti letali e costituire una minaccia alla sopravvivenza di tutti. Che cosa è razionale fare per assicurare la vostra sopravvivenza nel medio e lungo periodo? La mente che minaccia è una mente prigioniera di intenzioni paradossali?

Un dilemma di questo genere è stato a lungo familiare nel secolo scorso: quando si sono confrontate nello scenario ipotetico del conflitto le tecnologie di distruzione nucleare, i problemi che vi sono legati e le riflessioni che sono sorte non hanno mancato di appassionare esperti di strategia, teorici dei giochi e delle decisioni, consiglieri militari, studiosi di politica – talvolta con responsabilità politiche –, esperti di geopolitica, filosofi morali e della politica. (1) La riflessione strategica sull'uso degli armamenti nucleari ha avuto il suo avvio verso gli anni Cinquanta del secolo trascorso, in conseguenza delle stragi di civili largamente dimostrative di Hiroshima e Nagasaki, (2) e ha prodotto lavori pregevoli nel corso di più di un quarantennio, per affievolirsi successivamente, in conseguenza della dissoluzione dell'impero sovietico, che ha messo fuori gioco – momentaneamente? – uno dei due giocatori di quella partita. Presso il grande pubblico è divenuta nota con il poco rassicurante nome di *equilibrio del terrore* e con l'acronimo, profetico, ma fuorviante, di *MAD (Mutual Assured Destruction)*. (3) Rispetto al racconto dell'esempio iniziale, che conserva una sua utilità esplicativa, vi sono però variabili non di poco conto. La prima sembra essere derivata da un contesto empirico. Questo contesto ha generato quella che secondo molti – sicuramente secondo quell'entità dagli incerti confini e definizioni, ma non dall'incerta esistenza che è l'uomo della strada – rimane un assioma non bisognoso di dimostrazione. Si tratta di questo: poiché la capacità distruttiva delle armi nucleari si estende potenzialmente a tutto il genere umano, è irrazionale usarle, poiché tale uso genererebbe una *escalation* inarrestabile e incontrollabile. Questa certezza assiomatica, tuttavia, non è in disaccordo con un'altra convinzione, ossia che proprio a causa di un immenso potenziale distruttivo non è irrazionale minacciare l'uso degli ordigni nucleari. (4)

In un mondo di minacce asimmetriche, (5) quale sembra essere divenuto il nostro, il potenziale investigativo di quella riflessione sembra improvvisamente fuori moda. Nel nostro nuovo secolo, non esistono più due contendenti, entrambi rappresentati da Stati che si affrontano su un piano di approssimativa eguaglianza di risorse distruttive. Al momento attuale, e forse ancora per molto tempo a venire, (6) solamente uno ha capacità di intervenire in maniera coordinatamente vasta nella quasi totalità delle caselle dello scacchiere strategico (ma non bisogna dimenticare che il contendente dell'altro ieri ha conservato un suo arsenale nucleare di tutto rispetto). In un mondo di conflitti asimmetrici, cioè, non sembra sia possibile minacciare una massiccia ritorsione nucleare in caso di attacco al territorio del proprio Stato o di attacco ad alleati o di minaccia a propri interessi valutati come vitali, perché l'altro

contendente non è insediato in un territorio e non è rappresentato da un'istituzione statale.

Sostanzialmente per questi motivi, elencati in maniera così sintetica, la riflessione strategica originata dal confronto della guerra fredda sembra essere divenuta inattuale. Ma le cose stanno proprio in questo modo? Io penso che riflettere sulle condizioni di razionalità della minaccia e della deterrenza nelle interazioni conflittuali abbia ancora parecchi meriti e utilità, anche perché queste interazioni sono in definitiva relazioni di riconoscimento e parziale cooperazione. Uno di questi è di dirci qualcosa di importante sull'intreccio tra passioni e razionalità e sulla natura di tale intreccio in situazioni estreme.

2. *Il paradosso della deterrenza*

Le condizioni di razionalità di una minaccia, inoltre, possono implicare anche una dimensione etico-politica, se uno degli obiettivi della minaccia è la conservazione della pace fra i contendenti. L'ingiunzione *si vis pacem, para bellum* potrebbe, quindi, avere un contenuto meno paradossale – o nient'affatto paradossale (7) – di quello che appare a prima vista, anche se il legame tra due sostantivi opposti, *pace* e *guerra*, non è generalizzabile in altre ingiunzioni (non sembra esserci nessuna logica neppure paradossale in una ipotetica ingiunzione che dica “se vuoi preservare il tuo peso forma, cibati di torte al cioccolato”). La forma estrema della minaccia nucleare è la distruzione reciproca garantita. Su questa forma estrema si sono concentrate critiche molto forti, che hanno riguardato sia la moralità dell'azione sia la sua razionalità, poiché sembra essere l'esemplificazione, appunto estrema, di un paradosso, così espresso da Kavka: «There are cases in which, although it would be wrong for an agent to perform a certain act in a certain situation, it would nonetheless be right for that agent, knowing this, to perform the intention to perform that act in that situation». (8) Un atto da cui risulti come conseguenza la distruzione dei contendenti appare a molti chiaramente immorale, a maggior ragione se i contendenti non sono due persone, ma le popolazioni di due Stati. Quale *moralità* ci sarebbe nel provocare sofferenze in milioni di innocenti? Inoltre, la stessa razionalità della minaccia avanzata a scopo di deterrenza è dubbia. Ammettiamo che il nostro avversario, verso il quale abbiamo deciso di adottare un comportamento bilanciato di minaccia, ci attacchi per primo, nonostante noi si sia proclamata pubblicamente l'idea di reagire in maniera distruttiva. Una volta che il primo colpo sia stato lanciato l'idea stessa di deterrenza non sembra avere più senso, dal momento che si è dimostrata inefficace. Se fosse stata realmente efficace, infatti, non ci sarebbe stato nessun primo colpo. Quale *razionalità* e quale *utilità* quindi ci sarebbe nel provo-

care sofferenze in milioni di innocenti, quando già molti altri innocenti hanno sofferto? (9)

A differenza dell'ingiunzione che si ricordava sopra, nel concetto stesso di deterrenza sembra perciò annidarsi un genuino paradosso. Potrebbe essere razionale e morale *avvalersene* e irrazionale e immorale *attuarela*. Per questo alcuni hanno ritenuto che la rappresaglia sia tanto inutile quanto irrazionale quanto immorale e non soltanto nella sua forma maggiormente distruttiva – quella forma che, una volta attuata, non lascerebbe spazio per nessuna altra *escalation* possibile –, ma in qualsiasi forma. Anche perché (ma assolutamente *non* “proprio perché”) può condurre i contendenti a intraprendere un'*escalation* inarrestabile, la deterrenza non può essere assunta come opzione praticabile da un agente razionale, ossia non può fare parte dei fini che tale agente *deliberatamente* sceglie. (10)

L'idea che la rappresaglia sia tanto immorale quanto irrazionale e, cioè, che per avere attuazione deve essere eseguita da agenti o irrazionali o immorali o entrambi è diversa da altre critiche correnti, ad esempio, da quelle che sottolineano la difficoltà di comprendere le intenzioni dell'avversario o da quelle che ritengono che il concetto di deterrenza sia la sintesi della semplice descrizione di una situazione altamente instabile, venendo meno al fine per il quale la si è messa in campo, impedire il danno reciproco o la distruzione reciproca minacciandoli. (11)

Minacciare una rappresaglia, perciò, in quale senso è incompatibile con i fini di un agente razionale? Ogni agente ha molti fini e il loro perseguimento può travalicare, e il più delle volte travalica, i confini di una razionalità intesa in senso strumentale – anche se forse ci possono essere rari casi di agenti che fanno della stessa razionalità strumentale un fine (12) –, per cui, se non accogliamo l'idea troppo semplice che la deterrenza sia semplicemente irrazionale, potremmo avere bisogno di una diversa concezione della razionalità della minaccia. Ovviamente, si può osservare che le cose nel mondo reale possono andare in maniera molto diversa, ad esempio perché uno dei due contendenti può mancare in tutto o in parte dei requisiti di razionalità prudentiale o di moralità minimi per soddisfare le condizioni di minaccia efficace. Questa osservazione ci dice però semplicemente che il mondo reale è sempre più pericoloso e più complesso della teoria che lo descrive, ma può non essere un buon motivo per rinunciare a una teoria capace di individuare alcune condizioni plausibili di razionalità dell'azione in situazioni critiche. Forse queste condizioni, le condizioni di rappresaglia, non sono mai presenti nella purezza descritte dai teorici, ma questo non può essere considerato un ostacolo in generale alla modellizzazione. Assumiamo allora semplicemente che vi sia una minaccia effettiva e concreta alla pace e che la minaccia di rappresaglia (*si vis pacem, para bellum*) sia un mezzo efficace di conservarla, correndo rischi accettabili per preservare un fine della più grande importanza morale e assieme razionale dal punto di vista della concezione strumentale dell'agente.

3. *Paradosso performativo?*

Ci sono, *grosso modo*, due modi di affrontare queste questioni. Il primo sostiene la razionalità e la moralità della deterrenza; il secondo la sua immoralità e irrazionalità. La prima modalità ha una matrice utilitarista e consequenzialista; la seconda una matrice anticonsequenzialista. Fra i filosofi, chi ha indagato in maniera più estesa e profonda le implicazioni di queste posizioni e le possibilità di contaminarle è stato G. Kavka. (13) Posizioni e implicazioni si contaminano per dare origine a molteplici paradossi secondo Kavka, l'origine dei quali è individuata nel fatto che per certi versi entrambe le posizioni sembrano giuste. Ad esempio, gli anticonsequenzialisti sembrano avere torto a sostenere che le posizioni di deterrenza debbano necessariamente essere o irrazionali o immorali. (14) Sembrerebbe, infatti, che l'agente sia costretto a formarsi delle intenzioni paradossali, il che significa che esistono casi in cui, sebbene sia sbagliato per un agente agire in una determinata maniera, sarebbe tuttavia giusto per quell'agente formarsi l'intenzione di agire proprio in quel modo. Ossia in determinate circostanze sarebbe razionale, per così dire, munirsi di una certa dose di cecità per le conseguenze morali e/o irrazionali dei propri atti. (15) È dubbio che i primi teorici della deterrenza avrebbero accettato una descrizione di questo genere. (16) Piuttosto per loro, nonostante la mostruosità delle conseguenze, l'agente che avesse messo in atto una rappresaglia avrebbe agito in maniera pienamente razionale, perché non sarebbe solo pienamente razionale minacciare la rappresaglia in determinate circostanze, nonostante il fatto che ciò che si minaccia di fare sia irrazionale, ma anche avere la ferma e manifesta intenzione di attuarla effettivamente.

Per Kavka, invece, l'agente che minaccia la rappresaglia deve in certa maniera divenire irrazionale nelle sue intenzioni, attraverso processi di vario genere che possono anche comportare l'intenzionale corruzione di motivazioni morali e/o razionali. Queste circostanze di corruzione possono spingersi oltre il campo limitato della rappresaglia nucleare e riguardare anche situazioni più quotidiane di minaccia e deterrenza (e questo è il motivo per cui l'argomento possiede un interesse più vasto del *case study* più noto). In altre parole, siamo qui in presenza di paradosso performativo generale, perché, da un lato, la minaccia di attuare una rappresaglia è razionalmente e moralmente desiderabile, ma attuarla effettivamente, dall'altro, comporta la messa in pratica di procedure che è irrazionale e immorale attuare. Una minaccia per essere efficace deve essere credibile, e per essere credibile deve essere praticabile: perciò ogni agente (razionale e/o morale) per formarsi l'intenzione di essere effettivamente in grado di attuare la rappresaglia deve sottoporsi a un processo di corruzione cognitiva e/o morale, o essere già naturalmente corrotto. (17)

Così la questione della razionalità/irrazionalità di atti/intenzioni incrocia quella della moralità/immoralità delle conseguenze/intenzioni condizionali. La mia posizione è che le critiche di irrazionalità alla deterrenza non siano fondate e persuasive. (18) Ritengo che sia possibile mostrare la coerenza fra il formare l'intenzione condizionale di attuare la rappresaglia in circostanze appropriate e il mantenere l'intenzione pur sapendo che tale intenzione è condizionata da conseguenze non desiderabili per l'agente. Nello stesso tempo, se questo tentativo ha successo dovrebbe essere più facile mostrare che un agente che agisce razionalmente è più portato a conservare l'intenzione di applicare la rappresaglia in condizioni opportune.

4. *Agente e opzioni*

Ritorniamo al nucleo dei numerosi paradossi enunciati da Kavka. Per quale motivo dovremmo supporre che un agente razionale non è in grado di formare e mantenere nel corso del tempo una intenzione di attuare la rappresaglia? La struttura del problema è la seguente: a) l'agente è razionale; b) l'agente intende dare corso alla propria intenzione di fare una certa azione, se accadrà un certo evento (ad esempio, un *first strike*); c) risulta chiaro all'agente che se si verifica tale evento sarebbe irrazionale attuare la rappresaglia. Dal momento che l'insieme di queste condizioni è incoerente, allora se ne trae la conclusione che nessun agente razionale può realmente mantenere l'intenzione condizionale di rendere effettiva la sua minaccia, almeno con la consapevolezza e la conoscenza delle sue implicazioni. Questo significa che l'intenzione di mettere in atto la minaccia di rappresaglia, a meno di non agire irrazionalmente, non può essere attuata.

Tuttavia, all'interno del nucleo di questi paradossi si è insediato un altro nucleo, dove dovrebbero essere rappresentate le ragioni per le quali pensiamo che non sia coerente formarsi l'intenzione di dare corso alla minaccia e quella collaterale di sostenere la propria volontà di mantenere questa intenzione. Per comprendere cosa ci possa essere di sbagliato o di descrittivamente insufficiente in quella conclusione è necessario rendere conto delle tensioni effettive che esistono nella struttura del problema. E per fare questo è forse una buona strategia iniziare dal fondamento del problema, ossia: che cosa fa di un agente un agente razionale? Penso si possa fornire una risposta non eccessivamente complicata e in linea con il senso comune. La razionalità di un agente è una disposizione al comportamento, quella disposizione ad agire conformemente alle sue credenze giustificate sorrette dalle migliori evidenze disponibili e a non agire quando queste evidenze non sono disponibili, ossia a non agire quando la situazione presenta gradi elevati di incertezza cognitiva. (19) In altre parole, l'agente agisce razionalmente se sulla base dei suoi desideri e delle sue credenze giustificate sullo stato del mondo è disposto a intraprendere un corso

d'azione che è cognitivamente coerente con i suoi desideri. Questa disposizione, inoltre, non deve essere episodica, ossia non deve essere né presente solo condizionalmente a uno specifico corso d'azione futura né deve essere presente per caso, ma deve essere costante. (20)

In qualche modo, tutto ciò che qui viene descritto si riferisce a un agente ideale. Tuttavia, questa idealizzazione ha senso proprio in vista delle obiezioni che verranno sollevate. Infatti, non è tanto controversa questa descrizione iniziale, quanto piuttosto è ritenuta da alcuni complessivamente controversa la caratterizzazione delle proprietà dell'azione che l'agente deve perseguire. Abbiamo, perciò, teorie che ritengono che razionale è la massimizzazione diretta, altre che pensano che si deve massimizzare l'utilità attesa evidenziale oppure causale, (21) altre che ritengono che debbano essere preferite strategie subottimali. (22) Non è, tuttavia, necessario scegliere una in particolare fra tutte queste teorie, poiché l'obiezione è che la rappresaglia sia irrazionale strutturalmente, sin dal momento che la si dichiara, e quindi il bersaglio individuato dalla struttura del paradosso è semplicemente il fatto che avere l'intenzione di attuare la rappresaglia sia irrazionale, senza ulteriori specificazioni sulla teoria che sosterrrebbe quella razionalità presunta. Inoltre, non è controverso che la razionalità sia una proprietà di almeno alcune delle opzioni disponibili all'agente, ossia una proprietà che si applica alla scelta di un'azione, perché soddisfa determinati elementi specificati in maniera indipendente. Ne segue che una decisione può essere razionale se l'opzione scelta è razionale e se il processo che ha condotto a quella decisione è connesso in maniera affidabile alla scelta di opzioni razionali.

La razionalità dell'agente è perciò l'insieme delle condizioni che rendono possibile all'agente possedere la disposizione a scegliere razionalmente. Un agente pienamente razionale è quindi quello che decide la maggior parte delle volte razionalmente perché possiede la disposizione a farlo, come per Aristotele l'agente giusto è quello che possiede la disposizione a compiere atti giusti. (23) Sarebbe corretto, allora, dire che se l'agente sceglie razionalmente, mentre è la sua disposizione a rendere possibile la deliberazione adeguata per l'azione appropriata, è l'azione appropriata ad essere razionale più che l'agente, allo stesso modo in cui sarebbe corretto dire che è l'azione di un uomo giusto ad essere giusta più che l'agente che la compie? Sembrerebbe di no. Infatti, per mantenere la struttura basilare del paradosso, è proprio la razionalità centrata sulle opzioni, più che quella centrata sull'agente, che deve essere enfatizzata, perché quale sia la natura dell'agente ciò che risulta in determinate circostanze è proprio qualcosa di irrazionale (o immorale). (24)

Naturalmente, molto può essere detto a favore di una teoria basata sull'agente, specialmente se in campo morale abbracciamo una teoria perfezionista o un'etica della virtù. Tuttavia, queste specifiche teorie della motivazione morale non credo possano essere gran che illuminanti riguardo al nostro problema, che conviene ricordare an-

cora una volta: manifestare l'intenzione di fare qualcosa, se qualcos'altro si verificherà, è moralmente e razionalmente corretto, anche se fare quella cosa dopo che quel qualcos'altro si è verificato è sia immorale sia irrazionale. Quindi, il problema è il genere di conseguenze che tale paradosso ha sulle proprietà morali e razionali dell'agente. Per questo è francamente difficile comprendere come teorie morali basate sulle motivazioni dell'agente possano affrontare il paradosso. Piuttosto per queste teorie il problema non dovrebbe porsi affatto, poiché nel loro caso la moralità e la razionalità delle opzioni disponibili sono caratteristiche semplicemente derivate dalla moralità e dalla razionalità dell'agente. (25) Le teorie basate, invece, sulle caratteristiche dell'opzione anziché su quelle dell'agente concepiscono la scelta appropriata come una proprietà fondamentale dell'opzione. Questo può avvenire in base a criteri distinti, ad esempio facendo appello a determinate caratteristiche teleologiche, come potrebbe essere il caso delle varie teorie che invocano la massimizzazione dell'utilità oppure di quelle che si concentrano sugli esiti subottimali, oppure evidenziando l'adeguatezza di un certo insieme di opzioni ad altri criteri specificati indipendentemente, come potrebbe accadere perfino in certe teorie intuizionistiche. (26) La moralità dell'agente così come la sua razionalità è in queste versioni in sintonia con le caratteristiche delle opzioni disponibili.

5. *Disposizioni*

Si può obiettare sollevando l'accusa di irrazionalità alla disposizione di un agente che agisce in questo modo. Infatti, se un agente razionale è colui il quale ha la disposizione ad agire razionalmente, avere l'intenzione di compiere un atto irrazionale in circostanze specificate, ossia precisamente quelle dello schema paradossale che si è descritto, significa predisporre ad agire irrazionalmente. Quindi, l'agente razionale non può avere questa intenzione. Questa considerazione sembra essere molto plausibile, ma penso possa essere sottoposta a due critiche. In primo luogo, il legame tra intenzioni e disposizioni è tutt'altro che chiaramente definito. Io posso avere l'intenzione di fare qualcosa se si verificano determinate circostanze, e tuttavia l'insorgere di circostanze impreviste può modificare completamente i termini della questione e farmi recedere dall'attuare la mia intenzione, senza che la mia disposizione razionale ne sia intaccata. Non è affatto chiaro che non aver ragionevolmente previsto tutte le circostanze, che possono modificare le reazioni dell'agente possa mettere in questione la razionalità dell'agente stesso. Il mondo in cui viviamo non è, dal punto di vista del nostro problema, un sistema deterministico, e l'incapacità di prevedere tutte le circostanze che possono influenzare la nostra valutazione di un evento, in alcun modo può essere considerata una carenza di razionalità, anche perché, o se non altro perché, si tratta di una capacità universalmente condivisa.

Insomma: nessuno di noi è il demone di Laplace e questo è semplicemente uno degli elementi della nostra descrizione di agenti che scelgono. Questo per quanto riguarda il ventaglio di opzioni che possono sorgere senza alterare la razionalità della scelta che noi avremmo ritenuto opportuno compiere. Ma c'è un elemento forse più importante del quale tenere conto. Possiamo concedere che la presenza di una determinata intenzione renda una certa disposizione irrazionale, ma la presenza di questa stessa disposizione dice ben poco sulla razionalità dell'agente. E la ragione è analoga a quella appena svolta. La disposizione ad agire razionalmente comporta certamente la capacità immaginativa di pensare a circostanze che non si verifichino effettivamente nel futuro, ma non comporta affatto avere la capacità di considerare tutte le circostanze. I miei occhiali possono essere estremamente fragili, sebbene io possa immaginare circostanze in cui la fisica di questo oggetto sia così radicalmente alterata da renderli estremamente resistenti, ma la cosa importante è che nelle circostanze rilevanti gli occhiali si comportino effettivamente come oggetti delicati e inclini alla rottura, se non sono maneggiati con la dovuta cautela. Per questo stesso motivo, un agente può essere considerato razionale in tutte le circostanze rilevanti anche se occasionalmente in situazioni di alterazione comportamentale, ad esempio in seguito all'assunzione involontaria di droghe, può comportarsi irrazionalmente. (27)

Queste considerazioni non sono delle ipotesi *ad hoc*, poiché, se non vanno sottovalutate le circostanze che possono condurre ad alterare la personalità dell'agente, è chiaro che anche una frequenza episodica a scegliere le opzioni razionali non è semplicemente sufficiente a fare di colui che sceglie un agente razionale, dal momento che l'agente potrebbe aver scelto nella maniera giusta solo per caso. Allo stesso modo, se le circostanze che producono l'alterazione nella capacità decisionale dell'agente sono quel genere di circostanze che provocano una diminuzione involontaria delle capacità cognitive dell'agente, diminuzione per la quale l'agente non è quindi responsabile, e che magari tenta di evitare con tutti i mezzi disponibili, allora sembra che non ne sia coinvolto il nostro discorso sull'agente razionale come capace di scegliere le opzioni adeguate.

Ciò che rende ciascuno di noi un'approssimazione insoddisfacente dell'agente che agisce razionalmente non è il fatto che noi saremmo condotti ad agire irrazionalmente, se fossimo colpiti da una qualche circostanza che riduce le nostre capacità cognitive. Queste circostanze ipotetiche di riduzione o annullamento delle nostre capacità deliberative razionali mostrano solo che, se si verificano, allora noi agiamo irrazionalmente. La critica basata sulle disposizioni non costituisce quindi una accusa efficace all'idea di agente razionale. Se poi volgiamo l'argomentazione basata sulla razionalità in una basata sulla moralità, le cose sostanzialmente non cambiano. Un agente che sia tanto morale quanto razionale è ovviamente disposto e capace di

scegliere le opzioni razionali e morali, ma non in qualsiasi situazione controfattualmente immaginabile.

6. *Prevenzioni*

L'agente razionale può essere del tutto consapevole dell'irrazionalità di fare un certo atto al quale si è preventivamente vincolato, ossia non ne sarà semplicemente consapevole quando le circostanze controfattuali si verificheranno, ossia vi si è vincolato precisamente perché le circostanze controfattuali non si verifichino. La sua consapevolezza della struttura del paradosso è completa e riflessiva, ed è precisamente questo che la rende un paradosso. Forse è questa riflessività che deve essere appropriatamente colta. Immaginiamo un agente, che viene a conoscenza che in un momento t del prossimo futuro, se si verificherà un determinato evento, ad esempio l'esposizione a determinate sostanze tossiche in possesso di certi suoi avversari, comincerà a perdere tutte le sue capacità deliberative razionali e inizierà a esibire comportamenti patologici. L'agente sa anche che sono le conseguenze prevedibili di questo evento che trattengono i suoi avversari dall'esporsi a quella tossicità che lo renderebbe irrazionale. Esistono tuttavia delle cure preventive che renderebbero l'agente immune dagli agenti tossici in via definitiva. Tuttavia, l'agente decide di non sottoporvisi. In che senso, questo agente, che sa che si comporterà irrazionalmente in presenza di determinate circostanze, agisce irrazionalmente? Il fatto che accetti l'eventualità di agire in maniera irrazionale non rende irrazionale il suo proposito. Piuttosto mostra la sua capacità di indirizzare i tratti disposizionali del suo carattere a proprio vantaggio. Mentre, inoltre, chiaramente vede le possibilità patologiche del suo comportamento futuro, tuttavia non possiamo dire che adotta un comportamento patologico. Ma, allora, per quale motivo la struttura del paradosso sembra implicare un'inconsistenza? Si potrebbe dire che l'esempio che è appena stato fatto sia poco aderente al paradosso stesso, perché non è sensibile alla dinamica della scelta razionale, nel senso che il comportamento patologico del nostro esempio non è qualcosa che nasce riflessivamente dall'agente, ma è casomai causato da condizioni esterne. Nel paradosso le cose sembrerebbero andare diversamente, perché ciò che si afferma è che se si intende fare una certa azione in presenza di determinate condizioni, l'opzione deve dipendere dal fatto che l'agente (se è un agente razionale) non la riconosce come quella giusta da scegliere. Per questo motivo non può scegliere la rappresaglia, perché non si tratterebbe dell'azione razionale da fare. Ma secondo questa lettura la propensione condizionale a compiere un certo atto dovrebbe essere analizzata nei termini dell'assenso riflessivo dell'agente a compiere proprio quell'atto, ossia a comprenderlo non come condizionalmente razionale ma come incondizionatamente razionale. Tuttavia, se questa è la motivazione che

supporta tale linea di ragionamento, allora dovremmo essere piuttosto sospettosi ed il motivo è che questa procedura comporta una interpretazione particolare del modo in cui il soggetto dovrebbe decidere, ossia una specifica interpretazione della teoria delle decisioni, la quale è ben lungi dall'essere pacificamente accettata. Mi spiego: ciò che la teoria delle decisioni richiede è che le decisioni prese dell'agente siano in accordo sostanziale e sistematico con la teoria della razionalità che l'agente abbraccia e non invece che venga indicata a priori una specifica procedura di decisione riguardo alle opzioni in gioco. L'idea che le opzioni siano da intendersi realmente come non condizionali fornisce una visione erronea delle capacità di scelta dell'agente razionale. L'agente non sceglie determinando riflessivamente che le sue opzioni siano coerenti con una qualche procedura a priori, ma l'accordo riflessivo prodotto da una scelta strategica è la soddisfazione di ciò che viene identificato come un desiderio accompagnato da una motivazione. (28) Per questo motivo una difesa della strategia che abbiamo delineato non è particolarmente efficace se viene condotta in nome delle procedure che gli agenti dovrebbero seguire se vogliono conformarsi a una qualche regola delle decisioni.

Naturalmente è possibile che un agente decida di modellare lo schema della sua scelta delle opzioni, dopo accurato studio, a una qualche versione della teoria delle decisioni, ma non è affatto necessario, come tutti comprendono, che questo debba accadere. E se non è necessario che accada, allora talvolta può non essere nemmeno possibile, per cui, se non siamo in grado di individuare altre ragioni, sembrerebbe ragionevole supporre che il modo di selezionare le opzioni non deve necessariamente essere nei termini del processo descritto. Un agente razionale, quindi, semplicemente riconoscerebbe l'irrazionalità di un'intenzione condizionale, e questo lo renderebbe incapace di formare l'intenzione corrispondente? Se è possibile che l'agente sia in grado di concepire se stesso come un agente che in determinate circostanze agisce irrazionalmente, è sufficiente questo per farne un agente irrazionale nella disposizione a scegliere determinate opzioni? In termini generali, se l'ascrizione presuntiva fosse di irrazionalità, giocare il gioco della minaccia non porterebbe a nessun risultato simmetrico, ma questo è smentito anche dall'evidenza empirica. Effettivamente, del resto, quando si esaminano progetti futuri e linee d'azione proprie o altrui di soggetti sia individuali sia collettivi non è affatto privo di senso fare previsioni in base al comportamento razionale che attribuiamo all'agente. (29) In questo senso gli agenti che popolano la nostra vita ordinaria sono molto spesso razionali e questa ascrizione presuntiva come caratteristica dei loro comportamenti è fondamentalmente corretta, sia la si sia raggiunta per via empirica, attraverso una accurata osservazione dei comportamenti specifici di un certo agente o di un certo gruppo – osservazione che può ben essere anche un'auto-osservazione –, sia si tratti di un'assunzione in larga misura a priori – se io assumo che non è razionale per un istanza amministrativa inferiore agire lungo linee difformi da

un'istanza amministrativa superiore, sto facendo un'assunzione a priori sulla razionalità comportamentale di un certo settore burocratico -. Una imputazione di razionalità di questo genere, sia induttiva sia a priori, sembra essere appropriata sia per la nostra vita reale sia, a maggior ragione, per il caso ideale, anche in presenza di circostanze altamente sfavorevoli. Se un agente sia effettivamente dotato di razionalità in una determinata condizione di scelta dipende spesso dalle contingenze della situazione. Se è afflitto da una patologia cerebrale degenerativa, questa condizione lo può rendere nelle fasi non ultime della malattia perfettamente consapevole della possibilità di comportarsi irrazionalmente in futuro, senza che questo sia un argomento contro la razionalità dei suoi comportamenti presenti e può predisporre le cose in maniera da tutelarsi quando non sarà più presumibilmente in grado di agire razionalmente.

7. Integrità e desideri

Ammettiamo che si verifichi la condizione prevista dalla struttura del nostro paradosso. La natura di questo evento è eccezionale ed è quindi del tutto pacifico che l'agente sarà indotto a credere ora che la sua capacità di azione razionale sarà preservata anche in quelle circostanze eccezionali? Ovviamente, ci possono essere degli agenti che sanno bene che saranno incapaci di fare quello che si sono proposti, se le condizioni per lanciare una rappresaglia si verificheranno effettivamente. È possibile che questi agenti siano sufficientemente certi delle proprie motivazioni per sapere che non saranno mai in grado di vincolarsi al lancio di una rappresaglia, se le circostanze dovessero rendere disponibile questa opzione. Questi agenti sono pienamente razionali solo se hanno le conoscenze adatte a sapere che saranno proprio queste ad impedire loro di mettere in opera quella deterrenza, la cui minaccia è pienamente congruente ai loro interessi fondamentali? L'agente che compie questo esperimento mentale immagina che la scelta dell'opzione sarà determinata dal peso delle ragioni che militeranno a favore dell'opzione, precisamente perché si immagina come un agente che possiede un'integrità deliberativa rispetto alle circostanze possibili in cui avverrà la sua scelta. Per questo motivo il paradosso gli appare come autentico, perché formula delle condizioni che mettono in questione la sua integrità in scelte nel corso del tempo. In questo senso, formare l'intenzione di fare qualcosa che rispecchi il paradosso nell'azione richiede di più che credere, sapere, o anche causare il fatto che se si verificheranno determinate condizioni l'agente allora attuerà una rappresaglia, perché una scelta di questo genere costituirebbe una possibile violazione alla sua integrità deliberativa, che costituisce precisamente la descrizione dell'agente come agente razionale.

L'agente che ragiona in questa maniera mostra di identificarsi con determinate ragioni con le quali si identifica anche nelle sue scelte non condizionali. È, cioè, per lui parte del concetto di avere un'intenzione razionale che la disposizione a compiere l'atto intenzionato sia giustificata e causata in modi appropriati al modo in cui l'agente crede si dovrebbe ragionare e deliberare in circostanze simili o identiche. Potrebbe, in effetti, essere che poiché ritiene di avere ragioni conclusive per considerare l'argomento un paradosso non sia disposto ad ammettere che chi applica la rappresaglia agisca razionalmente. L'idea è che l'agente che sceglie la possibilità di mettere in atto la rappresaglia deliberi in una maniera tale che non gli permette di identificarsi con un agente razionale, poiché qualunque agente razionale non può trovare l'argomento persuasivo, a meno che non voglia rinunciare alla sua integrità deliberativa, decidendo di essere irrazionale.

Tuttavia, questo modello lascia da parte elementi importanti del processo decisionale, e precisamente il ruolo che le emozioni e le passioni giocano nella deliberazione. Ad esempio, una decisione può essere presa perché l'agente è stato guidato dall'ira, ma può ben essere che sia anche la decisione giusta da prendere su basi razionali. Se le cose stanno in questa maniera, allora non è l'emozione o la passione che fa sì che un agente decida razionalmente o irrazionalmente. Unicamente nel caso in cui la passione lo privi di capacità cognitive rilevanti, impedendogli di vedere quale opzione possieda le caratteristiche adeguate alla sua versione di razionalità, possiamo allora dire che ha agito in quel modo perché reso cieco dalla passione. Tuttavia, per lo stesso motivo, il fatto che un agente compia talvolta delle azioni perché viene mosso da determinate passioni, non è una ragione sufficiente per pensare che non si tratti di un agente razionale, poiché ciò che fa in modo che quell'agente agisca motivato dalla passione non determina necessariamente alcuna caratteristica irrazionale della sua opzione.

L'idea che l'agente razionale debba essere un soggetto completamente distaccato fa parte delle finzioni letterarie e non di quelle più riuscite. Ci è perfino difficile immaginare un agente che agisca al di fuori del contesto di una motivazione fornita da una qualche passione. Del resto, il concetto mutuato dalla scienza economica di "sovranità del consumatore", segnala, a mio modo di vedere, precisamente un intreccio di questo genere. La soddisfazione che il consumatore persegue è quella dei propri desideri alla luce delle proprie credenze sul mondo. Alcune di queste saranno giustificate, altre saranno vere senza essere giustificate, altre non saranno né vere né giustificate. Ossia: non possiamo dire a priori quali desideri siano quelli che caratterizzano un agente come irrazionale. È semplicemente una precondizione dell'azione avere dei desideri, e il possesso dei desideri si intreccia in maniera duplice con quello delle emozioni. In primo luogo, un gruppo numeroso dei desideri che coltiviamo, può essere descritto solo in termini di possesso di emozioni. Se raccolgo un gatto ammalato dalla strada, il mio desiderio per la sua sopravvivenza deve

essere descritto alla luce della mia emozione di pietà per il suo stato di animale ammalato. Il mio desiderio è quindi guidato dalla mia emozione, ma questo non mi impedisce affatto, e, anzi, mi consiglia, di portarlo da un veterinario, affinché gli somministri le cure adeguate al suo caso. In secondo luogo, se un agente decide che un determinato corso d'azione da intraprendere è quello che meglio si attaglia alle sue aspirazioni e desideri, perché dovremmo chiedergli di assumere un punto di vista neutrale rispetto a un altro agente che intenda promuovere un corso d'azione contrario e che impedirebbe la realizzazione dei propri? Il conflitto e il risentimento non è affatto detto che siano in questo senso irrazionali. Poiché possedere determinate categorie di emozioni rende possibile l'identificazione dell'agente con i suoi desideri, sembra essere una conclusione non fondata che questa identificazione necessariamente lo vincoli a una azione meno che razionale.

Queste considerazioni devono perciò guidarci verso una definizione di integrità deliberativa più ampia di quella utilizzata precedentemente. Il concetto di integrità deliberativa in realtà media fra deliberazione condizionale e non condizionale, nel senso che un agente che riflette su opzioni condizionali spesso non fa astrazione dalle sue dotazioni in termini di desideri e di emozioni e della conoscenza che di questo bagaglio hanno gli altri agenti.

Quando un agente coltiva intenzioni che potrebbero guidarlo verso la vendetta, ossia ha intenzione di mettere in gioco le sue capacità di deterrenza, segnala che le condizioni, condizioni magari per le quali l'agente stesso attivamente si adopera che non si verifichino, disponibili alla sua azione di rappresaglia riguardano il suo bagaglio emozionale in una maniera sostanziale. L'agente non è costretto a pensarsi come irrazionale anche avendo chiara la struttura del paradosso. Il punto è che nella situazione immaginativa l'agente non è affatto obbligato a immaginarsi come uno psicopatico le cui motivazioni non sono chiare, bensì piuttosto come agente che agisce in base alle intenzioni assunte come impegni condizionali a partire anche dal suo bagaglio emotivo. Tutto questo senza che ne sia intaccata l'integrità decisionale. Fa, anzi, parte della strategia della deterrenza che colui al quale tale strategia si rivolge comprende queste cose in maniera del tutto chiara, assieme al fatto ovvio che l'agente possiede le risorse per attuarla. Anzi: comunicare che, almeno sino a un certo punto, saranno legate a procedure automatiche accresce il valore e l'efficacia della deterrenza. La strategia nota come "bruciarsi i ponti alle spalle" rafforza nell'avversario l'idea che il bagaglio emotivo che guida la scelta della deterrenza è così forte da non lasciare spazio ad altre opzioni una volta che si verificassero le condizioni che rendono la rappresaglia possibile. Dal punto di vista dell'agente, il suo avversario deve essere messo nelle condizioni di avere la credenza che l'agente è pienamente razionale, ha integrità deliberativa, riconosce il proprio bagaglio emozionale come parte essenziale della propria auto-descrizione, e, eventualmente, mette in opera le strategie necessarie per evitare l'insorgere di sé successivi che po-

trebbero non essere vincolati alle sue procedure deliberative e alla sua concezione della razionalità. (30) Poiché la descrizione del proprio sé come vincolato a determinati valori che vengono pensati come del tutto essenziali alla percezione di se stessi non è affatto in contrasto con la propria descrizione in termini di agente razionale, la deterrenza è una scelta coerente con l'idea di agente razionale. Anzi: una sua maggiore o minore efficacia si può dire sia connessa alla capacità dell'agente di diffondere questo insieme di informazioni.

Si è detto che tenere in conto le proprie emozioni non significa affatto rinunciare a un agire strategico che persegua la soddisfazione dei propri desideri. Si potrebbe, però, obiettare che le emozioni sono delle distorsioni alla propria capacità deliberativa razionale. In fin dei conti, un agente razionale dovrebbe precisamente essere capace di ragionare a mente fredda, ossia dovrebbe essere in grado di mettere da parte quelle emozioni che offuscano la sua capacità di giudizio razionale, come accade molto spesso nel caso della vendetta. Ammettiamo che l'agente razionale possieda in effetti la capacità di mettere da parte le proprie emozioni o almeno quelle più forti, ossia quelle che generalmente si immagina influenzino, distorcendolo, il nostro giudizio. Sono, tuttavia, due cose chiaramente diverse possedere un'abilità e avere la necessità di usare questa capacità. Ci possono essere delle occasioni dove questa possibilità di fare uso della messa fra parentesi delle proprie emozioni può a ragione essere saggiamente esercitata, attraverso l'utilizzo di determinate procedure, ad esempio immaginative o attraverso lo scambio dialogico con altri agenti, particolarmente accreditati nelle decisioni che dobbiamo prendere. (31) In altre circostanze, invece questo utilizzo di una propria capacità non è necessario e potrebbe anche essere dannoso. Se l'assunzione di un farmaco blocca determinate mie reazioni che mi potrebbero portare a reagire a una circostanza che ritengo particolarmente dannosa per me, assumerlo mancherebbe di ottenere l'effetto di deterrenza che inibisce un altro agente dal produrre proprio quella circostanza per me dannosa. La capacità di agire di un soggetto razionale in circostanze controfattuali è guidata dalla sua capacità immaginativa, dalla capacità di costruire con il pensiero corsi d'azione e di pensarne le conseguenze. Ma questa stessa capacità è davvero una buona guida per agire se si verificano quelle circostanze indicate dalla struttura del paradosso? Ed è una buona guida in generale? La rappresentazione artistica, ad esempio, suscita in noi ideali di comportamento che anticipiamo con l'immaginazione, ma che il più delle volte ben difficilmente saremmo in grado di mettere in atto nella vita reale. Non sarebbe bene per un agente razionale essere scettico sull'insieme delle proprie capacità immaginative relativamente alla sua capacità di prevedere il corso futuro delle azioni e di anticipare quelle che saranno le proprie reazioni nelle situazioni di crisi? Le procedure di decisione automatiche in caso di attacco rispondono a problemi di questo genere.

Il problema è di notevole portata nel caso di alleanze fra l'agente e altri soggetti. Il peso di emozioni che sorgono nel caso in cui ad essere colpiti siano alleati e non l'agente è immaginabile non sia lo stesso di un attacco diretto verso l'agente stesso. Di questa diversità fa le spese la credibilità della deterrenza, quando l'agente si sia impegnato a difendere i propri alleati, poiché la presunzione è che il nostro avversario comprenda le nostre motivazioni arricchite dal nostro bagaglio emozionale, ossia che, in fondo, sia una sorta di specchio. (32) Questa è un'eventualità significativa, ma non riguarda realmente il caso che si sta esaminando, dove un attacco è effettivamente percepito come una volontà di colpire gli interessi e i desideri più profondi dell'agente. È del tutto sensato immaginare che questo possa accadere sia nel caso di un attacco diretto sia nel caso di un attacco a dei propri alleati. Non c'è per definizione una intersezione imperfetta di interessi ed è questa mancanza che rende la pubblicità del bagaglio emotivo dell'agente un indicatore della sua capacità di reagire.

8. *Rottura e integrità*

Tuttavia, in definitiva, il tuo obiettivo di essere percepito come un agente capace di attuare una rappresaglia può non avere buon esito. La tua capacità di deterrenza può essere percepita come imperfetta e alla fine la situazione condizionale può verificarsi. Se tu finisci per fare quello che ti sei ripromesso secondo lo schema del paradossale, agirai irrazionalmente, anche se dal punto di vista della motivazione sei razionale come l'agente la cui deterrenza ha invece successo. Quindi non c'è reale differenza fra te e chi nella deterrenza ha avuto successo. Eppure, tutto questo sembra essere molto controintuitivo. Che cosa si può rispondere? In primo luogo, è bene ricordare la differenza e la distanza che di solito sussiste fra intenzione ed esecuzione. La tua intenzione può essere razionale, ma quando l'evento condizionale si verifica può palesare una serie di implicazioni che ti erano oscure, senza per questo vanificare il tuo essere razionale, e quindi sconsigliarti di attuare quello che con la deterrenza ti eri proposto oppure rafforzandoti con ulteriori elementi nella tua intenzione precedente. Il mondo è notoriamente un posto complicato che dobbiamo affrontare con conoscenze imperfette: circostanze, fortuna, cattiva sorte, buona stella sono lo scenario entro il quale sempre avviene che si realizzino e progetti e condizioni per ora solo pensate come condizionali futuri. Pensare che queste metafore indichino nulla di più che corrispondenze analogiche significa dare un'immagine irrealistica tanto della razionalità quanto della moralità. Razionalità e moralità sono parte del mondo, non sono ciò che domina il mondo. È, anzi, forse proprio perché il mondo si incarica tanto spesso di smentire le nostre attese che ci siamo dotati di questi strumenti. Rimane il fatto che un agente realmente disposto

ad esercitare la deterrenza potrebbe in fin dei conti essere considerato un agente irrazionale in un qualche senso fondamentale della sua capacità di azione? A un'obiezione di questo genere non si può far altro che opporre la richiesta di esempi concreti che mostrino in atto una evidenza generalizzabile, dal momento che abbiamo ammesso che il nostro agente non è incline in maniera patologica alla vendetta e si è anche ammesso che non è affatto certo che queste emozioni spianeranno la strada a comportamenti irrazionali. In altre parole, coltivare deliberatamente una determinata dose di emozioni può essere del tutto razionale, anche se alcune condizioni ipotetiche, se realizzate, rendono l'esercizio di quelle emozioni un atto non desiderabile. Per questo non c'è motivo di pensare che l'agente non possa essere razionale anche se immagina le situazioni che potrebbero condurlo ad operare la rappresaglia e a mettere in atto una serie di minacce addizionali. Pensare che sia irrazionale in questo stadio delle sue capacità equivale ad accreditarlo come un agente che soffre di qualche disturbo cognitivo e non come un attore che massimizza la sua utilità secondo il suo sistema di valori.

La cosa interessante della struttura del paradosso è che siamo in presenza di un agente, che per motivi interni ai propri scopi deliberatamente affronta il rischio di un possibile processo di auto-corrruzione cognitiva? Ma si tratta di un processo che realmente si allontana dallo stato di possedere un'intenzionalità razionale? Se è parte del concetto di avere razionalmente l'intenzione di compier qualcosa, che la disposizione a compiere l'atto inteso sia causato (o giustificato) in maniera appropriata dalla concezione propria dell'agente delle ragioni per fare quell'atto, bisogna forse aggiungere tale concezione potrebbe andare bene per un agente che forma le proprie intenzioni per l'azione solo quando il mondo sembra andare come vuole lui. Certamente, tutti abbiamo la tendenza a volere che il mondo si adatti ai nostri desideri, ma la realtà si incarica piuttosto spesso di smentire questo nostro persistente narcisismo. Un agente di questo ultimo genere, l'agente che, in realtà, tutti talvolta siamo, può adottare delle strategie per contrastare questa inclinazione a formare motivazioni intenzionali solo quando le cose vanno per il verso giusto. Ad esempio, può esercitarsi a strutturare delle motivazioni secondarie per agire – anche per fare quello che comunque farebbe –, contribuendo così a complicare il proprio modello di razionalità nell'azione. È una sorta di terapia cognitiva che il soggetto disposto alla deterrenza deve aver messo precedentemente in atto in qualche maniera. Ma questo è realmente un processo di auto-corrruzione cognitiva? Da un lato, infatti, può condividere l'idea che la rappresaglia non può costituire il raggiungimento di nessuno dei suoi scopi ed anzi può contribuire a comprometterne alcuni di grande importanza, ma, da un altro, non manca di segnalare all'avversario che la capacità di reagire inclusa nella sua deterrenza riguarda il suo più profondo bagaglio emotivo.

La deterrenza è precisamente questo: un comportamento emotivamente comprensibile all'avversario e interlocutore. La minaccia dell'agente non mostra affatto, invece, che l'agente coltiva effettivamente il desiderio di attuare la rappresaglia. Questo darebbe una descrizione radicalmente sbagliata della natura e dello scopo della minaccia in situazioni limite, che è di essere al servizio delle preferenze più profonde e più funzionali all'auto-descrizione dell'agente, che coltivandola può continuare a rimanere un agente razionale. Quello che si deve concludere è che è errato supporre che nessun agente razionale può sensatamente avere l'intenzione di intraprendere una rappresaglia e coltivare quindi la corrispondente intenzionalità deterrente. Questo significa che effettivamente l'agente in questione manca di integrità deliberativa? Questo seguirebbe se si dimostrasse che il patrimonio emotivo, così importante per colui che attua la deterrenza, determina scelte irrazionali passo a passo, cosa che è invece non è affatto dimostrata dalla struttura stessa del paradosso. Schelling, nel suo studio pionieristico e così influente, parlava della deterrenza come una minaccia che lascia aperte le porte al caso, (33) ad esempio al ruolo delle emozioni, alla disinformazione, alla rottura della catena di comando, volendo forse significare che, anche se il patrimonio emotivo non qualifica affatto un'intenzione come razionale, tuttavia, fornisce una interpretazione plausibile della intenzione guida della deterrenza. Plausibile perché conserva all'agente la sua auto-descrizione assieme a elementi importanti della sua razionalità e della sua integrità deliberativa.

Note

(1) La bibliografia su questi temi è molto vasta; indico alcuni titoli che hanno costituito lo sfondo di questo lavoro: A. Wohlstetter, *The Delicate Balance of Terror*, Foreign Affairs, 1959, 37, 2; R. Aron, *Il grande dibattito. Introduzione alla strategia atomica* (1963), Bologna, Il Mulino, 1965; A. Baufre, *Introduzione alla strategia* (1963), Bologna, Il Mulino, 1966; H. Kahn, *La filosofia della guerra atomica. Esempi e schemi* (1965), Milano, Il Borghese, 1966; T. Schelling, *La diplomazia della violenza* (1966), Bologna, Il Mulino, 1968; T. Schelling, *The Strategy of Conflict*, Cambridge, Mass., Harvard University Press, 1980²; H. Kahn, *Thinking about the Unthinkable in the 1980s*, New York, Simon & Schuster, 1984; D. Gauthier, *Deterrence, Maximization, and Rationality*, in D. Gauthier, *Moral Dealing*, Ithaca, Cornell University press, 1990, pp. 298–321; R. Axelrod, *Giochi di reciprocità. L'insorgenza della cooperazione* (1984), Milano, Feltrinelli, 1985; D. Copp (ed.) *Nuclear Weapons, Deterrence, and Disarmament*, Calgary, University of Calgary Press, 1986; G. Kavka, *Moral Paradoxes of Nuclear Deterrence*, New York, Cambridge University Press, 1987; E. Luttwak, *Strategia* (1989), Milano, Rizzoli, 2001; J. De Nardo, *The Amateur Strategist*, New York, Cambridge University Press, 1995;

C. Buttasi, *Tecniche della guerra non convenzionale*, in E. Manganaro Favaretto (ed.), *La guerra. Una riflessione interdisciplinare*, Trieste, Edizioni Università di Trieste, 2003, pp. 225–246; R. Festa, *Minacce e promesse: logica e metodologia della deterrenza*, Trieste, Edizioni Università di Trieste, 2003, pp. 247–298.

(2) M. Zezima, *Salvate il soldato potere* (2000), Milano, Il Saggiatore, 2004.

(3) L'acronimo è fuorviante, perché implica che il risultato di qualsiasi conflitto atomico sia, appunto, la distruzione reciproca. Questa idea è, tuttavia, altamente speculativa e, in realtà, maggiormente speculativa di quella opposta, ossia che esista una tattica del conflitto atomico, dal momento che ipotizza un collasso di razionalità nei contendenti nel momento in cui il conflitto inizia. Fortunatamente, le evidenze empiriche dirette per sostenere l'una o l'altra ipotesi attualmente non sono disponibili, ma una generica presunzione di razionalità dovrebbe indirizzare verso la seconda ipotesi. Si vedano gli scenari ipotizzati in H. Kahn, *La filosofia della guerra atomica. Esempi e schemi*, cit.

(4) Questo è in accordo con la definizione di minaccia utilizzata da T. Schelling, *The Strategy of Conflict*, cit., p. 35–36: «When one threatens to fight if attacked or to cut his price if his competitor does, the threat is no more than a communication of one's own incentives, designed to impress on the other the automatic consequences of his act. And, incidentally, if it succeeds in deterring, it benefits both parties.»

(5) La minaccia asimmetrica coinvolge non solo dimensioni tattico-strategiche, ma anche importanti profili psicologici ed economici. Per i primi, si veda M. Juergensmeyer, *Terroristi in nome di Dio* (2000), Roma-Bari, Laterza, 2003; per i secondi, L. Napoleoni, *La nuova economia del terrorismo* (2003, Milano, Troppa, 2004, oltre a quello che si sta imponendo come uno dei testi di riferimento, ossia Q. Liang & W. Xiangsui, *Guerra senza limiti* (1999), Gorizia, Libreria Editrice Goriziana, 2001.

(6) Per un inquadramento generale, si vedano le analisi di C. Jean, *Guerra, strategia e sicurezza*, Roma-Bari, Laterza, 1997 e il recente C. Jean, *Geopolitica del XXI secolo*, Roma-Bari, Laterza, 2004.

(7) Per eliminare l'alone di paradossalità dal concetto di deterrenza è necessario, per prima cosa, renderlo un concetto elastico, come è esemplificato da H. Kahn, *Thinking about the Unthinkable in the 1980s*, cit., p. 109, il quale distingue tre generi di deterrenza: «Deterrence is not a single, or simple, concept. It is important to distinguish between three different and especially significant types of deterrence. "Type I" is a strategic deterrence of a massive direct attack against one's country or a large portion of its forces. The deterrence is the threat of a major retaliation. "Type II" is deterrence against extremely provocative acts other than a massive direct attack against one's country or its major forces (e.g., an attack, or threat of attack, on an ally). Strategic threats provide the deterrent. "Type III" is tit-for-tat

deterrence against limited military or nonmilitary actions. The deterrent is the fear that some kind of proportionate retaliation will make the aggression unprofitable.»

(8) G. Kavka, *Moral Paradoxes of Nuclear Deterrence*, cit., p. 19, che costituisce una diretta confutazione di ciò che Kavka chiama *Wrong Intentions Principle*. Kavka è incline a considerare le intenzioni come parti dell'azione e ad assegnare a ciascuna intenzione il medesimo contenuto di razionalità e moralità che si assegna alle azioni che incorporano quelle intenzioni.

(9) Come si esprime efficacemente Kahn, *Thinking about the Unthinkable in the 1980s*, cit., p. 59: «One Gedanken experiment that I have used many times and in many variations in talks given over the past twenty–five or thirty years begins with the statement: “Let us assume that the president of the United States has just been informed that a multimegaton bomb has been dropped on New York City. What do you think he would do?” When this was first asked in the mid–50s, the usual answer was: “Press every button for launching nuclear forces and go home”. At that time (and into the very early 1960s) the concept of a firm balance of terror was (correctly) not fully accepted because the U.S. had great strategic superiority. [...] In any case, the dialogue between the audience and myself continued more or less as follows, after they suggested “press every button”:

Kahn: “What happens next?”

Audience: “The Soviets do the same!”

Kahn: “And then what happens?”

Audience: “Nothing. Both sides have been destroyed.”

Kahn: “Why then did the American president do this?”

A general rethinking of the issue would follow, and the audience would conclude that perhaps the president should not launch an immediate, all–out retaliatory attack.»

(10) Come è spiegato da N. Chomsky, *The Rationality of Rational Suicide*, in D. Copp (ed.), *Nuclear Weapons, Deterrence, and Disarmament*, pp. 23–39, per ciò che riguardava il conflitto fra le due superpotenze. D'altra parte, Chomsky pensava, come molti, che la probabilità, anche bassa, di conflitto rimanesse una forma di “pazzia istituzionale”.

(11) E con ciò, venendo meno alla natura stessa del conflitto armato; cfr. C. Jean, *Guerra, strategia e sicurezza*, cit., p. 146: «Tutti sembrano dimenticare che le guerre sono profondamente diverse a seconda della natura della politica, che è all'origine del ricorso alla forza. È quindi dalla politica che si deve partire, non dagli strumenti tecnici che essa può impiegare.»

(12) Agenti, che, se esistono, si avvicinerebbero a quelli che D. Gauthier, *Morals by Agreement*, New York, New York, Oxford University Press, 1981 chiama massimizzatori diretti in quanto opposti ai massimizzatori vincolati.

(13) Il più volte citato G. Kavka, *Moral Paradoxes of Nuclear Deterrence*, analizza e distingue oltre a quello richiamato nel testo, altre forme del paradosso. Ad esempio, pp. 24–25: «There are situations (namely SDSs [= Special Deterrence Situations]) in which it would be right for agents, if they could, to perform certain actions (namely forming the intention to apply the sanction), and in which it is possible for some agents to perform such actions, but impossible for rational and morally good agents to perform them»; «In certain situations, it would be morally right for a rational and morally good agent to deliberately (attempt) to corrupt himself.»

(14) Questo sembra essere implicato dalle considerazioni di A. Gewirth, *Reason and Nuclear Deterrence*, in D. Copp (ed.) *Nuclear Weapons, Deterrence, and Disarmament*, cit., pp. 129–159. Per una posizione utilitarista si veda D. Lackey, *Missiles and Morals: A Utilitarian Look at Nuclear Deterrence*, «Philosophy and Public Affaire», 1982, pp. 189–231.

(15) Alcuni ne hanno dedotto che la moralità è superiore alla teoria dell'interesse personale anche in termini di interesse personale. Contro, si veda D. Parfit, *Ragioni e persone* (1984), Milano, Il Saggiatore, 1989, pp. 112–142. I dilemmi della cooperazione, che sorgono ad esempio da situazioni equivalenti al dilemma del prigioniero, sono suscettibili di soluzione – o dissoluzione – perché esistono molti casi in cui abbiamo bisogno di soluzioni morali basate su motivazioni come: affidabilità, riluttanza ad essere dei *free rider*, desiderio di soddisfare il test kantiano, altruismo sufficiente.

(16) Per una sintesi si veda M. Leffler, *The American Conception of National Security and the Beginnings of the Cold War*, «American Historical Review», 1984, pp. 346–400.

(17) Non si tratta delle uniche posizioni in campo, naturalmente, nemmeno fra i filosofi. Ad esempio, D. Gauthier, *Deterrence, Maximization, and Rationality*, cit., pp. 320–321, ha sostenuto la razionalità della rappresaglia se la minaccia dovesse dimostrarsi inefficace: «In itself, of course, nothing could be less cooperative, less directed at mutual advantage, than the use of nuclear weapons. But a retaliatory, deterrent policy is directed at preventing such use – directed at maintaining those conditions in which societies may be brought to recognize the benefits of cooperation. A policy of nuclear deterrence clearly has failed if a nuclear exchange occurs, but the serious alternative to such a policy, in the absence of agreement to eschew all threat behavior, can only be the willingness to accept victimization, to suffer passively a nuclear strike or to acquiesce in whatever the potential striker demands as the price of its avoidance.»

(18) Anche perché la risposta, se razionale, a un attacco deve essere proporzionata e segnalare la possibilità di un'ulteriore cooperazione, come esaminato da H. Kahn, *La filosofia della guerra atomica. Esempi e schemi*, cit.

- (19) Questo presuppone una idea di sistematicità della ragione nei tre grandi contesti della credenza, dell'azione, della valutazione, quale quella che è offerta sinteticamente da N. Rescher, *La razionalità* (1988), Roma, Armando, 1999, pp. 133–146.
- (20) Una definizione analoga la si può rintracciare, ad esempio, in D. Wiggins, *Needs, Values, and Truth*, Oxford, Oxford University Press, 2002³, pp. 215–237, che non a caso commenta testi aristotelici.
- (21) R. Nozick, *La natura della razionalità* (1993), Milano, Feltrinelli, 1995, pp. 71–72, all'interno di una discussione sulla massimizzazione dell'utilità attesa e del principio di dominanza collegati all'analisi del paradosso di Newcomb.
- (22) Come sarebbe il caso del *maximin* rawlsiano incorporato nel principio di differenza" di J. Rawls, *Una teoria della giustizia* (1971), Milano, Feltrinelli, 1981, pp. 77–83.
- (23) Aristotele, *Etica Nicomachea*, libro III.
- (24) Naturalmente, la presunzione sulla razionalità degli agenti che si contendono deve continuare ad essere mantenuta, ma ciò che è rilevante è l'oscillazione fra questa e la razionalità delle opzioni. Una analisi psicologica e empirica sembra rafforzare l'idea che agenti reali non percepiscono realmente il paradosso, come veniva suggerito dall'aneddoto narrato da Kahn. Si veda l'analisi della *irrational consistency* in R. Jervis, *Perception and Misperception in International Politics*, Princeton, Princeton University Press, 1976.
- (25) Ho in mente soprattutto le analisi di T. Nagel, *La possibilità dell'altruismo* (1970), Bologna, Il Mulino, 1994 e T. Nagel, *Uno sguardo da nessun luogo* (1986), Milano, Il Saggiatore, 1991.
- (26) W. D. Ross, *Il giusto e il bene* (1930), Milano, Bompiani, 2004.
- (27) Rimane poi il fatto che possono esistere degli agenti le cui motivazioni ci sembrano così culturalmente distanti dalle nostre da rendere la deterrenza inapplicabile. Si veda il profetico saggio di T. Schelling, *Thinking about Nuclear Terrorism*, in T. Schelling, *Choice and Consequence*, Cambridge, Mass., Harvard University Press, 1984, pp. 309–327.
- (28) Una procedura iterata di questo genere produce probabilmente qualcosa di molto simile a quello che la letteratura indica come "equilibrio riflessivo". L'accostamento non è soltanto analogico, poiché, se nell'equilibrio riflessivo si ha un bilanciamento delle circostanze della scelta con i principi generali sostenuti dall'agente, non accade qualcosa di diverso nel caso che stiamo esaminando. Si tratta di una forma di intuizionismo moderato, quale è quella delineata da R. Audi, *Moral Knowledge and Ethical Character*, Oxford, Oxford University Press, 1981.
- (29) Per la logica dei gruppi si veda il classico di M. Olson, *La logica dell'azione collettiva* (1971), Milano, Feltrinelli, 1983.

(30) È, cioè, la continuità psicologica dell'agente a fornirgli integrità deliberativa. Su questi problemi si veda Michele Di Francesco, *L'io e i suoi sé*, Milano, Cortina, 1998.

(31) È quanto Hume pensava accadesse nel caso dei giudizi influenzati dalle cosiddette “passioni calme”, le quali si conformano al “punto di vista comune”. Su questo problema, mi permetto di rinviare a P. Marrone, *L'io delle passioni. Indagini su Hume*, Trieste–Roma, Goliardiche, 2000, pp. 166–169.

(32) L'idea della logica strategica come uso cosciente del paradosso in E. Luttwak, *Strategia*, pp. 23–42 è fondamentalmente motivata da un assunto del genere.

(33) T. Schelling, *The Strategy of Conflict*, cit., pp. 53–80.