Università di Pisa

Facoltá di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea in Matematica

Anno Accademico 2007/2008

Tesi di Laurea Specialistica

LOGICA MODALE E FONDAMENTI DELLA TEORIA DEI GIOCHI

Candidato

Umberto Grandi

Relatori

Alessandro Berarducci

Hykel Hosni

Controrelatore

Mauro Di Nasso

Indice

Introduzione

La teoria dei giochi nasce con lo scopo di fornire un ambiente matematico unitario per l'analisi di situazioni in cui più individui razionali interagiscono. I modelli utilizzati per descrivere queste situazioni sono appunto i "giochi", come introdotti da Von Neumann-Morgenstern in [vNM47].
In questa tesi ci concentreremo su due particolari classi di giochi non cooperativi, in cui non sono ammessi accordi vincolanti tra i giocatori: i *giochi in forma strategica*, che rappresentano tramite giochi con un solo turno situazioni in cui tutti i giocatori scelgono contemporaneamente la propria strategia, e i *giochi in forma estesa*, che aggiungono ai primi una struttura sequenziale di mosse successive dei diversi giocatori.

La teoria dei giochi è di recente stata oggetto di un sempre crescente interesse di ricerca, grazie anche ad un'opportuna sistematizzazione e a numerose applicazioni nei più svariati ambiti. In questo contesto si è reso necessario formalizzare ed analizzare in dettaglio le ipotesi ed i metodi su cui si fonda l'analisi proposta dalla teoria dei giochi.
Lo sviluppo della teoria ha portato infatti alla definizione di numerosi concetti risolutivi, proposti a soluzione delle diverse classi di giochi. Nello studio sui fondamenti della teoria dei giochi vengono proposte diverse interpretazioni, allo scopo di esprimere le varie ipotesi implicite nella definizione dei concetti risolutivi.

La prima parte della tesi è dedicata a presentare questi concetti: vengono definiti i giochi in forma strategica e i giochi in forma estesa, e presentati i due più semplici concetti risolutivi, l'*equilibrio di Nash* e gli *equilibri perfetti nei sottogiochi*.
Viene posta particolare attenzione nell'esporre con chiarezza due diverse interpretazioni dei concetti risolutivi: *l'interpretazione descrittiva* in cui si indaga sulle condizioni di razionalità e di conoscenza che portano all'effettivo realizzarsi degli esiti indicati, e l'*interpretazione normativa*, che vede il concetto risolutivo associare, ad ipotesi di razionalità sui giocatori, un insieme di esiti ideali del gioco.

A partire dalla metà degli anni ottanta si è incominciato ad utilizzare strumenti di logica modale nello studio dei fondamenti della teoria dei giochi, ottenendo grandi semplificazioni nella formulazione, e nella conseguente

analisi, di questi concetti che altrimenti sarebbero difficilmente formalizzabili.
La logica modale è un'estensione della logica proposizionale, le cui formule sono interpretate su particolari strutture relazionali dette *modelli di Kripke*. Grazie alla versatilità di queste strutture la logica modale viene utilizzata con successo in diversi ambiti, primo fra tutti la ricerca nel campo dell'intelligenza artificiale. Da un punto di vista metateorico la logica modale è interpretabile nella logica del prim'ordine, e la maggior parte delle teorie modali risultano decidibili.
Lo scopo di questa tesi è di mostrare come la logica modale può essere utilizzata per modellizzare entrambe le interpretazioni dei più semplici concetti risolutivi proposti dalla teoria dei giochi.

Per quanto riguarda l'interpretazione descrittiva, viene presentata in dettaglio la teoria modale $\mathbf{S5}_n^C$ e la sua interpretazione come *logica epistemica*, in grado di fornire un modello per le conoscenze dei giocatori. Viene mostrato come le formule di questa teoria possono essere interpretate sulle *strutture di Aumann*, che sono lo strumento classico di dimostrazione di teoremi fondazionali sotto l'interpretazione descrittiva. In modo particolare viene mostrato come l'utilizzo della logica epistemica sia perfettamente equivalente all'utilizzo delle strutture di Aumann, risultando molto più agevole e permettendo inoltre di tradurre i risultati dimostrati in formule di $\mathbf{S5}_n^C$, che è una teoria decidibile.

Nel capitolo seguente vengono esposte due teorie modali sufficientemente espressive da esprimere in una formula la consistenza di un concetto risolutivo visto come "raccomandazione", che nell'interpretazione normativa formalizza la nozione di "esito razionale".
Ad ogni gioco in forma strategica viene associato un modello relazionale su cui si interpreta un linguaggio le cui formule esprimono la struttura strategica dei giochi. Risulta immediato osservare in questo caso l'equivalenza tra la formula che esprime la consistenza e la definizione di equilibrio di Nash.

Il caso dei giochi in forma estesa è più complesso: ad ogni gioco si associa un modello della teoria **BTA** con un linguaggio temporale sufficientemente espressivo a rappresentare la struttura sequenziale del gioco. Si estende poi il modello con l'aggiunta di una raccomandazione come una seconda relazione temporale, ottenendo una modello che diremo modello normativo del gioco. Un modello normativo è un modello di $\mathbf{BTA_p}$, una teoria che si dimostra completa rispetto alle strutture temporali ad albero con previsione.
In questo linguaggio viene espressa la nozione di consistenza interna per una previsione, e viene dimostrato che gli unici modelli normativi soddisfacenti la formula di consistenza interna sono quelli associati all'induzione a ritroso, dove la raccomandazione consiste in un equilibrio perfetto nei sottogiochi.

Capitolo 1

Teoria dei giochi e logica modale: presentazioni

1.1 Teoria dei giochi

Lo scopo della teoria dei giochi è di fornire un linguaggio adatto a descrivere situazioni di interazione tra individui razionali, e fornire adeguati strumenti matematici per l'analisi di questi problemi. Scopo di questo primo capitolo è dare un senso e una definizione ben precisa alle parole "razionali" ed "interazione"; per valutare con precisione i limiti e la forza di un modello dalle ambizioni così generali sono necessarie definizioni molto chiare.
I modelli della teoria dei giochi sono rappresentazioni molto astratte di classi di situazioni "reali": sebbene la teoria si sia evoluta ad una teoria matematica a sé stante, la fonte principale di ispirazione rimane nelle applicazioni.

In questa tesi ci preoccuperemo di esplicitare nella maniera più chiara possibile tutte le ipotesi assunte dall' "ambiente di lavoro" della teoria dei giochi. Per poter far ciò, e per poter chiarire e distinguere gli scopi della teoria, è utile chiarire la sua posizione all'interno delle varie correnti del pensiero scientifico.
La teoria dei giochi si propone di fondare una teoria della decisione in condizioni di interazione, fondata sulla teoria della decisione (semplice), che verrà accennata nella prossima sezione. Analogamente ad essa, i risultati della teoria dei giochi possono essere considerati sotto due aspetti differenti. Il primo, l'approccio *descrittivo*, vede la teoria rilevare e descrivere le regolarità del comportamento, costruendo un modello consistente con ciò che è generalmente osservabile.
L'approccio *normativo* considera invece la teoria come una raccomandazione, che indica ad un individuo che soddisfi certe ipotesi di razionalità il modo "ideale" di comportarsi. Se con RAT indichiamo le ipotesi concernenti la

razionalità e con ACT delle azioni, i tipici risultati descrittivi ricercano le condizioni necessarie per ottenere (dunque descrivere) certe azioni:

$$RAT \Leftarrow ACT;$$

mentre teoremi che esplicitano le condizioni di razionalità sufficienti ad implicare un determinato comportamento sono tipicamente interpretate dall'approccio normativo

$$RAT \Rightarrow ACT.$$

I due approcci non sono in alcun modo slegati nè incompatibili; il loro continuo intreccio permette anzi di ottenere una visione completa della teoria in oggetto. I risultati che si ottengono hanno comunemente la forma

$$RAT \Leftrightarrow ACT$$

dove entrambe le frecce saranno passibili di analisi sotto entrambi gli aspetti descrittivo e normativo.
Cercheremo di evitare la confusione generata dall'intrecciarsi delle due interpretazioni soffermandoci spesso ad analizzare sotto entrambe le interpretazioni i risultati che verranno enunciati.

In teoria dei giochi le terminologie variano leggermente, ma la confusione resta[1]. Scopo dichiarato in [OR94] è infatti la massima chiarezza tra le due interpretazioni, denominate rispettivamente "steady state" e "deductive interpretation".
La prima è un'interpretazione standard in economia, che vede la teoria dei giochi modellare le regolarità di interazione che si osservano nella vita reale. Questo è proprio l'approccio descrittivo, in cui un concetto risolutivo come l'equilibrio di Nash riassume le regolarità ("steady states") osservate in una medesima classe di situazioni di interazione.
In particolare si suppone che ogni giocatore deduca il comportamento degli altri giocatori dall'osservazione e dalla comprensione di queste regolarità data dall'esperienza. Si prenda ad esempio la nozione di equilibrio: la lettura descrittiva vede l'equilibrio come una descrizione del comportamento osservabile di individui razionali, che deducendo dall'esperienza la stabilità o l'instabilità dei diversi esiti possibili, tendono a preferire situazioni di maggior stabilità.
L'approccio normativo è invece impersonato dall'interpretazione deduttiva

[1]Citiamo ad esempio il seguente brano di B.de Finetti: "...qui è un po' difficile vedere se affiori la consueta confusione e commistione tra punto di vista descrittivo e normativo nell'economia, o si tratti semplicemente di accantonamento di difficoltà insolute correttamente individuate;...la deprecata commistione di descrittivo e normativo, nelle incrostazioni da cui il subcosciente del pensiero economico appare ancora irretito, è probabilmente una derivazione di tre concetti...", tratto da [dF63b].

della teoria dei giochi: un gioco è considerato nella sua astrattezza come un evento singolo (one-shot), e la teoria come la raccomandazione teorica per giocatori razionali. Ogni giocatore deduce il comportamento degli altri dall'ipotesi che anch'essi siano conformi agli stessi assunti di razionalità e che dunque seguano la stessa raccomandazione. Vedremo come la stabilità dell'equilibrio è legata ad opportune condizioni di consistenza per la raccomandazione.

Questi concetti faranno la loro comparsa in quasi tutti i paragrafi successivi, e rappresentano una delle principali motivazioni a supporto dell'utilizzo di strumenti di logica all'interno dei fondamenti della teoria dei giochi.

1.1.1 Comportamento razionale: breve introduzione alla teoria della decisione

Prima di studiare il modello di decisione in situazioni complesse proposto dalla teoria dei giochi, è fondamentale stabilire un primo contatto con il concetto di "giocatore razionale" descrivendone il comportamento in situazioni di decisione semplice, quando l'esito dipende solamente dalle scelte del singolo giocatore.
I risultati che verranno esposti in seguito seguiranno indicativamente il medesimo schema:

- definizione di un insieme su cui il giocatore esprime preferenze;
- presentazione di una lista di assiomi che descrivano il comportamento di (o siano ammissibili per) un giocatore razionale;
- dimostrazione di un teorema di rappresentazione, che faccia discendere dagli assiomi una rappresentazione numerica (la terminologia sarà "una rappresentazione in termini di utilità") delle preferenze espresse su X.

Farà eccezione il teorema di rivelazione delle preferenze, dove l'intreccio di approccio normativo e descrittivo farà la sua prima comparsa.
Ci restringeremo sempre al caso di insiemi numerabili, in certi casi anche finiti, da un lato per semplificare la trattazione e dall'altro poiché in seguito non avremo bisogno di risultati più generali. Quasi tutti i risultati esposti si possono comunque generalizzare in diverse direzioni. Una trattazione completa è reperibile su [Fis70], da cui sono presi la maggior parte dei risultati che seguono.

Relazioni di Preferenza

Sia X un insieme non vuoto, e sia $\prec$ una relazione binaria su X. Scriviamo $x \not\prec y$ se non vale $x \prec y$. L'approccio classico alla rappresentazione delle preferenze consiste nel supporre che le preferenze di un individuo siano conformi alla seguente definizione:

Definizione 1.1. Una relazione binaria $\prec$ si dice *relazione di preferenza* se:

i) è antisimmetrica: $x \prec y \Rightarrow y \not\prec x$;
ii) è negativamente transitiva: $x \not\prec y$ e $y \not\prec z \Rightarrow x \not\prec z$.

Dimostriamo alcune semplici proprietà:

Lemma 1.2. *Una relazione binaria $\prec$ è negativamente transitiva se e solo se $x \prec y \Rightarrow \forall z\ x \prec z$ o $z \prec y$ (o entrambi).*

Proposizione 1.3. *Una relazione di preferenza $\prec$ è transitiva, irriflessiva ed aciclica (se $x_1 \prec x_2 \prec \cdots \prec x_{n-1} \prec x_n$ allora $x_1 \neq x_n$).*

Dimostrazione. Il lemma è di facile dimostrazione, visto che $x \prec y \Rightarrow \forall z\ x \prec z$ o $z \prec y$ è la contronominale della definizione di negativamente transitiva: $x \not\prec y$ e $y \not\prec z \Rightarrow x \not\prec z$.
Per dimostrare la transitività basta notare che se $x \prec y$ ed $y \prec z$, usando il Lemma 1.2 per z stesso si ottiene che $x \prec z$ o $z \prec y$ ma dato che per ipotesi $y \prec z$ e la $\prec$ è antisimmetrica, ne segue che $x \prec z$.
Per antisimmetria se $x \prec x$ allora $x \not\prec x$, dunque $\prec$ è irriflessiva.
Supponiamo che esistano $x_1 \dots x_n$ tali che $x_1 \prec x_2 \dots x_{n-1} \prec x_n$ ed $x_1 = x_n$: per transività otterremmo $x_1 \prec x_n = x_1$. Assurdo per irriflessività, dunque $\prec$ è aciclica. □

Dimostriamo ora che esprimere una relazione di preferenza su X equivale ad esprimere un ordine lineare sui suoi elementi.
Definendo $x \sim y \Leftrightarrow x \not\prec y$ e $y \not\prec x$, e $x \preceq y \Leftrightarrow y \not\prec x$.

Proposizione 1.4. *Valgono le seguenti proprietà:*
i) $\sim$ è una relazione di equivalenza;
ii) $\preceq$ è un ordine lineare;
iii) $x \preceq y \Leftrightarrow x \prec y$ o $x \sim y$.

Dimostrazione. $\sim$ è simmetrica per definizione ed è riflessiva per irriflessività di $\prec$. Se inoltre $x \sim y$ ed $y \sim z$ allora dalla definizione si ha che $x \not\prec y$ e $y \not\prec z$. Per negativa transitività $x \not\prec z$. Allo stesso modo da $y \not\prec x$ e $z \not\prec y$ si ottiene $z \not\prec x$ e dunque $x \sim z$.
Presi x ed y allora o $x \prec y$ allora per antisimmetria $y \not\prec x$ dunque $x \preceq y$, oppure $x \not\prec y$ e allora $y \preceq x$: $\preceq$ è completa. La transitività di $\preceq$ è esattamente la negativa transitività di $\prec$.
Dimostriamo ora che $x \preceq y$ ed $y \preceq x \Rightarrow x \sim y$: dalle definizioni ottengo che $y \not\prec x$ e $x \not\prec y$, che è proprio la definizione di $x \sim y$. Questo dimostra anche la terza proprietà. □

Partendo all'inverso da un ordine lineare $\preceq$ su X, e definendo $x \prec' y \Leftrightarrow y \not\preceq x$ e $x \sim' y \Leftrightarrow x \preceq y \wedge y \preceq x$, si ottiene una relazione di preferenza.

Definendo un ordine totale $\preceq'$ da $\prec'$ col procedimento appena esposto si ottiene $\preceq'=\preceq$.

Dunque esprimere preferenze strette antisimmetriche e negativamente transitive su X è equivalente ad esprimere un ordine totale su X. In generale si preferisce definire una relazione di preferenza nel primo modo (detto anche ordine debole), dando alla relazione $x \sim y$, che esprime la mancanza di preferenza, le seguenti interpretazioni:

- x ed y sono inconfrontabili;
- x è egualmente preferito a y;
- l'individuo è incerto se $x \prec y$ oppure $y \prec x$.

Questa abbondanza di significati genera diversi problemi legati soprattutto alla transitività di $\sim$: se x ed y sono inconfrontabili ed y è egualmente preferibile a z, non è chiaro quale conclusione si debba trarre dal fatto che per transitività $x \sim z$.
Un'altra problematica nasce dal tentativo di giustificare le proprietà di una relazione di preferenza (che ricordiamo sono le assunzioni di razionalità sui giocatori). Se da un punto di vista normativo esse sono piuttosto ragionevoli (si può facilmente concordare nel voler esprimere preferenze antisimmetriche, dunque non contradditorie, e negativamente transitive) proprio la negativa transitività crea grossi problemi ad una lettura descrittiva del modello.
Si pensi infatti al seguente esempio: l'individuo esprime preferenze su $X = \{990€, 995€, 1000€\}$, è abbastanza ragionevole pensare che sebbene 990 sia preferito a 995 non ci sia chiara preferenza tra 900 e 1000 (dunque $990 \sim 1000$), nè tra 995 e 1000. Da queste preferenze si ottiene che $990 \prec 995$ senza che nè $990 \prec 1000$ nè $1000 \prec 995$, in contraddizione con il Lemma 1.2.
Per ovviare a queste difficoltà si possono definire le relazioni di preferenza come ordini stretti parziali, definendo due diversi concetti di eguaglianza, uno per l'equivalenza di preferenza ed un altro per l'inconfrontabilità. Su [Fis70] vengono riportati alcuni risultati in questo ambito.

Rivelazione delle preferenze

A supporto della definizione di relazioni di preferenza come data al paragrafo precedente si può dimostrare come essa possa essere ricava tramite un approccio totalmente descrittivo, dove l'ingrediente base sono le scelte dell'individuo. Non sono infatti le preferenze di un individuo a ricadere nel dominio dell'osservabile, ma piuttosto le sue scelte.

Sia dunque X l'insieme delle alternative, possiamo modellizzare la scelta come una funzione $c : \mathscr{P}(X) \longrightarrow \mathscr{P}(X)$ dalle parti di X in sé, tale che

$c(B) \neq \emptyset$ e $c(B) \subseteq B$ per ogni $B \subseteq X$. La funzione c indica per ogni sottoinsieme B le alternative scelte (dunque preferite) dall'individuo. Nel caso di spazi numerabili considereremo la funzione c sulle parti finite di X. Presa una relazione binaria $\prec$ su X, si può definire una funzione di scelta $c_\prec$ come $c_\prec(B) = \{x \in B \mid \forall y \; x \not\prec y\}$.

Le condizioni che vengono imposte a c per essere una "buona" funzione di scelta sono le seguenti proprietà α *e* β *di Sen*[2]:

α: $x \in B \subseteq D$ e $x \in c(D)$ allora $x \in c(B)$
la spiegazione informale di Sen: "se il campione del mondo di un certo sport è pachistano allora è anche il campione del Pakistan";

β: $A \subseteq B$ e $x, y \in c(A)$ e $y \in c(B)$ allora anche $x \in c(B)$;
la spiegazione informale di Sen: "se il campione del mondo è pachistano tutti i campioni del Pakistan sono campioni del mondo".

L'approccio normativo sarà interessato a sapere sotto quali condizioni $c_\prec$ è una funzione di scelta (la risposta è se $\prec$ è aciclica). Seguendo invece l'approccio descrittivo, cerchiamo le condizioni necessarie perché una relazione $\prec$ sia tale che $c_\prec = c$.
Dimostriamo dunque il teorema detto di rivelazione delle preferenze:

Teorema 1.5 ([Kre88])**.** *Se $\prec$ è una relazione di preferenza allora $c_\prec$ è una funzione di scelta che verifica le proprietà α e β di Sen.*
Viceversa, se c è una funzione di scelta che verifica le proprietà α e β allora esiste una relazione di preferenza $\prec$ tale che $c = c_\prec$.

Dimostrazione. Supponiamo che c sia una "buona" funzione di scelta e definiamo $x \prec y \Leftrightarrow c(\{x, y\}) = \{y\}$; dimostriamo che $\prec$ è una relazione di preferenza.
Innanzitutto è antisimmetrica, infatti se $c(\{x, y\}) = \{y\}$, quindi $x \prec y$, non può anche essere il caso che $y \prec x$, altrimenti $c(\{y, x\}) = \{x\}$ che però per ipotesi è uguale a $\{y\}$.
Per mostrare che è negativamente transitiva supponiamo che $x \not\prec y$ e $y \not\prec z$, dunque che $c(\{x, y\}) \neq \{y\}$ e $c(\{y, z\}) \neq \{z\}$. Se per assurdo fosse $x \prec z$ allora $c(\{x, z\}) = \{z\}$, e allora $x \notin c(\{x, y, z\})$, altrimenti per la proprietà α si avrebbe che $x \in c(\{x, z\})$.
Dato che $c(\{x, y\}) \neq \emptyset$ e $c(\{x, y\}) \neq \{y\}$ allora $x \in c(\{x, y\})$. Se ora fosse $y \in c(\{x, y, z\})$ otterrei per la proprietà α che $c(\{x, y\}) = \{x, y\}$, e per la proprietà β che $x \in c(\{x, y, z\})$, possibilità già scartata in precedenza.
Dunque $x \notin c(\{x, y, z\})$, $y \notin c(\{x, y, z\})$, e dal fatto che $y \in c(\{y, z\})$ ottengo allo stesso modo che $z \notin c(\{x, y, z\})$: assurdo poiché $c(\{x, y, z\})$ deve

[2]Seguendo la notazione di [Kre88].

essere non vuoto.
Per mostrare che $c = c_{\prec}$, mostriamo che se $x \in c(A)$, allora per ogni $z \in A$ si ha che $x \not\prec z$. Se infatti esistesse un tale z, allora $c(\{x, z\}) = \{z\}$ contraddicendo la proprietà α.
Supponiamo d'altra parte che $x \notin c(A)$, e sia $z \in c(A)$. Per la proprietà β, $c(\{x, z\}) = \{z\}$, altrimenti x dovrebbe appartenere a $c(A)$. Dunque $x \prec z$ ed $x \notin c_{\prec}(A)$.

Supponiamo ora che $\prec$ sia una relazione di preferenza. Che $c_{\prec}(B) \subseteq B$ è ovvio. Se supponiamo che $c_{\prec}(B) = \emptyset$, allora per ogni x esiste z a lui preferito. Posso costruire allora una catena $x \prec y \dots$ che per la finitezza di B si deve chiudere in un ciclo: assurdo poiché una relazione di preferenza è aciclica.
Vediamo ora le due proprietà: α: se $x \in B \subseteq D$ e $x \in c_{\prec}(D)$ allora $\forall y \in D\ x \not\prec y$ dunque anche $\forall y \in B\ x \not\prec y$. Dunque $x \in c_{\prec}(B)$.
β: sia $B \subseteq D$, $x, y \in c_{\prec}(B)$ e $x \in c_{\prec}(D)$ allora in particolare $x \not\prec y$ e $y \not\prec x$ e $\forall z \in D\ x \not\prec z$; per negativa transitività ottengo che $\forall z \in D\ y \not\prec z$ e dunque che $y \in c_{\prec}(D)$. □

L'ipotesi di razionalità "il giocatore razionale massimizza la propria utilità" è espressa nella sua forma più generale dal teorema precedente: la scelta di un individuo tra un certo insieme di alternative è rappresentata (o indicata) dai punti di massimo della sua relazione di preferenza.

Funzioni di utilità

Dimostriamo il primo risultato di rappresentazione numerica delle preferenze, tramite una funzione $u : X \longrightarrow \mathbb{R}$ che viene detta funzione di utilità:

Teorema 1.6. *Se $\prec$ è una relazione di preferenza su X allora esiste una funzione $u : X \longrightarrow \mathbb{R}$ tale che*

$$x \prec y \ \Leftrightarrow\ u(x) < u(y) \qquad \forall x, y \in X$$

Inoltre tale u è unica a meno di cambiamenti di variabile strettamente crescenti.

Dimostrazione. Sia $\sim$ definita come ai paragrafi precedenti, e si consideri su $X/_{\sim}$ la relazione $x \prec' y$ sse esiste un $x \in \bar{x}$ e un $y \in \bar{y}$ tali che $x \prec y$ (dove con $\bar{x}$ si intende la classe di x modulo $\sim$). $\prec'$ è un ordine stretto.
Sia ora $x_0, x_1 \dots x_n \dots$ una numerazione di $X/\sim$, definiamo u per induzione su n:

- $u(x_0) = 0$

- supponendo $u(x_i)$ definite per $i < n$ si dividono i seguenti casi:

  se $x_n \prec' x_i \ \forall i < n$ allora $u(x_n) = -n$;

  se $x_i \prec' x_n \ \forall i < n$ allora $u(x_n) = n$;

  se esistono i, j tali che $x_i \prec' x_n \prec' x_j$ e per nessun $k < n$ si ha che $x_i \prec' x_k \prec' x_j$ allora $u(x_n) = \frac{u(x_j)-u(x_i)}{2}$.

Chiaramente $x \prec' y$ se e solo se $u(x) < u(y)$. Definiamo $u : X \longrightarrow \mathbb{R}$ come $u(x) = u(\bar{x})$ (con un lieve abuso di notazione la funzione si chiama sempre u), ottenendo così anche che $x \sim y \Leftrightarrow u(x) = u(y)$. □

Viceversa data una funzione $u : X \longrightarrow \mathbb{R}$ è chiaro che essa induce una relazione di preferenza su X.

Nella dimostrazione si è ottenuta in realtà una funzione a valori in $\mathbb{Q}$; l'utilizzo di funzioni di utilità reali è legato al fatto che il teorema rimane valido per spazi non numerabili. In tal caso le ipotesi si complicano leggermente, e una naturale richiesta è che u sia continua; il teorema rimane valido per spazi metrici separabili e preferenze continue (in modo che le successioni di elementi di X rispettino la relazione), ottenendo funzioni u continue da X in $\mathbb{R}$ (si veda ad esempio [Fis70], Cap.3, in cui sono discusse diverse generalizzazioni).

Decisione in condizioni di rischio: teoria dell'utilità attesa di von Neumann-Morgenstern

Nel modello sinora esposto lo spazio X rappresenta uno spazio di alternative il cui esito è certo, tra le quali l'individuo compie la sua scelta. Fronteggiando situazioni in cui l'agente non è completamente informato, o in generale eventi dall'esito incerto, è necessario inserire una nozione di incertezza nel modello. Innanzitutto è fondamentale separare due diverse cause di incertezza (che danno luogo a due differenti modelli):

- l'*incertezza statistica*, in cui ciò che si conosce, su cui dunque si esprimono preferenze, sono distribuzioni di probabilità su un certo insieme di eventi. Per esempio, esprimere una preferenza tra le seguenti proposizioni: "un gelato se esce testa al lancio di una monetina o continuare a studiare se esce croce", "tirare un dado a quattro facce e mangiare tanti gelati quanto il numero che esce" e "eseguire il lancio precedente se in un turno di roulette esce un numero rosso, continuare a studiare se esce nero, tirare il dado a quattro facce se esce lo zero";

- l'*incertezza soggettiva*, in cui l'agente si trova a scegliere tra eventi probabilistici la cui probabilità è però dettata soltanto dalle proprie convinzioni. Per esempio una situazione in cui uscire a prendere un

gelato se piove può essere peggio che continuare a studiare (in una stanza senza finestre come l'aula Ciampa).

In questa sezione accenniamo al modello per l'incertezza statistica, ossia la teoria dell'utilità attesa di von Neumann-Morgenstern.

Sia dunque X il solito insieme di oggetti, esiti, conseguenze, alternative, a seconda della situazione. L'incertezza oggettiva viene modellizzata esprimendo preferenze (o presentando una funzione di scelta) sull'insieme $\overline{X}$ delle misure di probabilità su X. Secondo una terminologia usuale, l'agente esprime preferenze su "lotterie" con premi in X. Nel caso di X finito una tal definizione non crea problemi, mentre nel caso numerabile è necessario restringere $\overline{X}$ all'insieme delle probabilità semplici (concentrate su un numero finito di punti).
Il modello comprende il caso di eventi certi, identificando gli elementi $x \in X$ con la probabilità δ_x corrispondente, che dà probabilità 1 ad x e 0 a tutto il resto. $\overline{X}$ è inoltre chiuso per combinazione convessa, ossia se $\mu, \nu \in \overline{X}$ allora per ogni $a \in [0,1]$ $a\mu + (1-a)\nu$ è una probabilità in $\overline{X}$.

L'idea in questo caso è di rappresentare numericamente le preferenze $\prec$, espresse su $\overline{X}$, attraverso il valore atteso di una funzione di utilità su X. Le condizioni sufficienti per ottenere questo risultato sono detti assiomi di von Neumann-Morgenstern:

($A1$): $\prec$ è una relazione di preferenza;

($A2$): se $\mu, \nu \in \overline{X}$ e $a \in [0,1]$ allora
$\mu \prec \nu \Rightarrow a\mu + (1-a)\eta \prec a\nu + (1-a)\eta \quad \forall \eta \in \overline{X}$;

($A3$): $\mu \prec \nu \prec \eta \in \overline{X} \;\Rightarrow\; \exists\, a, b \in [0,1]$ tale che
$a\mu + (1-a)\eta \prec \nu \prec b\mu + (1-b)\eta$.

Il secondo assioma è detto assioma di indipendenza o di sostituzione poiché permette di valutare lotterie complesse scorporandole nelle singole componenti. Meno ragionevole può sembrare il terzo, che implica la non esistenza di superpremi o superlotterie infinitamente preferibili (nel senso che qualsiasi combinazione convessa di una superlotteria con un altra, anche a pesi infinatamente bassi, rimane strettamente preferibile ad ogni altra lotteria). Questa assunzione si presta a diversi attacchi (tra i quali il più noto è il paradosso di Allais, [All53]), ma è dal punto di vista normativo che questo modello esprime tutta la sua potenza. Enunciamo infatti il seguente teorema[3]:

Teorema 1.7. *Sia $\prec$ una relazione binaria su $\overline{X}$, $\prec$ verifica gli assiomi* A1, A2, A3 *se e solo se esiste una funzione $u : X \longrightarrow \mathbb{R}$ tale che*

$$\mu \prec \nu \Leftrightarrow \mathbb{E}_\mu[u] < \mathbb{E}_\nu[u] \quad \forall \mu, \nu \in \overline{X}$$

[3]Per una dimostrazione e diverse generalizzazioni rimandiamo sempre a [Fis70].

u è unica a meno di trasfomazioni affini positive.

Dove con $\mathbb{E}_\mu[u]$ si intende il valore atteso di u sotto la probabilità μ, e con trasformazione affine positiva una funzione della forma $f(x) = ax + c$ con $a > 0$.
Dal punto di vista normativo è sufficiente accettare gli assiomi di von Neumann-Morgenstern ed esprimere preferenze sugli esiti per ottenere grandi semplificazioni nel calcolo delle preferenze sulle lotterie. Riprendiamo infatti l'esempio di inizio paragrafo: supponendo che piú gelati siano meglio di uno e che andare a mangiare un gelato (G) sia preferibile che continuare a studiare (S), si ottiene $S \prec G \prec 2G \prec 3G \prec 4G$; per il teorema di rappresentazione dell'utilità possiamo scegliere $u(S) = 0$, $u(G) = 4$, $u(2G) = 8$, $u(3G) = 11$, $u(4G) = 14$. Concordando con gli assiomi di vN-M è dunque sufficiente calcolare il valore atteso dei 3 eventi in considerazione per "capire" le preferenze sugli eventi incerti:

i) nel caso della moneta la probabilità che esca testa è $\frac{1}{2}$ dunque $\mathbb{E}_1 = \frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 0 = 2$;

ii) nel caso del dado $E_2 = \frac{1}{4} \cdot 4 + \frac{1}{4} \cdot 8 + \frac{1}{4} \cdot 11 + \frac{1}{4} \cdot 14 = \frac{37}{4}$;

iii) nel caso della roulette siccome i numeri sono 37 e lo zero non è nè rosso nè nero: $E_3 = \frac{18}{37} \cdot E_1 + \frac{18}{37} \cdot 0 + \frac{1}{37} \cdot E_2 = \frac{36}{37} + \frac{1}{4} < 2 = E_1$

Quindi il complicato meccanismo associato alla roulette non è preferibile al lancio della moneta per decidere se continuare a studiare.

Decisione in condizioni di incertezza: teoria di Savage

La situazione a cui ci si riconduce[4] per modellizzare situazioni di decisione in condizioni di incertezza soggettiva, dipendente solamente dalle convinzioni del giocatore, è composta da:

- un insieme C (lo spazio delle conseguenze);
- un insieme Ω di "stati del mondo". Il "mondo" è, secondo le parole di Savage in [Sav54], "the object about which the person is concerned", ed ogni stato è "a description of the world, leaving no relevant aspect undescribed". Gli stati incorporano tutte le informazioni rilevanti per il problema di decisione su cui l'individuo è incerto. Gli stati devono essere formulati in modo tale che siano mutualmente esclusivi, in modo che l'agente sappia di essere in uno e un unico stato, pur senza sapere quale realmente "accade". Inoltre, non devono in alcun modo incorporare o dipendere dalla scelta dell'individuo.

[4]Segue una trattazione schematica della teoria di Savage di decisione in condizioni di incertezza soggettiva, come esposta in [Kre88]; per le dimostrazioni ed una trattazione più completa si rimanda a [Fis70] o direttamente a [Sav54].

A partire da questi due oggetti possiamo definire l'insieme delle azioni X come l'insieme delle funzioni dagli stati di natura nell'insieme C delle conseguenze: $X = C^{\Omega}$. L'idea è che la conseguenza di una determinata azione non è nota all'agente e dipende dallo stato di natura.
La scelta di un'azione, all'interno di un sottoinsieme di azioni possibili, determina dunque una funzione consequenziale dall'insieme degli stati di natura nell'insieme C. L'agente sceglie sull'insieme delle azioni X, dunque la relazione di preferenza sulle azioni è una relazione di preferenza $\prec$ su X.

È possibile dimostrare un risultato di rappresentazione nella forma seguente, come analogo "soggettivo" del Teorema 1.7:

sotto certe condizioni [5] *esiste un unica misura di probabilità* μ *su* Ω *e una funzione* $u : C \longrightarrow \mathbb{R}$ *tale che*

$$\forall f, g \in X \quad f \prec g \Leftrightarrow \mathbb{E}_{\mu \circ f^{-1}}[u] < \mathbb{E}_{\mu \circ g^{-1}}[u]$$

dove con $\mathbb{E}$ è indicato il valore atteso su C e con $\mu \circ f^{-1}$ è indicata la probabilità immagine di μ secondo f. Al solito u è unica a meno di funzioni affini positive.

L'importanza di questo risultato risiede nel fatto che sia la funzione di utilità u, sia la probabilità μ sono totalmente soggettive, dipendono unicamente dalla relazione di preferenza $\prec$ (che dunque "rivela" le convinzioni dell'agente sugli stati di natura, oltre che le preferenze sulle conseguenze in C).

Vediamo un esempio: un individuo si sta organizzando per un pic-nic l'indomani, e deve scegliere di portare alcuni oggetti tra i seguenti: pantaloni corti, giacca a vento, macchina fotografica, frisbee, pallone, ombrello. Non essendo in grado di consultare le previsioni del tempo, gli stati di natura sono (e proprio di natura si parla):

$$\Omega = \{\text{☀}, \text{⛈}, \text{☁}\}$$

Se consideriamo l'insieme degli oggetti da portare come l'insieme delle conseguenze (la conseguenza di portare i pantaloni corti è di avere i pantaloni corti) non risulta ben chiaro il motivo per cui si debba considerare tutto l'insieme delle funzioni da Ω in C come insieme delle azioni. Il risultato di rappresentazione vale per tutto l'insieme C^{Ω}, ma possiamo restringerci e considerare solamente le azioni costanti, che sono in effetti le azioni realmente attuabili.
A questo punto il lato descrittivo della rappresentazione di Savage afferma che esprimendo le preferenze sull'insieme delle azioni, l'agente ha in realtà in mente (e rivela attraverso le preferenze) una misura di convinzione sugli stati di natura e una funzione di utilità sull'insieme delle conseguenze: se

[5] Gli assiomi di Savage. Cfr. per esempio [Fis70], cap.14.

l'agente preferisce portare l'ombrello a portare un frisbee, significa che ritiene più probabile che piova; se preferisce la macchina fotografica alla giacca a vento crede più probabile che ci sia sole o nuvolo, oppure ama tanto fotografare che l'importanza del tempo è minima.
Dal punto di vista normativo invece, se l'agente "aderisce" agli assiomi di Savage, grazie al teorema di rappresentazione gli sarà sufficiente esprimere la proprie preferenze sulle conseguenze (per esempio rappresentando le conseguenze come un sottoinsieme di $\mathbb{R}^2$: la prima componente indicante il livello di divertimento, la seconda il livello di asciuttezza associati a un azione costante e ai diversi stati di natura) ed indicare le proprie convinzioni sugli stati di natura. Col calcolo dell'utilità attesa si ottiene la raccomandazione desiderata su quale è l'azione "ideale" da compiere.

1.1.2 Ipotesi della teoria dei giochi

Possiamo ora esaminare in dettaglio le due ipotesi su cui si fonda la teoria dei giochi:

- i giocatori sono *razionali*;
- i giocatori *ragionano strategicamente*: nel fare la loro scelta ogni giocatore tiene in considerazione la propria conoscenza o le proprie convinzioni rispetto alle scelte degli altri giocatori.

Che un giocatore sia razionale significa che[6] "he is aware of his alternatives, forms expectations about any unknowns, has clear preferences, and chooses his action deliberately after some process of optimization": tutte espressioni formalizzabili nella teoria della decisione esposta ai paragrafi precedenti[7]. Dunque un giocatore è razionale si è conforme alla teoria della decisione (semplice).

In assenza di incertezza il problema di decisione che il giocatore razionale fronteggia è dato da:

- un insieme A di azioni a disposizione del giocatore;
- un insieme C di conseguenze
- una mappa consequenziale $g : A \longrightarrow C$, che rappresenta le conseguenze delle azioni in A;
- una relazione di preferenza $\prec_i$ su C, o indifferentemente per il Teorema 1.6, una funzione di utilità $u_i : C \longrightarrow \mathbb{R}$.

Il problema di decisione per il giocatore consiste nella scelta di quelle azioni che risultano preferite dato un sottoinsieme $B \subseteq A$ di azioni realizzabili. Per

[6] [OR94], pag.4.

[7] Da un punto di vista della teoria dei giochi come teoria descrittiva questo significa che essa rimane vulnerabile agli attacchi portati alla teoria della decisione (principalmente da psicologi ed economisti comportamentali, che ne sottolineano giustamente i limiti).

la teoria della decisione la funzione di scelta "razionale" associa al sottoinsieme $B \subseteq A$ il sottoinsieme $c(B) = \{x \in B \mid \forall y \; x \not\prec y\}$. Per il Teorema 1.6 questo è equivalente a risolvere il problema di massimo per $b \in B$ di $u(g(b))$; se con "argmax" indichiamo il punto di massimo della funzione u[8] possiamo scrivere:

$$c(B) = \arg\max_{b \in B} u(g(b))$$

che, come già sottolineato, esprime il concetto "il giocatore razionale massimizza la propria utilità".

Le condizioni di incertezza possono derivare da diversi fattori. Per esempio il giocatore può essere incerto sull'esito di azioni degli altri giocatori che non siano deterministiche, o decidere di giocare egli stesso una strategia non deterministica; può non essere certo della razionalità degli altri giocatori, oppure non essere a conoscenza delle utilità degli altri giocatori o non avere una conoscenza perfetta delle "regole del gioco".
Come suggerito dagli esempi le teorie di decisione in condizioni di incertezza su cui si fonda la teoria dei giochi sono quella di von Neumann-Morgenstern e quella di Savage.
Nel primo caso, in condizioni di incertezza dovuta ad azioni e parametri non deterministici ma oggettivi, la funzione consequenziale g associa ad ogni azione del giocatore in A una distribuzione di probabilità sull'insieme delle conseguenze, e scriveremo $g : A \longrightarrow \overline{C}$. Per il Teorema 1.7 le preferenze del giocatore razionale sulle lotterie su C sono espresse dal valore atteso di una funzione di utilità su C.
Dunque i dati del problema rimangono invariati e la frase "il giocatore razionale massimizza la propria utilità attesa" è espressa dal fatto che dato un sottoinsieme $B \subseteq A$, la scelta del giocatore risolve il problema di massimo dell'utilità attesa sulle lotterie associate a $b \in B$:

$$c(B) = \arg\max_{b \in B} \mathbb{E}_{g(b)}[u]$$

dove con $\mathbb{E}_{g(b)}$ si intende il valore atteso di u su C con misura di probabilità $g(b)$.
Nel caso in cui invece l'incertezza dipenda dalla mancanza di informazione del giocatore, e sia dunque legata alle sue convinzioni riguardo certi determinati parametri, seguendo la teoria di Savage si suppone che il giocatore rappresenti le proprie convinzioni tramite una misura di probabilità μ su un insieme di stati di natura Ω, che elenchi tutte le possibilità date dai parametri non noti. Sull'insieme C delle conseguenze si suppone siano espresse le preferenze (o la funzione di utilità) del giocatore e che egli abbia in mente una mappa consequenziale $g : A \times \Omega \longrightarrow C$. La mappa g associa ad ogni

[8] Stiamo sempre considerando il caso di insiemi finiti o numerabili; in ogni caso B è finito dunque il massimo esiste. In casi più generali si richiedono ipotesi più forti su u e su B.

azione in A l'esito corrispondente a seconda dello stato di natura.
Nella notazione ai paragrafi precedenti, stiamo supponendo che il giocatore restringa l'insieme di tutte le azioni ad un insieme di azioni attuabili $A \subseteq C^{\Omega}$, e che dunque la funzione consequenziale $g(a, w) = a(w)$.
In questo modo l'ipotesi di comportamento razionale consiste nello scegliere, dato un sottoinsieme B di azioni in A, quelle azioni che massimizzino l'utilità attesa della funzione di utilità u secondo la probabilità immagine delle convinzioni μ secondo la mappa consequenziale $g(a, -) = a : \Omega \longrightarrow C$:

$$c(B) = \arg\max_{b \in B} \mathbb{E}_{\mu \circ g(a,-)^{-1}}[u].$$

Questa è una trattazione esaustiva dell'ipotesi di razionalità alla base della teoria dei giochi. Nel seguito della tesi ci concentreremo solamente sul primo caso, in cui l'ipotesi di razionalità è la massimizzazione dell'utilità.
Nelle situazioni di interazione la conseguenza di un azione dipende dal comportamento degli altri giocatori. Dunque a differenza delle situazioni alle sezioni precedenti, in questo caso esprimere preferenze non porta ad una funzione di scelta.

1.1.3 Giochi in forma strategica

Un *gioco in forma strategica* è un modello di situazioni di interazione a una mossa, in cui i giocatori scelgono simultaneamente un'azione da compiere all'interno di dati insiemi di azioni ammissibili. Simultaneamente non è da intendersi in senso temporale: si intende che non trapeli informazione sulle scelte dei vari giocatori prima che tutti abbiano fatto la propria scelta. L'esito del gioco è completamente determinato una volta che ogni giocatore ha fatto la sua scelta.

Definizione 1.8. Definiamo *gioco in forma strategica* $G = \langle N, A_i, \prec_i \rangle$:

- un insieme N di giocatori;
- un insieme non vuoto A_i per ogni $i \in N$;
- una relazione di preferenza $\prec_i$ sull'insieme $\prod_i A_i$ per ogni $i \in N$ (le preferenze del giocatore i-esimo sull'insieme degli esiti).

Per il Teorema 1.6 le relazioni di preferenza possono equivalentemente essere espresse da funzioni di utilità $u_i : \prod_{i \in N} A_i \longrightarrow \mathbb{R}$.

Diciamo profilo di strategie un elemento del prodotto cartesiano $\sigma \in \prod_i A_i$ per $i \in N$; l'insieme $\prod_i A_i$ dei profili di strategie viene detto insieme

degli esiti. Indicheremo con σ_i la strategia del giocatore i-esimo in σ, e con σ_{-i} le restanti strategie.
Come già sottolineato, la differenza tra un problema di decisione e un problema di teoria dei giochi risiede nel fatto che le preferenze di ogni giocatore non sono espresse sull'insieme delle proprie azioni, ma sull'insieme degli esiti, che dipende anche dalle azioni scelte dagli altri giocatori.

In certi casi è più conveniente inserire un insieme di esiti esterno C su cui esprimere le relazioni di preferenza $\prec_i$, e una mappa consequenziale $g : \prod_i A_i \longrightarrow C$. Le preferenze su $\prod_i A_i$ vengono riportate tramite g^{-1}.
Nel caso in cui ad ogni profilo di strategie sia associata una distribuzione di probabilità su un insieme di esiti C (quindi la funzione consequenziale $g : \prod_i A_i \longrightarrow \overline{C}$) verrà utilizzata la teoria dell'utilità attesa per riportare il gioco alla Definizione 1.8. Allo stesso modo se l'esito del gioco è incerto si utilizza la teoria di Savage per costruire uno spazio probabilizzato Ω di stati di natura e una funzione consequenziale $g : \Omega \times \prod_{i \in N} A_i \longrightarrow C$, per riportare le preferenze espresse su C all'insieme degli esiti.

Il più famoso e importante concetto risolutivo in teoria dei giochi è sicuramente l'equilibrio di Nash.
L'equilibrio di Nash, come introdotto in [Nas50], cattura una regolarità, uno "*steady state*", nelle situazioni di interazione tra decisori razionali:

Definizione 1.9. Un profilo di strategie σ^* è un *equilibrio di Nash* se per ogni $i \in N$:

$$\forall a_i \in A_i \;\; u_i(\sigma^*_{-i}, a_i) \leqslant u_i(\sigma^*)$$

dove con (σ^*_{-i}, a_i) si intende il profilo di strategie σ^* con la strategia a_i sostituita al posto di σ^*_i.

In un equilibrio di Nash nessun giocatore può cambiare strategia, fissate le azioni degli altri giocatori, senza ottenere un esito peggiore.
È utile considerare una formulazione equivalente della Definizione 1.9. Per ogni giocatore i, preso un profilo di strategie (degli altri) σ_{-i}, definiamo l'insieme

$$B(\sigma_{-i}) = \{\bar{a}_i \in A_i \mid u_i(\sigma_{-i}, \bar{a}_i) \geqslant u_i(\sigma_{-i}, a_i) \; \forall a_i \in A_i\}$$

come l'insieme delle "migliori risposte" di i alle azioni dei restanti giocatori. Un profilo di strategie σ^* è un equilibrio di Nash se e solo se per ogni i $\sigma^*_i \in B(\sigma^*_{-i})$.
Dunque se definiamo la funzione $G : \prod_i A_i \longrightarrow \mathscr{P}(\prod_i A_i)$, che ad ogni profilo di strategie associa il prodotto degli insiemi "miglior risposta" $G(\sigma) = \prod_i B(\sigma_{-i})$, otteniamo l'insieme degli equilibri di Nash come punto fisso di una funzione[9]:

[9] Questo è anche il metodo per la dimostrazione di esistenza (e di ricerca) di un equilibrio di Nash: si veda ad esempio [OR94], teorema di Kakutani.

σ^* è un equilibrio di Nash se e solo se $\sigma^* \in G(\sigma^*)$.

L'idea di equilibrio di Nash è incentrata sul concetto di incentivo: la possibilità di ottenere con una propria azione un esito migliore di quello conseguente un certo profilo, rappresenta un incentivo a spostarsi per decisori razionali che massimizzano le proprie preferenze, rendendo dunque il profilo instabile. Il carattere interattivo impone inoltre che nessun altro giocatore abbia incentivi a deviare dall'equilibrio, altrimenti si instaura una catena di riconsiderazioni e dunque di nuovi incentivi a cambiare strategia: nel concetto di equilibrio di Nash è riassunta l'assenza di incentivi a deviare. Ritorneremo su questo punto quando analizzeremo un concetto risolutivo sotto il punto di vista dell'interpretazione normativa.

Esempi

Esponiamo in questa sezione alcuni esempi significativi di giochi in forma strategica ed equilibri di Nash. Sono tutti esempi molto semplici (2 giocatori, 2 strategie ciascuno), ma catturano situazioni chiave che si ripetono sovente in situazioni più complesse.

BoS[10]: due amiche desiderano andare a un concerto di musica classica, e i due teatri della città questa sera danno Bach al teatro Verdi, e Stravinski al Sant'Andrea. La prima delle due preferirebbe ascoltare il concerto di Bach, mentre la seconda preferisce Stravinski. In ogni caso entrambe preferiscono andare insieme a teatro, piuttosto che ritrovarsi sole ad ascoltare il proprio concerto preferito.
I giocatori sono dunque due, e per entrambi le azioni effettuabili sono: B per "andare al teatro Verdi" ed S per "andare al Sant'Andrea". Le funzioni di utilità sono indicate nella tabella sottostante:

	B	S
B	(2,1)*	(0,0)
S	(0,0)	(1,2)*

Tabella 1.1: Bach o Stravinski.

nelle righe le azioni di 1 e nelle colonne le azioni di 2; la prima componente del vettore guadagno corrisponde al guadagno di 1, la seconda quello di 2.
Il gioco rappresenta una situazione in cui due giocatori con interessi contrastanti hanno necessità di cooperare, ed ha due equilibri di Nash (indicati in tabella con una *) (B,B) ed (S,S): per entrambi la prima coordinata è la massima a colonna fissata (miglior risposta di 1 a 2) e la seconda coordinata

[10]Classicamente denominato "Battaglia dei Sessi"; concordiamo con l'atteggiamento politicamente corretto di [OR94] nel ridefinirlo "Bach or Stravinski".

è il massimo a riga fissata (miglior risposta di 2 a 1). La domanda ovvia in questo caso è: come scegliere tra i due equilibri? Una risposta parziale si può ottenere con le strategie miste che verranno esposte in seguito.

Dilemma del Prigioniero: questo è forse il più famoso esempio di gioco in forma strategica a due giocatori, ed è riassunto nella tabella

	C	NC
C	(-3,-3)*	(0,-5)
NC	(-5,0)	(-1,-1)

Tabella 1.2: Dilemma del prigioniero.

Sono numerosissime le situazioni[11] in cui il dilemma del prigioniero si ripresenta, ma la classica (che giustifica il nome) è la seguente:

"Due persone sospettate di un crimine sono rinchiuse in due celle separate e separatamente vegono interrogate; entrambe hanno la possibilità di confessare o di mantenere il silenzio. Nel caso in cui entrambi confessino, il giudice accorderà uno sconto di pena dando ad entrambi 3 anni di prigione. Se invece nessuno dei due confessa verranno incriminati per qualche reato minore e resteranno in carcere per un solo anno. Nel caso in cui uno solo dei due confessasse, egli verrà liberato ed utilizzato a testimone contro l'altro che prenderà 5 anni di galera[12]"

Il gioco è simmetrico e qualsiasi cosa faccia un giocatore, l'altro preferisce Confessare a Non Confessare, dunque l'equilibrio di Nash in questo caso è (C,C). Il guagagno (si fa per dire) associato all'equilibrio è di 3 anni di galera a testa, quando avrebbero potuto scontarne uno soltanto non confessando entrambi.

Come già sottolineato in precedenza l'equilibrio di Nash cattura una regolarità, dunque nessun problema dal punto di vista descrittivo (sempre che i dati confermino).

Mozart o Mahler: di nuovo ad un concerto, ma questa volta le due ragazze concordano sulle preferenze ed entrambe preferiscono Mozart; al solito preferiscono andare nello stesso posto piuttosto che da sole.

Il gioco è detto *gioco di coordinazione* e ha due equilibri Nash: uno ovvio in cui entrambe vanno a vedere il loro artista preferito, ma anche l'altro in cui entrambe vanno a un concerto di Mahler[13].

[11] Per sottolineare la varietà di applicazioni della TdG (ma è solo un esempio): questo gioco rientra in una serie di "situazioni chiave" modellizzate come giochi strategici simmetrici, utilizzate dal matematico e psicologo Laszlo Mero per analizzare diverse definizioni di razionalità nel suo primo libro sull'argomento [Mer96].

[12] Due storie alternative e divertenti su [dF63a], pag.65.

[13] Questo esempio mostra come l'equilibrio di Nash non elimini situazioni in cui il vettore

	Mozart	Mahler
Mozart	(2,2)*	(0,0)
Mahler	(0,0)	(1,1)*

Tabella 1.3: Gioco di coordinazione.

Carta, forbice, sasso: le mosse sono tre e per entrambi gli esiti sono "vince, perde o pareggio". Il gioco viene detto *a somma zero* poiché il guadagno del secondo giocatore è l'opposto di quello del primo (se uno vince l'altro deve perdere, il pareggio vale 0). In tabella viene dunque indicato soltanto il guadagno del primo giocatore.

	Carta	Forbice	Sasso	
Carta	0	1	-1	$\bar{1}$
Forbice	1	0	-1	$\bar{1}$
Sasso	-1	1	0	$\bar{1}$
	$-\underline{1}$	$-\underline{1}$	$-\underline{1}$	

Tabella 1.4: Gioco a somma zero.

Questi giochi sono anche denominati giochi di competizione stretta, poiché gli interessi dei due giocatori sono completamente opposti. Sono i giochi introdotti e studiati per la prima volta da von Neumann e Morgenstern in [vNM47].
Cè un metodo generale per scoprire se un gioco a somma zero ha un equilibrio di Nash: consiste nel calcolare il massimo di ogni riga, ed evidenziarne il minimo, detto valore superiore del gioco (in Tabella 1.4 è $\bar{1}$); calcolare il minimo di ogni colonna ed evidenziarne il massimo (in questo caso $-\underline{1}$), detto valore inferiore del gioco. I due valori rappresentano il minimo che entrambi i giocatori si possono garantire: il gioco ha un equilibrio di Nash se questi due valori sono uguali, ed in ogni equilibrio il guadagno è dato da questo valore, detto valore del gioco.
In questo caso non c'è nessun equilibrio di Nash poiché il valore inferiore è -1 mentre quello superiore 1.

guadagno è inferiore per entrambi i giocatori. In termini tecnici essere equilibrio di Nash non implicha essere anche un ottimo paretiano, ed il dilemma del prigioniero mostra che essere un ottimo paretiano (C,C) non implica essere equilibrio di Nash (fatto piuttosto plausibile).

Strategie miste

A partire dall'equilibrio di Nash si possono definire molti altri concetti risolutivi, sia nel tentativo di risolvere alcuni dei problemi accennati negli esempi precedenti, sia nel tentativo di trovare un sostituto nei casi in cui non esista equilibrio di Nash.

A titolo di esempio introduciamo il più semplice dei concetti risolutivi che si basano sull'equilibrio di Nash, e ne diamo un'interessante interpretazione in termini di convinzioni.
Definiamo *estensione mista* di un gioco $G = \langle N, A_i, u_i \rangle$ il gioco $G' = \langle N, \overline{A}_i, \bar{u}_i \rangle$ dove gli $\overline{A}_i$ sono gli insiemi di tutte le distribuzioni di probabilità sugli A_i. Ogni giocatore sceglie un'azione in $\overline{A}_i$ dunque un profilo di strategie misto è una n-pla $(\bar{\alpha}_1, \ldots, \bar{\alpha}_n)$. Si dicono strategie pure quelle strategie miste che associano ad una azione in A_i probabilità 1 e 0 alle altre. Le funzioni $\bar{u}_i(\bar{\alpha})$ sono, per la teoria vN-M, l'utilità attesa di u_i sotto la probabilità data dal profilo di strategie misto $(\bar{\alpha}_1, \ldots, \bar{\alpha}_n)$.

Definizione 1.10. Un equilibrio in stategie miste per il gioco G è un equilibrio di Nash dell'estensione mista.

Il risultato fondamentale che si dimostra a proposito delle strategie miste è la seguente:

Teorema 1.11. *Ogni gioco finito G ha un equilibrio in strategie miste.*

Il Teorema è stato dimostrato da Nash in [Nas54] ed è un'applicazione del teorema del punto fisso di Kakutani.
Carta, Forbice Sasso: riprendendo l'esempio in tabella 1.4, vediamo qual'è l'equilibrio in strategie miste del gioco. Una strategia mista si indica con un vettore (p_1, p_2, p_3) che indica la probabilità di giocare ognuna delle strategie possibili. Un profilo di strategie miste che sia equilibrio di Nash necessita che ogni strategia sia la miglior risposta tra tutte le strategie miste di un giocatore. Per verificare ciò è sufficiente mostrare che il valore atteso (il guadagno con la strategia mista) è maggiore o uguale al guadagno che si otterrebbe giocando una qualsiasi strategia pura.
L'equilibrio di Nash in questo caso è per entrambi $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$: fissata questa strategia per uno dei due giocatori, l'altro è indifferente a giocare qualsiasi strategia pura che darà guadagno nullo, così come la strategia mista di equilibrio. Viceversa per l'altro giocatore.
Il fatto che il guadagno sia lo stesso per le strategie pure non implica nulla: se venissero giocate si romperebbe l'equilibrio poiché l'altro giocatore avrebbe un incentivo a spostarsi e come abbiamo visto il gioco in strategie pure non ha equilibrio.

Diverse sono le interpretazioni e conseguentemente le "giustificazioni" portate a favore dell'utilizzo di strategie miste. La seguente interessante interpretazione di carattere descrittivo dell'utilizzo delle strategie miste è anche

un buon esempio di quanto profondo (e complicato) sia l'uso delle convinzioni nei fondamenti della teoria dei giochi ([OR94], cap.3).
Si può vedere un equilibrio di Nash in strategie miste α^* come un "profilo di convinzioni", interpretando la distribuzione di probabilità α_i^* su A_i come la *convinzione comune*[14] degli *altri* giocatori, su quale azione sarà intrapresa da i. Sotto questa interpretazione il giocatore sceglie una singola azione piuttosto che una distribuzione di probabilità, scelta poco giustificabile se il gioco non viene ripetuto. L'equilibrio di Nash è dunque uno "steady state" delle convinzioni dei giocatori, non delle loro azioni.
I lati positivi di questo approccio sono molteplici, ma sopravviene la necessità di *definire* in che modo i giocatori considerano e *conoscono* le convinzioni altrui, oltre che definire la nozione di convinzione comune.

1.1.4 Giochi in forma estesa

Un *gioco in forma estesa* è un modello più dettagliato di una situazione di interazione, e ne rappresenta la struttura sequenziale oltre alle caratteristiche strategiche. Nei giochi in forma strategica le strategie erano scelte una volta per tutte all'inizio del gioco, mentre in un gioco in forma estesa esse possono essere riconsiderate e variate durante lo svolgimento del gioco.
Restringeremo l'attenzione in questa tesi al caso dell'*informazione perfetta*: ad ogni momento del gioco in cui i giocatori devono prendere una decisione, ognuno di essi è perfettamente informato di tutte le scelte e le azioni proprie e degli altri giocatori avvenute in precedenza.
Innanzitutto un *albero con radice* è composto da un insieme Ω di nodi (Ω è finito se si richiede che l'albero sia finito) e da una relazione binaria $\rightarrowtail$ su Ω tale che:

- esiste un unico $w_0 \in \Omega$ senza predecessore, ossia tale che $\nexists w' \in \Omega$ t.c. $w' \rightarrowtail w_0$;
- $\forall w \in \Omega$ esiste un unico cammino $w_0, w_1, \ldots, w_m = w$ tale che $w_j - 1 \rightarrowtail w_j \ \ \forall j \leqslant m$.

L'insieme dei nodi senza successore (tali che $\nexists w'$ tale che $w \rightarrowtail w'$) è l'insieme dei nodi terminali Z, che è non vuoto se Ω è finito.
Definiamo per prima cosa:

Definizione 1.12. Una *struttura di gioco finito in forma estesa*[15] è composta da un insieme di giocatori N e da un albero con radice $(\Omega, \rightarrowtail)$, etichettato da una funzione $\iota : \Omega \setminus Z \longrightarrow N$.

Definizione 1.13. Un *gioco in forma estesa* $G = \langle N, (\Omega, \rightarrowtail), \iota, \prec_i \rangle$ si ottiene da una struttura di gioco in forma estesa con l'aggiunta di relazioni

[14] Common belief.
[15] Extensive game form.

di preferenza $\prec_i$ su Z per ogni giocatore $i \in N$, o equivalentemente una funzione di utilità $u : Z \longrightarrow \mathbb{R}^N$.

Da qui in avanti supporremo sempre che i giochi siano finiti.
Vediamo un esempio:

Divisione dei beni: due giocatori devono dividersi due beni egualmente preferiti e indivisibili nella seguente maniera: il primo giocatore propone una suddivisione, ed il secondo ha la possibilità di accettare o rifiutare. In caso di rifiuto i beni non vengono distribuiti.
La situazione è formalizzabile nel seguente modo:

- $N = \{I, II\}$;
- $(\Omega, \rightarrowtail)$ composto da: w_0 collegato a 3 nodi a, b, c corrispondenti alle 3 suddivisioni che I può proporre: entrambi per sè $(2, 0)$, equamente divisi $(1, 1)$ o entrambi al secondo $(0, 2)$; 6 nodi terminali, due per ognuno dei 3 nodi precedenti, ad indicare l'accettazione o il rifiuto di II;
- $\iota(w_0) = I$, $\iota(a) = \iota(b) = \iota(c) = II$;
- le funzioni di utilità sui nodi terminali sono riportate in Figura 1.1, la prima componente è u_1, la seconda u_2:

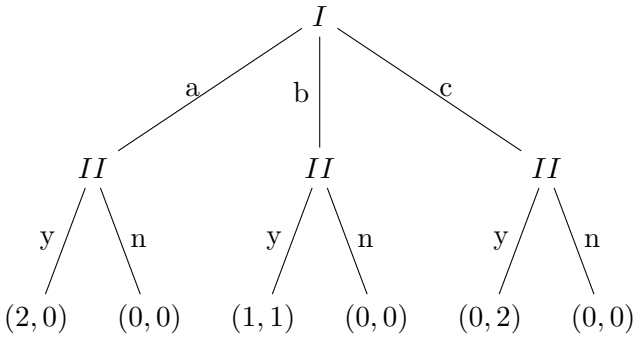


Figura 1.1: Suddivisione di due beni

A differenza dei giochi in forma strategica, nei modelli in forma estesa i concetti di azione e di strategia differiscono. Preso un nodo $w \in \Omega \setminus Z$ (che viene detto nodo decisionale) tale che $\iota(w) = i$, consideriamo l'insieme $A_i(w) = \{w' \in \Omega \mid w \rightarrowtail w'\}$: questo è l'insieme delle azioni disponibili per il giocatore i al nodo w. Una *strategia* per il giocatore i è una funzione che ad ogni nodo w tale che $\iota(w) = i$ associa un elemento di $A_i(w)$.

Nell'esempio precedente le strategie di I corrispondono alle 3 suddivisioni che può proporre, dunque $S_1 = \{a = (2,0), b = (1,1), c = (0,2)\}$. Le strategie per II sono tutte le funzioni dai suoi 3 nodi decisionali in y, n: $S_2 = \{yyy, yny, nyy, nnn, nny, yyn, ynn, nyn\}$ se accetta o meno le diverse proposte di I.

Diremo *profilo di strategie* $s = (s_1, \ldots, s_n)$ la scelta di una strategia per ogni giocatore.

Osservazione: ogni profilo di strategie s porta ad un unico nodo terminale. Definiamo infatti partendo da w_0 la sequenza $w_h = s_i(w_h)$ se $\iota(w_h) = i$. Dato che ogni nodo non terminale è associato ad un giocatore, ed una strategia è una mappa completa dall'insieme dei propri nodi decisionali nei nodi successivi, si giungerà per forza ad un nodo terminale. Questa sequenza rappresenta lo svolgimento effettivo del gioco. Scriveremo $O(s) = z$ se il profilo di strategie s "porta" al nodo z.

Osservazione: le strategie determinano il comportamento dei giocatori anche in nodi che non vengono raggiunti dallo svolgimento del gioco.
Anche nel caso dei giochi in forma estesa si definisce la nozione di equilibrio di Nash, ma non risulta pienamente soddisfacente (principalmente a causa di quest'ultima osservazione):

Definizione 1.14. Sia $\langle N, (\Omega, \rightarrowtail), \iota, \prec_i \rangle$ un gioco finito in forma estesa, il profilo di strategie s^* è un equilibrio di Nash se e solo se per ogni $i \in N$ e per ogni strategia s_i di i

$$u_i(O(s^*_{-i}, s_i)) \leqslant u_i(O(s^*_{-i}, s^*_i))$$

Vediamo cosa succede in alcuni esempi:

Il più facile gioco di entrata [Kre90]: i due giocatori sono due aziende, e la prima si trova a fronteggiare la possibilità di uscire da uno status quo inserendosi in un mercato (In o Out). La seconda azienda in caso di entrata può decidere se collaborare (C), o competere (con atteggiamento A-ggressivo). Il gioco è schematizzato nella figura 1.1.4.
Entrambi i profili di strategie (Out, A) ed (In, C) sono equilibri di Nash: fissata la scelta A di II, alla prima azienda conviene giocare Out, mentre fissata la scelta di Out per la prima, la seconda non è chiamata a giocare ed è indifferente tra le due opzioni. Allo stesso modo per l'altro profilo.
Nel profilo (Out, A) è però contenuta una scelta irrazionale di II: nel caso in cui II fosse chiamata a giocare non è giustificabile la scelta di A, che gli porta guadagno minore. In questi casi, con una terminologia introdotta da Schelling in [Sch60], si parla di *minaccia non credibile*.

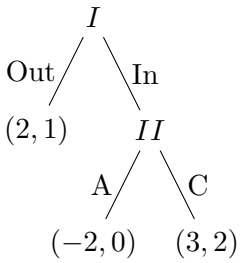


Figura 1.2: Il più facile gioco di entrata

Riadattamento del cavallo di Selten: il gioco è a 3 giocatori ed è descritto nella Figura 1.1.4.

Le strategie per I sono $\{a, d\}$, per II sono $\{A, D\}$ e per III sono date da

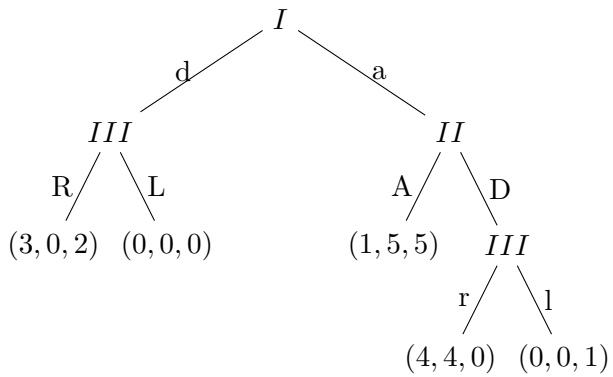


Figura 1.3: Riadattamento cavallo di Selten

$\{Rr, Rl, Lr, Ll\}$. Consideriamo per esempio il profilo (a, A, Ll): III non è chiamato a giocare ed è dunque indifferente al gioco; fissato il suo comportamento però, e fissata la scelta a di I, a due conviene scegliere A, e non portare III in gioco; allo stesso modo fissate le strategie di II e di III, la scelta migliore per I è a.
Anche in questo esempio, nel caso in cui III fosse chiamato a giocare, egli dovrebbe compiere un azione irrazionale. L'idea in questo caso è quella di una collusione tra II e III, ma la *promessa* di III di giocare L non è credibile.

In effetti l'equilibrio di Nash così definito annulla la struttura sequenziale del gioco in forma estesa, richiedendo che le strategie siano decise una volta per tutte all'inizio del gioco.
Definiamo il gioco in forma strategica $G^* = \langle N, A_i, u \rangle$ associato al gioco

in forma estesa $G = \langle N, (\Omega, \rightarrowtail), \iota, u \rangle$ mantenendo l'insieme N invariato, gli insiemi A'_i come gli insiemi delle strategie dei giocatori i nel gioco in forma estesa, e la funzione di utilità sui profili di strategie come la funzione u sui nodi terminali associati ai profili di strategie. Un profilo di strategie s è un equilibrio di Nash per il gioco in forma estesa G se e solo se è un equilibrio di Nash nel gioco in forma strategica G^* associato.

Si rende dunque necessario definire un raffinamento del concetto di equilibrio di Nash, che tenga conto della struttura sequenziale del gioco e lasci la possibilità di riformulare la propria strategia durante lo svolgimento. Cominciamo con la seguente definizione:

Definizione 1.15. Sia $G = \langle N, (\Omega, \rightarrowtail), \iota, \prec_i \rangle$ un gioco in forma estesa a informazione perfetta, per ogni $w \in \Omega$ si dice *sottogioco* con radice in w il gioco $G(w) = \langle N, (\Omega', \rightarrowtail), \iota_{|\Omega'}, \prec_i \rangle$ dove Ω' è il sottoalbero generato a partire da w, dunque l'insieme dei nodi w' tale che esiste una successione di nodi $w, w_1, \ldots, w'$ che lo collega a w tale che per ogni j $w_j \rightarrowtail w_{j+1}$. $\iota_{|\Omega'}$ è la restrizione della funzione ι ad Ω' e anche l'utilità è da considerarsi ristretta all'insieme $\Omega' \cap Z$.

Se s è un profilo di strategie per G allora indichiamo con $s_{|w}$ il profilo di strategie per il sottogioco $G(w)$, ottenuto restringendo le strategie s_i ad Ω'.

Definizione 1.16. Un profilo di strategie s^* si dice *equilibrio perfetto nei sottogiochi* per il gioco in forma estesa a informazione perfetta G se e solo se $s^*_{|w}$ è un equilibrio di Nash per ogni sottogioco $G(w)$.

Questa nozione di equilibrio è particolarmente soddisfacente per tre motivi:

i) mantiene la struttura sequenziale del gioco, tenendo in considerazione la possibilità di riconsiderare le proprie scelte durante lo svolgimento del gioco;
ii) scarta tutti i profili contententi nodi in cui il giocatore, se chiamato a giocare, dovrebbe fare scelte irrazionali: nell'esempio del più facile gioco di entrata, il profilo (Out, A) *non* è un equilibrio perfetto nei sottogiochi; allo stesso modo nel cavallo di Selten l'unico equilibrio perfetto nei sottogiochi è (D, A, Rl);
iii) vale il teorema di Kuhn[16]: ogni gioco finito in forma estesa a informazione perfetta ha un equilibrio perfetto nei sottogiochi.

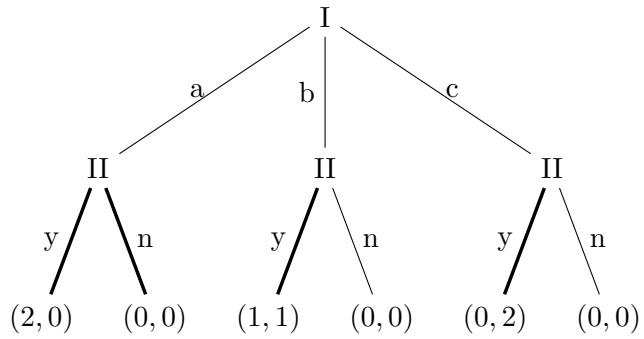
[16]Inizialmente dimostrato da Zermelo per il gioco degli scacchi: gli scacchi sono un gioco a due giocatori a informazione perfetta a somma zero, ed il teorema afferma che esiste un profilo di strategie perfetto nei sottogiochi che permette o a uno dei due giocatori di vincere, o ad entrambi di pareggiare. Fortunatamente la complessità di calcolo di una tale strategia è sufficientemente elevata da rendere il gioco degli scacchi ancora interessante.

La dimostrazione del teorema di Kuhn si basa su un procedimento costruttivo fondamentale detto *induzione a ritroso*. L'idea è la seguente:

- partendo da un nodo w pre-terminale (ossia un nodo i cui successori siano tutti nodi terminali), se $\iota(w) = i$ scegliere il ramo che massimizza l'utilità di i (scelta arbitraria se ve n'è più di uno);
- riportare il vettore utilità corrispondente alla scelta fatta e cancellare tutti i successori di w rendendolo un nodo terminale;
- ripetere il procedimento per ogni nodo pre-terminale fino alla radice.

Le scelte fatte ad ogni nodo corrispondono ad un profilo di strategie perfetto nei sottogiochi.
Vediamo l'algoritmo in azione nell'esempio della suddivisione dei beni:

- nel primo nodo pre-terminale i guadagni di due sono uguali, dunque riportiamo entrambe le scelte; nei due nodi successivi invece scegliamo l'azione che massimizza l'utilità di II. In figura riportiamo in grassetto le frecce che corrispondono alla scelta dell'induzione a ritroso.



- ora riportiamo i guadagni nei due casi distinti, e procediamo alla massimizzazione per I, ottenendo due equilibri perfetti nei sottogiochi, come nelle Figure 1.4 e 1.5:

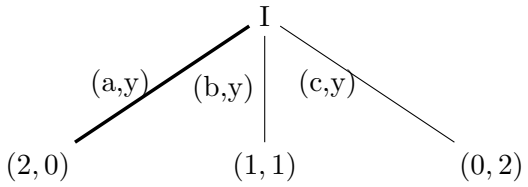


Figura 1.4: (a, yyy)

La soluzione per induzione a ritroso non è affatto unica, come si vede dall'esempio, ma si può garantire richiedendo che nessun giocatore sia indifferente rispetto a nessun nodo terminale:

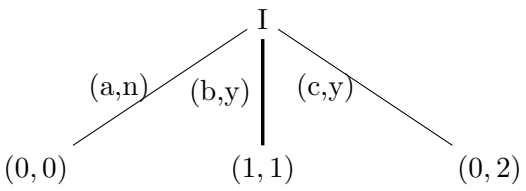


Figura 1.5: (a, nyy)

Definizione 1.17. Un gioco in forma estesa a informazione perfetta $G = \langle N, (\Omega, \rightarrowtail), \iota, \prec_i \rangle$ si dice *generico* se per ogni giocatore i e per $z, z' \in Z$ nodi terminali si ha che $u_i(z) = u_i(z') \Rightarrow z = z'$.

Osservazione: ogni gioco generico finito ha un unico equilibrio perfetto nei sottogiochi.

1.2 Fondamenti della teoria dei giochi

Riassumendo abbiamo definito sinora due classi di giochi: i giochi in forma strategica e i giochi in forma estesa (a informazione perfetta). Inoltre abbiamo definito due concetti risolutivi, l'equilibrio di Nash e gli equilibri perfetti nei sottogiochi. Diamo dunque una definizione generale di concetto risolutivo:

Definizione 1.18. Un *concetto risolutivo* è una funzione F che ad ogni gioco associa un sottoinsieme dei profili di strategie dunque

- per giochi in forma strategica $F(G) \subseteq \prod_{i \in N} A_i$;
- per giochi in forma estesa $F(G) \subseteq \prod_{i \in N} S_i$.

Indicheremo con N il concetto risolutivo che ad ogni gioco in forma strategica associa gli equilibri di Nash, con I la funzione che ad ogni gioco in forma estesa associa gli equilibri perfetti nei sottogiochi.

Come già accennato più volte, la teoria dei giochi si propone di fondare una teoria delle decisioni in condizioni di interazione, fondata sulla teoria classica delle decisioni ("semplici"). L'ipotesi di razionalità sui giocatori è dunque, nel caso di assenza di incertezza, la massimizzazione della propria utilità.
Dalla teoria delle decisioni, la teoria dei giochi eredita due interpretazioni per i concetti risolutivi, che sono il sottoinsieme di fondamenti della teoria dei giochi di cui ci occuperemo in questa tesi:

- *interpretazione descrittiva*: il concetto risolutivo descrive regolarità di comportamento dei giocatori, come sono state osservate. La domanda fondazionale in questo caso è quali condizioni (di conoscenza delle regole del gioco, delle convinzioni e della razionalità degli altri giocatori...) portano giocatori *razionali* a scegliere un determinato concetto risolutivo.

- *interpretazione normativa*: il concetto risolutivo associa un insieme di esiti ideali del gioco, a cui giocatori razionali (massimizzatori di utilità) dovrebbero conformarsi. Sotto questa interpretazione, si ricercano quali sono le condizioni *necessarie* perché un concetto risolutivo possa essere proponibile a giocatori razionali.

Lo scopo di questa tesi è di mostrare come la logica modale può essere utilizzata per modellizzare entrambe le interpretazioni.

1.3 Logica Modale

Inseriamo in seguito una breve introduzione alla logica modale.
Inizialmente la logica modale è servita ad introdurre e a formalizzare i concetti di necessità e possibilità, per estendere la logica classica oltre i concetti di vero e falso. La logica modale parte dunque dalla logica proposizionale (che denoteremo con **PC**) e ne estende il linguaggio con nuovi operatori:

- il *linguaggio* modale di base $\mathcal{L} = (Var, \tau)$ è composto da un insieme di variabili proposizionali Var, dagli usuali connettivi $\rightarrow$, $\neg$, e da un vocabolario τ; $\Box \in \tau$ viene detto operatore modale (unario);
- le *formule* sono costruite per induzione: le variabili proposizionali in Var sono formule ben formate, se ψ, φ sono formule anche $\neg\psi$ e $\psi \rightarrow \varphi$ ed anche $\Box\psi$ per $\Box \in \tau$ sono formule ben formate.

Anche le regole di dimostrazione sono un'estensione delle regole di **PC** (questo implica che ogni teorema dimostrabile nel calcolo proposizionale è dimostrabile in logica modale):

$$\textit{Modus Ponens:} \quad \text{da } \psi \text{ e } \psi \rightarrow \varphi \text{ dedurre } \varphi; \qquad \frac{\psi, \psi \rightarrow \varphi}{\varphi}$$

$$\textit{Nec:} \quad \text{da } \varphi \text{ dedurre } \Box\varphi; \qquad \frac{\varphi}{\Box\varphi}$$

Dato un insieme Γ di formule ben formate, una *dimostrazione* in Γ è una sequenza finita di formule $\varphi_1, \ldots, \varphi_n$ tali che per ogni i:
- φ_i è un'istanza di un assioma in Γ oppure

- esistono φ_j, $\varphi_k = \varphi_j \to \varphi_i$ con $k, j < i$ (φ_i è ottenuta per MP da formule precedenti) oppure
- $\varphi_i = \Box\varphi_j$ con $j < i$ (φ_i è ottenuta per Nec da formule precedenti).

Aggiungiamo una definizione per ogni operatore $\Box \in \tau$, a definire un nuovo operatore $\Diamond$, duale di $\Box$:

$$Def\colon\ \Diamond\psi \leftrightarrow \neg\Box\neg\psi$$

Gli assiomi logici (le tautologie) sono i seguenti schemi di assiomi:

K	
PC	...
K	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$

Diciamo *teoria* un insieme di assiomi Γ contenente gli assiomi logici. Denotiamo con **K** la teoria composta dai soli assiomi logici. Se Γ è un insieme di assiomi, scriveremo $\vdash_\Gamma \psi$ se esiste una dimostrazione in Γ con $\varphi_n = \psi$ e diremo che ψ è un teorema di Γ.

A titolo di esempio elenchiamo in seguito tre teorie tra le più importanti (senza riscrivere gli assiomi logici) nel linguaggio $\mathcal{L} = (Var, \Box)$:

T	$T : \Box p \to p$
S4	$T : \Box p \to p$
	$4 : \Box p \to \Box\Box p$
S5	$T\Box p \to p$
	$5 : \Diamond p \to \Box\Diamond p$

Mostriamo come esempio di dimostrazione che $\vdash_{\mathbf{S5}} \Box p \to \Box\Box p$, scrivendo nelle colonne di sinistra le regole o gli assiomi utilizzati:

	Col.1		Col.2		Col.3
T	$\Box p \to p$	**5**	$\Diamond p \to \Box\Diamond p$	**T + 5**	$\Box\Diamond p \leftrightarrow \Diamond p$
Def	$p \to \Box p$	$p/\Box p$	$\Diamond\Box p \to \Box\Diamond\Box p$	$p/\neg p$	$\Box\Diamond\neg p \leftrightarrow \neg\Diamond\neg p$
$p/\Box p$	$\Box p \to \Diamond\Box p$	**T**	$\Diamond\Box p \leftrightarrow \Box\Diamond\Box p$	Def	$\Diamond\Box p \leftrightarrow \Box p$

Semantica

I modelli su cui si interpretano le formule modali sono detti modelli di Kripke, e sono formati da un insieme di stati o mondi e da una valutazione sulle variabili per ogni stato. Definiamo innanzitutto:

Col.1+Col.2	$\Box p \rightarrow \Box\Diamond\Box p$
Col.3	$\Box p \rightarrow \Box\Box p \quad = \mathbf{4}$

Tabella 1.5: **4** è dimostrabile in **S5**

Definizione 1.19. Un *frame o struttura* $\mathcal{F} = (\Omega, R_\Box \mid \Box \in \tau)$ per il linguaggio $\mathcal{L} = (Var, \tau)$ è dato da:

- un insieme Ω;
- una relazione binaria $R_\Box$ su Ω per ogni $\Box \in \tau$.

Definizione 1.20. Un *modello* $\mathcal{M} = (\Omega, R_\Box, V)$ per il liguaggio $\mathcal{L} = (Var, \tau)$ è dato da:

- un frame $(\Omega, R_\Box)$ per $\mathcal{L}$;
- una valutazione $V : \Omega \times Var \longrightarrow \{0, 1\}$, o equivalentemente $V : Var \rightarrow \mathscr{P}(\Omega)$.

Le relazioni $R_\Box$ sono dette relazioni di accessibilità, e l'insieme Ω insieme degli stati (dei punti, dei mondi). Quest'ultima terminologia deriva dall'interpretazione classica di $\Box\psi$ come "ψ è necessario", in cui gli elementi di Ω svolgono il ruolo di "mondi possibili".
L'interpretazione V si estende per induzione ad un interpretazione su tutte le formule:

•) $V(w)(\neg\psi) = 1$ sse $V(w)(\psi) = 0$;

•) $V(w)(\psi \rightarrow \varphi) = 1$ sse $V(w)(\psi) = 0$ o $V(w)(\varphi) = 1$;

•) $V(w)(\Box\psi) = 1$ sse $\forall v \in \Omega$ tale che $wR_\Box v$ $V(w)(\psi) = 1$

L'idea sottostante è che ogni stato in Ω valuta le formule proposizionali (non compaiono infatti quantificatori su Ω), mentre la formula $\Box\psi$ è vera se ψ è vera in tutti gli stati $R_\Box$-accessibili da w.

Dato un modello $\mathcal{M} = (\Omega, R, V)$, il valore di verità di una formula φ può essere considerato a diversi livelli:

i) φ è *vera* nello stato w, e si scrive $\mathcal{M}, w \models \varphi$, se $V(w)(\varphi) = 1$;
ii) φ è *globalmente vera* in $\mathcal{M}$, e si scrive $\mathcal{M} \models \varphi$, se $\mathcal{M}, w \models \varphi$ per ogni $w \in \Omega$;
iii) φ è *valida* sul frame $\mathcal{F}$, e si scrive $\mathcal{F} \models \varphi$, se per ogni valutazione $V : Var \longrightarrow \mathscr{P}(\Omega)$ si ha che $(\mathcal{F}, V) \models \psi$. In altre parole se φ è globalmente vera in ogni modello basato sul frame $\mathcal{F}$.

A differenza che nel caso proposizionale, il collegamento tra sintattica e semantica dato dai teoremi di completezza avviene in questo caso a livello dei frame, ed è il concetto di validità a giocare il ruolo principale.

Correttezza

Data una teoria Λ, richiedere che sia *corretta* rispetto ai suoi modelli significa richiedere che tutte le conseguenze della teoria continuino a valere sulla classe dei propri modelli, dunque che le regole di deduzione dei teoremi preservino la "verità". Per le logiche modali, come anticipato, nel ruolo dei modelli ci sono i frame e alla verità si sostituisce la validità:

Λ è *corretta* rispetto alla classe di frame $\mathscr{F}$ sse $\vdash_\Lambda \varphi \Rightarrow \mathcal{F} \models \varphi$ per ogni $\mathcal{F} \in \mathscr{F}$.

Innanzitutto, K è un assioma logico perché:

Lemma 1.21. *K è valido su tutti i frame.*

Grazie al fondamentale lemma che esponiamo in seguito, è possibile dimostrare la correttezza di ogni teoria modale rispetto alla propria classe di frame:

Lemma 1.22. *Sia Γ un insieme di formule ben formate ed $\mathcal{F}$ un frame su cui sono valide tutte le formule di Γ. Ogni teorema di $K \cup \Gamma$ è valido su $\mathcal{F}$.*

Dimostrazione. La dimostrazione è per induzione sulle dimostrazioni: se α è un istanza di un assioma in $K \cup \Gamma$ allora è valido in $\mathcal{F}$ per ipotesi.
Se α è ottenuta per modus ponens da β e $\beta \to \alpha$ (valide per ipotesi in tutto il modello) allora β è valida per le regole di V.
Infine il caso di α ottenuta per necessitazione (ossia $\alpha = \Box\beta$ per β valida). Se per assurdo α non fosse valida questo implicherebbe che esiste un modello e un mondo dove β è falsa, ma per ipotesi induttiva β è verificata in tutto il frame. □

Per ottenere la correttezza di una teoria Λ è dunque sufficiente controllare la validità degli assiomi su una certa classe di frame: per il Lemma 1.22 tutti i teoremi di Λ saranno automaticamente validi in quella classe.
Mostriamo per esempio che:

Proposizione 1.23. ***S5*** *è corretta rispetto ai frame riflessivi, simmetrici e transitivi (frame dove R è una relazione di equivalenza).*

Dimostrazione. Dimostriamo che gli assiomi di **S5** sono validi su questi frame. L'assioma T: $\Box p \to p$ è valido poiché essendo il frame riflessivo, Rww vale per ogni $w \in W$ e dunque se $\mathcal{M}, w \models \Box p$ allora $\mathcal{M}, w \models p$.
Per **5**: $\Diamond p \to \Box\Diamond p$, notiamo che se $\mathcal{M}, w \models \Diamond p$ allora esiste w' tale che $\mathcal{M}, w' \models p$. Ora per transitività e simmetria ogni mondo con cui il primo w è in relazione può vedere w', dunque vale che $\mathcal{M}, w'' \models \Diamond p$ per ogni w'' in relazione con w e così $\mathcal{M}, w \models \Box\Diamond p$. □

Allo stesso modo si dimostra che **T** è corretta rispetto ai frame riflessivi ed **S4** è corretta rispetto ai frame riflessivi e simmetrici. Ovviamente **K** è corretta rispetto alla classe di tutti i frame.

1.3.1 Completezza tramite canonicità

Data una classe di modelli si può in generale definire la nozione di *conseguenza logica*, scrivendo che "φ segue logicamente da ψ" se in ogni modello in cui vale ψ vale anche φ. Nel caso della logica modale, data una classe $\mathscr{F}$ di frame, scriveremo $\psi \models_{\mathscr{F}} \varphi$ per indicare che per ogni $\mathcal{F} \in \mathscr{F}$ si ha che $\mathcal{F} \models \psi \Rightarrow \mathcal{F} \models \varphi$.

Ciò che si richiede con la completezza[17] di una teoria Λ rispetto a una classe $\mathscr{F}$ di frames è che:

$$\psi \models_{\mathscr{F}} \varphi \Rightarrow \vdash_{\Lambda} \psi \rightarrow \varphi$$

ossia che conseguenza logica e conseguenza sintattica coincidano (la freccia inversa è data dalla correttezza). Dimostreremo in realtà un enunciato equivalente: se φ è vera in ogni frame di $\mathscr{F}$ allora φ è dimostrabile da Λ.

Esponiamo ora un metodo di dimostrazione piuttosto generale detto modello canonico. Non è un metodo universale, nel senso che non dimostra la completezza di ogni teoria modale, ma è sufficiente per tutte quelle che prenderemo in considerazione.
Data una teoria Λ, l'idea consiste nel costruire un modello[18], che chiameremo $\mathcal{M}^c_\Lambda$, tale che se una formula ψ è globalmente vera su $\mathcal{M}^c_\Lambda$ allora ψ è dimostrabile in Λ. A questo punto per ottenere la completezza è sufficiente mostrare che il frame $\mathcal{F}^c$ su cui è definito il modello canonico appartiene alla classe di correttezza $\mathscr{F}$ di Λ. Infatti se φ è valida su tutti i frame di $\mathscr{F}$, in particolare è valida sul frame del modello canonico e dunque globalmente vera su $\mathcal{M}^c_\Lambda$, e così è dimostrabile in Λ.

Diamo innanzitutto qualche definizione: diciamo che un insieme di formule α è Λ-*inconsistente* se esistono $\gamma_1 \dots \gamma_n \ \in \alpha$ tali che $\vdash_\Lambda \neg(\gamma_1 \wedge \dots \wedge \gamma_n)$; diciamo che α è Λ-consistente altrimenti. Inoltre, α è Λ-*consistente massimale* se è massimale rispetto all'inclusione.
Le proprietà fondamentali di questi insiemi sono elencate nel seguente lemma:

[17] Completezza debole; verrà in realtà dimostrata anche la completezza forte o compattezza, che sostituisce a ψ un insieme di formule di cardinalità arbitraria. Le teorie qui considerate sono tutte compatte, ma le due nozioni non sono equivalenti, per esempio se W è $\Box(\Box\psi \rightarrow \psi) \rightarrow \Box\psi$, **KW** non è compatta ma è completa rispetto ad una classe di frames finiti, si veda ad esempio [HC96] pag.139.

[18] Le dimostrazioni sono prese da [HC96] e da [BdRV01] e non sono svolte in tutti i dettagli.

Lemma 1.24. *Se α è Λ-consistente massimale allora:*

i) per ogni formula ben formata β, o $\beta \in \alpha$ o $\neg\beta \in \alpha$;

ii) $\alpha \vee \beta \in \alpha \Rightarrow \alpha \in \alpha$ oppure $\beta \in \alpha$;

iii) $\alpha \wedge \beta \in \alpha \Rightarrow \alpha \in \alpha$ e $\beta \in \alpha$;

iv) α è chiuso per modus ponens.

Per dimostrare l'esistenza di un insieme Λ-massimale consistente si utilizza:

Lemma 1.25 (Lindenbaum)**.** *Ogni insieme Λ-consistente può essere esteso a un insieme Λ-consistente massimale.*

Dimostrazione. Sia α un insieme Λ-consistente, è sufficiente considerare la famiglia di Σ sottoinsiemi di formule $\{\Sigma/\alpha \subseteq \Sigma,\ \Sigma\ \Lambda\text{-consistente}\}$, e concludere la dimostrazione usando il lemma di Zorn. □

Diamo ora la seguente definizione generale:

Definizione 1.26. Definiamo *modello canonico* di Λ il modello $\mathcal{M}^c_\Lambda = \langle \Omega^c, R^c, V^c \rangle$ dove:

- Ω^c sia formato da tutti gli insiemi Λ-consistenti massimali di formule;
- se indichiamo con $\Box^{-1}(\alpha) = \{\varphi \mid \Box\varphi \in \alpha\}$,

$$\alpha R^c \beta \text{ per } \alpha \text{ e } \beta \text{ in } \Omega^c \text{ sse } \Box^{-1}(\alpha) \subseteq \beta.$$

Nel seguente lemma è contenuta l'idea per la costruzione della valutazione V^c:

Lemma 1.27. *$\Diamond\psi \in \alpha \Rightarrow$ esiste β Λ-consistente massimale tale che $\psi \in \beta$ ed $\alpha R \beta$.*

Dimostrazione. Mostriamo che $\Box^{-1}(\alpha) \cup \psi$ è Λ-consistente. Se così non fosse infatti esisterebbero $\gamma_1 \dots \gamma_n$ tali che $\Box\gamma_1 \dots \Box\gamma_n \in \alpha$ tali che $\vdash_\Lambda \neg(\gamma_1 \wedge \dots \gamma_n \wedge \psi)$

ossia	$\vdash_\Lambda (\gamma_1 \wedge \dots \wedge \gamma_n) \rightarrow \neg\psi$
per Nec:	$\vdash_\Lambda \Box((\gamma_1 \wedge \dots \wedge \gamma_n) \rightarrow \neg\psi)$
per K:	$\vdash_\Lambda \Box(\gamma_1 \wedge \dots \wedge \gamma_n) \rightarrow \Box\neg\psi$
per distributività:	$\vdash_\Lambda (\Box\gamma_1 \wedge \dots \wedge \Box\gamma_n) \rightarrow \neg\Diamond\psi)$
ossia:	$\vdash_\Lambda \neg(\Box\gamma_1 \wedge \dots \wedge \Box\gamma_n) \vee \neg\Diamond\psi)$
dunque:	$\vdash_\Lambda \neg(\Box\gamma_1 \wedge \dots \wedge \Box\gamma_n \wedge \Diamond\psi)$

ma $\Diamond\psi \in \alpha$ e dunque α sarebbe Λ-inconsistente, contro le ipotesi.

□

Osservazione: è possibile definire in maniera del tutto equivalente la relazione R^c_Λ in funzione dell'operatore duale $\Diamond$, infatti se $\Diamond(\alpha) = \{\Diamond\psi | \psi \in \alpha\}$, vale che: $\alpha R^c \beta \Leftrightarrow \Diamond(\alpha) \subseteq \beta$.

Possiamo dunque definire la valutazione V:

$$V^c(\alpha)(p) = 1 \text{ sse } p \in \alpha$$

per ogni $\alpha \in \Omega^c$ e per ogni variabile proposizionale p.
Con argomenti simili ai precedenti si dimostra il teorema fondamentale:

Teorema 1.28. *Per ogni φ formula ben formata, $V^c(\alpha)(\varphi) = 1 \Leftrightarrow \varphi \in \alpha$.*

Da questo teorema segue che se ψ è globalmente vera su $\mathcal{M}^c_\Lambda$ allora che $\vdash_\Lambda \psi$. Supponiamo infatti che $\nvdash_\Lambda \psi$, allora $\{\neg\psi\}$ è Λ-consistente, dunque estendibile per il Lemma di Lindenbaum ad un sottoinsieme Λ-consistente massimale w. Per la definizione di Ω^c, α è un punto del modello canonico, e dato che $\neg\psi \in \alpha$ otteniamo $\mathcal{M}^c_\Lambda, \alpha \models \neg\psi$ per il Teorema 1.1, contro l'ipotesi di validità di ψ. Abbiamo dimostrato la seguente:

Proposizione 1.29. *Ogni teoria Λ è completa rispetto al proprio modello canonico $\mathcal{M}^c_\Lambda$.*

Per ottenere la completezza di Λ rispetto alla propria classe di correttezza $\mathscr{F}$ è dunque sufficiente dimostrare che il frame canonico appartiene a questa classe, ossia che $\mathcal{F}^c_\Lambda \models \Lambda$.

Definizione 1.30. Una teoria Λ si dice *canonica* se $\mathcal{F}^c_\Lambda \models \Lambda$.

In particolare:

Definizione 1.31. Una formula ψ si dice canonica, se per ogni teoria Λ, $\psi \in \Lambda$ implica che $\mathcal{F}^c_\Lambda \models \psi$.

Dunque un assioma è canonico se forza il frame del modello canonico a validarlo. In certi casi diremo che un assioma ψ è canonico rispetto a una certa proprietà, seguendo il ragionamento seguente: ψ è valido su un frame $\mathcal{F}$ che verifica questa proprietà, e ψ forza il modello canonico a godere di questa proprietà, dunque ψ è canonico.
Dimostrare la completezza di una logica Λ tramite canonicità significa dimostrare che tutti gli assiomi di Λ sono validi sul frame canonico:

Proposizione 1.32. *Λ è canonica se e solo se tutti i suoi assiomi sono canonici.*

1.3.2 Completezza di S5

Mostriamo per esempio che $\mathcal{F}^c_{\mathbf{S5}}$ appartiene alla classe di correttezza di **S5** (insiemi con una relazione di equivalenza):

Proposizione 1.33. *$R^c_{\mathbf{S5}}$ è riflessiva, simmetrica e transitiva.*

Dimostrazione. Per la riflessività: preso $w \in \Omega^c$, se $\Box\varphi \in w$ allora per T e per chiusura di w tramite MP si ottiene che anche $\varphi \in \Omega$, dunque $\Box^{-1}(w) \subseteq w$.
Per la simmetria: supponiamo che wRw', ossia che $\Box^{-1}(w) \subseteq w'$. Se per assurdo $\Box^{-1}(w') \not\subseteq w$, e dunque R non fosse simmetrica, esisterebbe una formula ψ tale che $\Box\psi \in w'$ ma $\psi \notin w$. Per la massimalità di w si avrebbe che $\neg\psi \in w$. Ora $p \to \Box\Diamond p$ è un teorema di **S5** e w è chiuso per MP, dunque $\Box\Diamond\neg\psi \in w$. Per concludere, $\Diamond\neg\psi = \neg\Box\psi \in w'$, mentre avevamo supposto il contrario.
Per la transitività: se wR^c_v e vR^cz allora per 4 e chiusura per MP, se $Box\varphi \in w$ si ottiene che $\Box\Box\varphi \in w$. Dunque $\Box^{-1}(w) \subseteq z$. □

Corollario 1.34. S5 *è canonica, dunque completa rispetto alla classe dei frame* (Ω, R) *dove* R *è una relazione di equivalenza.*

1.3.3 Logica multiagente

Nella definizione di linguaggio modale data nei paragrafi precedenti ci si è limitati a considerare operatori modali unari $\Box$. La definizione si può generalizzare:

Definizione 1.35. Un vocabolario $\tau = (\mathcal{O}, \rho)$ è dato da un insieme non vuoto $\mathcal{O}$ e da una funzione $\rho : \mathcal{O} \longrightarrow \mathbb{N}$. Gli elementi di $\mathcal{O}$ sono detti operatori modali e la funzione ρ arietà. Un linguaggio modale $\mathcal{L} = (Var, \tau)$ è dato da un insieme di variabili proposizionali Var e da un vocabolario τ.

Se indichiamo un elemento di $\mathcal{O}$ di arietà $\rho(\triangle) = n$ con $\triangle$, e $\psi_1, \ldots, \psi_n$ sono formule ben formate, allora anche $\triangle(\psi_1, \ldots, \psi_{\rho(\triangle)})$ è una formula ben formata.
I modelli in cui sono interpretabili le formule del linguaggio $\mathcal{L} = (Var, \tau)$ sono strutture di Kripke $(\Omega, (R^{\rho(\triangle)+1}_{\triangle} \mid \triangle \in \mathcal{O}), V)$, composte da un insieme di stati Ω, una valutazione V da Var nelle parti di Ω, e una relazione $n+1$-aria per ogni operatore modale n-ario in $\mathcal{O}$.
La valutazione V si estende alle formule modali come nel caso semplice, con in più:

$$V(w)(\triangle^{\rho(\triangle)}(\psi_1, \ldots, \psi_n)) = 1 \Leftrightarrow$$

$$\forall w_1, \ldots w_n \text{ t.c. } R_{\triangle}(w, w_1, \ldots, w_n) \text{ si ha che } V(w_i)(\psi_i) = 1.$$

Preso un linguaggio $\mathcal{L} = (Var, \tau)$ diremo *linguaggio multiagente* (per n agenti) $\mathcal{L}_n = (Var, \tau_n)$ dove Var è lo stesso insieme, e τ_n è dato da n copie di τ indicizzate con indici da 1 a n. Dunque se $\tau = (\mathcal{O}, \rho)$ allora $\tau_n = (\mathcal{O}_1, \ldots, \mathcal{O}_n, \rho_1, \ldots, \rho_n)$; considerando l'unione degli $\mathcal{O}_i$ così come l'unione delle mappe ρ_i e le formule sono costruite come alla precedente definizione.

Definizione 1.36. Se Λ è una $\mathcal{L}$-teoria, diciamo *teoria multiagente* la corrispondente $\mathcal{L}_n$-teoria per n agenti Λ_n composta da n copie degli assiomi di Λ, per gli operatori modali in $\mathcal{L}_n$.

Dato un frame $\mathcal{F} = (\Omega, R_\triangle)$ per Λ (ossia su cui Λ è corretta), il frame $\mathcal{F}_n = (\Omega, R_{\triangle,1}, \ldots, R_{\triangle,n})$ è un frame per Λ_n. Dunque se $\mathscr{F}$ è la classe di correttezza di Λ, la classe di correttezza di Λ_n è $\mathscr{F}_n$, composta dai frames $\mathcal{F}_n$ per ogni $\mathcal{F} \in \mathscr{F}$. Per teorie canoniche questo risultato si estende alla completezza:

Proposizione 1.37. *Se Λ è una $\mathcal{L}$-teoria, e Λ è canonica rispetto alla classe $\mathscr{F}$ di frames, allora Λ_n è canonica rispetto alla classe $\mathscr{F}_n$ di frames.*

Dimostrazione. Preso un assioma φ_i di Λ_n, gli operatori che vi compaiono sono tutti indicizzati dallo stesso i, e le relazioni $R^C_{\triangle,i}$ del modello canonico sono costruite separatamente per ogni i. La canonicità di Λ si può applicare dunque alla sua copia i-esima, ottenendo la canonicità di φ_i, e allo stesso modo di tutti gli altri assiomi di Λ_n. □

Corollario 1.38. $\mathbf{S5}_n$ *è completa rispetto ai frame $(\Omega, R_1, \ldots, R_n)$ dove le R_i sono relazioni di equivalenza.*

Capitolo 2

Interpretazione descrittiva: $\mathbf{S5}_n^C$ e strutture di Aumann

Uno dei primi articoli riguardanti l'interpretazione descrittiva di concetti risolutivi è [AB95], di cui riportiamo i primi passaggi a titolo di esempio:

"Preliminary Observation: Suppose that each player is rational, knows his own payoff function, and knows the strategy choices of the others. Then the players' choices constitute a Nash equilibrium in the game being played."

Sebbene possa sembrare un risultato banale, è esplicativo di ciò che si ricerca in questo ambito: una relazione che leghi condizioni di conoscenza (espresse in un appropriato modello) ad un concetto risolutivo, data la solita ipotesi di razionalità come massimizzazione dell'utilità[1].

In questo capitolo dimostreremo l'equivalenza tra il modello per la conoscenza classicamente utilizzato in teoria dei giochi (le strutture di Aumann, introdotte in [Aum76]) e l'approccio proveniente dalla logica modale (la logica epistemica di [Hin62] estesa al caso di più giocatori).

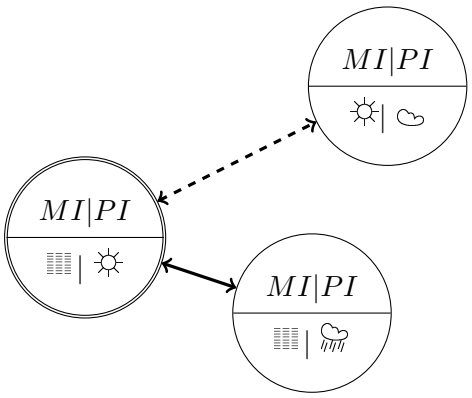
2.1 Logica epistemica: un modello per la conoscenza

Knowledge is power, arm yourself.
(Propagandhi)

Una delle possibili definizioni di conoscenza che sono state proposte, nel tentativo di costruire un modello per la conoscenza di un individuo, suggerisce in maniera diretta l'utilizzo della logica modale.

[1] Un articolo che riassume i principali risultati ottenuti in questo campo è [BB99].

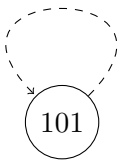
Esponiamo il modello tramite un esempio: mi trovo a Milano in una nebbia fittissima, *so* che c'è nebbia, qualsiasi cosa possa immaginare che succeda nel mondo intorno a me, non vedo nulla quindi c'è nebbia. Nel contempo posso immaginare un mio amico che a Pisa sta prendendo il sole in terrazza, ma potrebbe benissimo essere rintanato in casa a causa della pioggia: non *so* se a Pisa piove, considero possibile uno stato del mondo in cui piove a Pisa e un altro in cui non piove.
Nell'esempio la conoscenza di un individuo è identificata con la quantità di informazione che l'individuo possiede riguardo alle possibilità del mondo che lo circonda: "sa che x se x si verifica in tutti gli stati del mondo che ritiene possibili" (tra i quali è certamente presente lo stato in cui si trova realmente).
Raccogliendo in un insieme Ω tutti gli stati possibili del mondo (limitandosi a considerare soltanto certi aspetti, certe variabili su cui si è incerti), si definisce una relazione tra i mondi, indicando per ogni stato del mondo quali sono quegli stati ritenuti "plausibili" se ci si trova in quel determinato stato.
Tornando all'esempio: se mi trovo nello stato in cui io sono a Milano, a Milano c'è nebbia e a Pisa cè il sole, allora considero possibile lo stato in cui sono a Milano e a Milano c'è nebbia ma a Pisa piove; non considero possibile lo stato in cui sono a Milano e a Milano splende il sole.



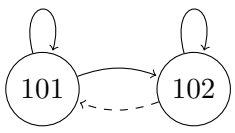
La freccia tratteggiata indica la relazione mancante, mentre lo stato in cui mi trovo è segnalato con un doppio contorno. La relazione nell'esempio è volutamente biunivoca, infatti il modello di conoscenza che esponiamo richiede che le relazioni di accessibilità siano relazioni di equivalenza.
Proviamo a giustificare questo fatto con un esempio, interpretando la relazione che lega due mondi possibili in termini di informazione: lo stato w è accessibile dallo stato v se dalle informazioni in suo possesso un agente razionale non sa dedurre in quale dei due si trova (risulterà fondamentale il fatto che tra le informazioni in possesso dell'individuo c'è anche propria conoscenza).
L'esempio è il seguente: l'insieme dei mondi è un albergo, ed ogni stanza è

un mondo. Fissiamo un insieme di informazioni rilevanti per l'osservatore (le variabili del linguaggio proposizionale), ad indicare tutte le proprietà delle stanze che gli permettono di distinguere l'una dall'altra. L'assegnamento descrive come è fatto l'albergo nella realtà. Grazie ad ingegnosi giochi di specchi l'individuo è in grado di vedere contemporaneamente diverse stanze, senza capire in quale si trova.
Per l'interpretazione data (e quozientando due stanze se sono identiche), tutte le stanze legate da specchi sono collegate dalla relazione di accessibilità dell'individuo: dalle informazioni in suo possesso non sa distinguere dove si trova.
Un buon candidato a descrivere la conoscenza di una certa proposizione ψ è dunque che essa sia vera in ogni stanza visibile. Dal fatto che in ogni stanza che vede contemporaneamente l'abat-jour è accesa, è un ottimo ragionamento dedurre che nella stanza in cui si trova veramente l'abat-jour è accesa, dunque l'individuo *sa* che nella sua stanza l'abat-jour è accesa.
Esaminiamo le proprietà di questa relazione di accessibilità:

- Se non fosse riflessiva (specchi assorbenti in ogni stanza); nei modelli contenenti "punti morti", stati non in relazione con nessun altro, si avrebbe che per condizione vacua l'agente saprebbe tutto e il contrario di tutto. Nell'albergo qui sotto ad esempio egli saprebbe che nella sua stanza c'è un tappeto, ma anche che non c'è, la qual cosa non rappresenta una proprietà desiderabile.



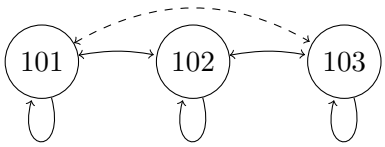
- Simmetria (specchi monodirezionali); sembrerebbe plausibile che per esempio dalla stanza 101 vede la 102 (la considera possibile) e dunque non saprebbe dire dove si trova, mentre la 102 è priva di specchi. Ma si consideri il seguente albergo:



  Se a è una proprietà di una stanza che le distingue (esiste perché non esistono due stanze uguali), allora l'individuo è in grado di distinguere

in quale stanza si trova *sulla base della propria conoscenza.* Infatti se è nella 101 allora $101 \models \Diamond a \wedge \Diamond \neg a$, ossia considera entrambe possibili; nella 102 invece non è così: $102 \models \Box \neg a$.
Più semplicemente: se in w considero possibile essere in v, ma dal mondo v non considero possibile essere in w, questa sola informazione mi permette di sapere dove mi trovo realmente, dunque non avrebbe senso considerare possibile nel mondo v essere in w. La relazione di accessibilità deve rappresentare la medesima informazione, e non può essere modificata da essa stessa.

- Se non fosse transitiva (specchi non collegati), allo stesso modo che per la simmetria, ma con qualche ragionamento in più, si incappa negli stessi problemi incontrati con la proprietà di simmetria. Si consideri il seguente albergo:



nella stanza 101 non cè nessun tappeto e l'abat-jour è accesa, nella 102 anche ma cè un tappeto, mentre nella 103 cè un tappeto e l'abat-jour è accesa. Se dalla stanza 101 si vedesse la 102 e da questa la 103, ma non si potesse dalla 101 accedere alla 103, di nuovo la relazione di accessibilità permetterebbe all'individuo di distinguere in quale stanza si trova, rendendo inutile la relazione stessa. Infatti:
$101 \models \Box a \wedge \neg \Box\ b$
$102 \models \neg \Box a \wedge \neg \Box b$
$103 \models \neg \Box a \wedge \Box b$
leggendo con a l'abat-jour è accesa e con b c'è un tappeto, l'agente può distinguere la stanza in cui è veramente sulla base della propria conoscenza.

Filosofi (e non) hanno scritto e dibattuto apporfonditamente se questo sia o meno un buon modello per la conoscenza (cfr. ad esempio [Hin62], [Len78]). Concordando con l'opinione espressa in [FHMV95], consideriamo inutile dare un'unica definizione di conoscenza, ma consideriamo la definizione sopra enunciata *un* buon modello per la conoscenza, perlomeno per le applicazioni per cui verrà utilizzata in seguito.

Possiamo dunque modellizzare problemi riguardanti la conoscenza con strutture formate da un insieme di stati legati da una relazione di equivalenza, e in questo modo utilizzare la logica modale precedentemente introdotta per esprimere e studiare le proprietà della conoscenza così definita.

È stato in precedenza dimostrato che **S5** è completa rispetto a strutture di Kripke con relazioni di equivalenza e dunque essa è la teoria adatta per "parlare" di conoscenza.
Se interpretiamo le formule del linguaggio $\mathcal{L} = (Var, \Box)$ come formule che si riferiscono alla conoscenza di un individuo, allora le variabili Var descrivono le "variabili rilevanti" degli stati del mondo, e $\Box\psi$ si legge "l'individuo sa ψ". Dai paragrafi precedenti sappiamo che le validità esprimibili in questo linguaggio sono tutte e sole i teoremi di **S5**.

I problemi filosofici nascono dalla lettura degli assiomi di **S5** in questo contesto. Enunciamo infatti una traduzione in termini epistemici delle regole e degli assiomi che concernono l'operatore conoscenza $\Box$:

- *Nec*: se ψ è valida, cioè vera in ogni stato di ogni modello, allora l'individuo sa ψ;
- **K** (assioma di onniscenza logica): $\Box(\psi \to \varphi) \to (\Box\psi \to \Box\varphi)$. Afferma che l'agente sa tutte le conseguenze della propria conoscenza: se sa che $\psi \to \varphi$ e sa anche ψ allora sa anche φ: L'agente è "dotato" di Modus Ponens;
- **T** (assioma di verità): $\Box\psi \to \psi$. Asserisce che un individuo sa soltanto cose vere. Secondo certi filosofi è qui che risiede la differenza tra conoscenza e convinzione: mentre si possono credere cose che sono false, non si possono sapere falsità;
- **4** (assioma di introspezione positiva[2]): $\Box\psi \to \Box\Box\psi$. Asserisce che l'individuo sa di sapere;
- **5** (assioma di introspezione negativa): $\neg\Box\psi \to \Box\neg\Box\psi$. Asserisce che l'individuo è "socratico": sa di non sapere tutto ciò che non sa.

Il primo e l'ultimo assioma sono al primo posto nel generare complicazioni di carattere interpretativo. Da una parte supporre che un individuo sappia tutte le conseguenze della propria conoscenza sembra una richiesta troppo forte; allo stesso modo per soddisfare l'assioma **5** l'individuo dovrebbe essere capace di considerare tutte le proposizioni riguardanti cose che non sa.

Questo dovrebbe far sovvenire i due approcci normativo e descrittivo: l'interpretazione sopra proposta vede **S5** come modello di come un individuo dovrebbe comportarsi o si comporta riguardo alla sua conoscenza, quindi soffre delle medesime confusioni di cui ai capitoli iniziali.
Dal punto di vista normativo le discussioni qui sopra sono al solito inutili: se un individuo si considera disposto a ragionare come **S5** per quel che concerne la sua conoscenza, può sfruttare la semplicità con cui in logica epistemica sono trattate situazioni di conoscenza piuttosto intricate (fatto che sarà più

[2]Riportato classicamente tra gli assiomi della logica epistemica anche se è un teorema, come mostrato alla Sezione 1.3.

esplicito quando considereremo la conoscenza incrociata di più individui). Dal punto di vista descrittivo i problemi restano: in certi casi quello appena esposto risulta essere un buon modello per la conoscenza, ma le ipotesi di razionalità che stanno alla base sono difficilmente giustificabili.

2.1.1 Logica epistemica multiagente

Se **S5** è interpretabile come logica epistemica, la teoria multiagente $\mathbf{S5}_n$ è interpretabile come modello per la conoscenza di n agenti. Il vocabolario è composto da n operatori unari $\Box_i$, interpretati come "l'agente i sa" , e gli assiomi sono **K**, **T** e **5** per ogni operatore $\Box_i$. Le formule di $\mathbf{S5}_n$ sono interpretate su frame $(\Omega, R_1, \ldots, R_n)$ dove le R_i sono relazioni di equivalenza. Dal Corollario 1.38 sappiamo che $\mathbf{S5}_n$ è completa rispetto ai frame $(\Omega, R_1, \ldots \ldots R_n)$ dove le R_i sono relazioni di equivalenza.

Nell'analisi di situazioni strategiche capita spesso di dover considerare conoscenze incrociate tra agenti, situazioni difficilmente rappresentabili non solo nel linguaggio comune, ma anche tramite l'analisi matematica. Nel linguaggio della logica epistemica multiagente risulta molto semplice formalizzare frasi complicate proprio riguardo a situazioni di conoscenze incrociate: "De Giorgi non sa se Nash sa che de Giorgi sa che Nash sa che qualcun altro ha già risolto il diciannovesimo problema di Hilbert" diventa

$$\neg\Box_g\Box_n\Box_g\Box_n H \wedge \neg\Box_g\neg\Box_n\Box_g\Box_n H$$

leggendo $\Box_g$ per "de Giorgi sa che", $\Box_n$ per "John Nash sa che" e H per "qualcuno ha risolto il diciannovesimo problema di Hilbert".

2.1.2 Common Knowledge

Nelle due sezioni precedenti è stato costruito:

- •) un modello per la conoscenza di n individui, composto da un insieme di mondi possibili che riassumono tutte le possibili configurazioni di un certo insieme di parametri incerti, e da n relazioni binarie, una per ogni individuo, a rappresentare la "possibilità" relativa dei vari mondi;

- •) un linguaggio adatto ad esprimere le proprietà della conoscenza così modellizzata ed un insieme di assiomi che esprimono *tutte* le validità della conoscenza in questo linguaggio.

Nel caso di un singolo agente il modello considerato risulta soddisfacente per esprimere tutti i concetti rilevanti per l'analisi della conoscenza. Nella generalizzazione a più individui nasce invece la necessità di ampliare il linguaggio per poter esprimere nuovi concetti che insorgono nel considerare conoscenze condivise e conoscenze incrociate di più agenti.

Innanzitutto si vuole esprimere in una formula il fatto che, data una formula ψ, "tutti sanno ψ". Aggiungiamo la seguente definizione:

$$\textit{Def: } E\psi \leftrightarrow \bigwedge_{i \in N} \Box_i \psi$$

e così $w \models E\psi$ se e solo se $w \models \Box_i \psi$ per ogni individuo i.

Molto più delicata è invece la definizione del concetto di conoscenza comune (common knowledge): poter esprimere che una certa proposizione è conoscenza comune a tutti si rivela di importanza fondamentale in numerose applicazioni, soprattutto nel caso in cui si vogliano analizzare i legami tra conoscenza ed azione.

Un classico esempio è lo studio della nascita di convenzioni all'interno di un gruppo di persone (studiato ampiamente in [Lew69]); il tipico caso è l'associare al verde di un semaforo il via libera, ed al rosso il fermarsi. Se anche supponessimo che all'interno del gruppo degli automobilisti queste associazioni fossero note a tutti ($E\psi$), un individuo difficilmente si fiderà a passare col verde, dato che non sa se queste regole sono note a tutti o meno. Ma anche se supponessimo che le regole sono note a tutti e che tutti sanno che queste regole sono note a tutti ($EE\psi$), neanche allora sarebbe sufficiente per considerarlo una convenzione (sebbene in questo modo gli automobilisti si sentono sicuri: nessuno passa col rosso). Infatti nessuno sa che tutti sanno che tutte le regole sono note a tutti, quindi si considera possibile l'esistenza di un individuo che non sa che tutti sanno le regole, e che dunque si fermerà per sicurezza al verde.

Questo regresso all'infinito suggerisce la seguente definizione:

- *Definizione informale:* $C\psi \Leftrightarrow \bigwedge_{k \in \mathbb{N}} E^k \psi$

dove con E^k si intende E ripetuto k volte.
Essendo una congiunzione infinita, questa definizione non è nè ammissibile nè operativa, non essendo traducibile in una definizione formale all'interno del linguaggio. Dimostriamo però alcune proprietà che si riveleranno utili nel suggerire un'assiomatizzazione effettiva per la conoscenza comune.
Se $\mathcal{F} = (\Omega, R_1, \ldots, R_n)$ è un frame per $\mathbf{S5}_n$, definiamo su $\mathcal{F}$ una nuova relazione R_C:

Definizione 2.1. Diremo che w è *raggiungibile* da v, e scriveremo wR_Cv, se esiste $k \in \mathbb{N}$ e k stati $w_1, \ldots, w_k$ in Ω, con $w_1 = w$ e $w_k = v$ tali che per ogni $1 \leqslant j \leqslant k-1$ esiste un $i \leqslant n$ tale che $w_j R_i w_{j+1}$.

La relazione R_C è la composizione iterata delle n relazioni di accessibilità del modello. Utilizzando la definizione informale per C, vale il seguente:

Lemma 2.2. *$(\mathcal{M}, w) \models C\psi$ se e solo se per ogni v raggiungibile da w si ha che $(\mathcal{M}, v) \models \psi$.*

Dimostrazione. Dimostriamo che $(\mathcal{M}, w) \models E^k\psi$ se e solo se per ogni v raggiungibile da w in k passi si ha che $(\mathcal{M}, v) \models \psi$.
Per induzione su k: se $k = 0$ non c'è nulla da dimostrare; supponiamo invece $(\mathcal{M}, w) \not\models E(E^{k-1})\psi$, allora esiste un i e uno stato w' tale che wR_iw' e $(\mathcal{M}, w') \models \neg E^{k-1}\psi$. Per ipotesi induttiva esiste un v raggiungibile in $k-1$ passi da w' tale che $(\mathcal{M}, v) \models \neg\psi$, e dunque v è raggiungibile in k passi da w.
Se invece supponiamo che esista v raggiungibile da w in k passi tale che $(\mathcal{M}, v) \models \neg\psi$, dalla definizione di E è immediato vedere che $(\mathcal{M}, w') \models \neg E^k\psi$. □

Definiamo R_E come l'unione delle R_i: wR_Ev se esiste un i tale che wR_iv. La relazione R_C è la chiusura transitiva di R_E, che è simmetrica e riflessiva, dunque R_C è una relazione di equivalenza; l'operatore C gode dunque delle stesse proprietà degli operatori $\Box_i$ di conoscenza. Inoltre:

Lemma 2.3. *Sia $\mathcal{F} = (\Omega, R_1, \ldots, R_n)$ un frame per $\mathbf{S5}_n$, per ogni formula ψ e φ:*
i) $\mathcal{F} \models C\varphi \leftrightarrow E(\varphi \wedge C\varphi)$;
ii) se $\mathcal{F} \models \psi \rightarrow E(\psi \wedge \varphi)$ allora $\mathcal{F} \models \psi \rightarrow C\varphi$.

Dimostrazione. Se $(\mathcal{M}, w) \models C\varphi$ allora φ è vera in ogni punto v raggiungibile da w; dunque se $\mathcal{F} \models C\varphi$ allora $\mathcal{F} \models \varphi$, e dunque anche $\mathcal{F} \models \Box_i\varphi$ ed $\mathcal{F} \models \Box_iC\varphi$, e dunque la tesi.
Viceversa se $E(\varphi \wedge C\varphi)$ è valida, fissato un modello e un generico stato w, dimostriamo che φ è vera per ogni v raggiungibile da w. Se $w = v$ si ottiene che $\varphi \wedge C\varphi$ è vera in w per riflessività delle R_i. Se invece v è raggiungibile in k passi da w, basta considerare w_{k-1}; dato che $E(\varphi \wedge C\varphi)$ è valida, è vera anche in w_{k-1} e questo implica che φ è vera in v.

Per quanto riguarda *ii*), supponiamo che $\mathcal{F} \models \psi \rightarrow E(\psi \wedge \varphi)$ e che $\mathcal{F} \models \psi$ e dimostriamo che $\mathcal{F} \models C\psi$. Sia $\mathcal{M}$ un modello su $\mathcal{F}$ w uno stato in Ω e v raggiungibile in k passi da w. Dimostriamo che φ è vera in v.
Sia $w_1 \ldots w_k$ la sequenza con $w_1 = w$ e $w_k = v$; se $k = 0$ da $w \models E(\psi \wedge \varphi)$ allora $w \models \varphi$ per riflessività delle R_i. Altrimenti si consideri w_{k-1}, per ipotesi $E(\psi \wedge \varphi)$ è vera e dunque ogni punto raggiungibile tramite R_i in un passo da w_{k-1} verifica $\psi \wedge \varphi$, dunque φ è vera in v. □

L'operatore $C\psi$ è dunque il punto fisso della mappa $f(\psi, x) = E(\psi \wedge x)$ che manda la formula x in $E(\psi \wedge x)$.

Riassumendo queste proprietà, possiamo enunciare un'assiomatizzazione adeguata ad includere nel vocabolario un operatore per la conoscenza comune:

- il linguaggio è $\mathcal{L}_n^C = (Var, \Box_i \mid i \leqslant n, C)$ con $\Diamond$ ed E operatori definiti.
- la teoria $\mathbf{S5}_n^C$ formata dagli assiomi in Tabella 2.1.2.

$\mathbf{S5}_n^C$	
$\mathbf{S5}_n$	...
K	$C(\varphi \to \psi) \to (C\varphi \to C\psi)$
T	$C\psi \to \psi$
4	$C\psi \to CC\psi$
5	$\neg C\psi \to C\neg C\psi$
$C1$	$C\psi \to E(\psi \wedge C\psi)$
$C2$	$(\psi \to E(\psi \wedge \varphi)) \to (\psi \to C\varphi)$
Regole	**MP**, *Nec*

Per quanto visto finora $\mathcal{F} = (\Omega, R_1, \ldots, R_n, R_C)$, dove R_C è la chiusura transitiva di R_E, è un frame per $\mathbf{S5}_n^C$. Inoltre si dimostra che:

Teorema 2.4 ([FHMV95], Teorema 3.3.1)**.** *$\mathbf{S5}_n^C$ è completa[3] rispetto alla classe di frame $(\Omega, R_1, \ldots, R_n, R_C)$ dove le R_i sono relazioni di equivalenza ed R_C la chiusura transitiva dell'unione delle R_i.*

Dunque tramite questa assiomatizzazione è posssibile tradurre la definizione informale data in precedenza.

Una teoria modale si dice decidibile se esiste una procedura effettiva per decidere se data una formula ψ del linguaggio è un dimostrabile oppure no. Se la teoria è completa rispetto ad una classe $\mathscr{F}$ di strutture, il problema è equivalente ad esibire una procedura effettiva per decidere se una formula ψ è valida su tutte le strutture di $\mathscr{F}$ o meno[4]. Nel nostro caso il problema ha risposta affermativa:

Teorema 2.5 ([FHMV95], Teorema 3.3.1)**.** *$\mathbf{S5}_n^C$ è decidibile.*

2.2 Strutture di Aumann e strutture di Kripke

Esponiamo ora il modello per la conoscenza utilizzato in teoria dei giochi. Alla base del modello c'è il solito insieme Ω di stati. Per ogni giocatore è

[3]Il metodo per dimostrare la completezza è simile a quello del modello canonico, e viene portato in [Aum99] come argomento a favore dell'interpretazione epistemica di queste teorie: dato ad un agente il vocabolario $\mathcal{L}$, il modello canonico di **S5** rappresenta esattamente l'insieme dei mondi immaginabili dal giocatore in questione tramite il vocabolario $\mathcal{L}$. Gli insiemi di formule consistenti massimali sono esattamente tutti i possibili mondi che può immaginare con il linguaggio a sua disposizione.

[4]Detto appunto validity problem, [FHMV95].

definita una funzione di informazione $\mathcal{P}_i : \Omega \longrightarrow \mathscr{P}(\Omega)$, che ad ogni stato w di Ω associa un sottoinsieme di stati indistinguibili dal giocatore i in w. Si richiede che le funzioni $\mathcal{P}_i$ soddisfino le seguenti proprietà:

P1: $w \in \mathcal{P}(w)$
P2: se $w \in \mathcal{P}(w')$ allora $\mathcal{P}(w) = \mathcal{P}(w')$

Questo implica che ogni funzione di informazione induce una partizione su Ω; indicheremo con $\mathcal{P}_i$ indifferentemente la partizione associata o la funzione di informazione, scrivendo $\mathcal{P}_i(w)$ per indicare il sottoinsieme della partizione contenente w.

Definizione 2.6. Una *Struttura di Aumann* $\mathcal{A} = (\Omega, \mathcal{P}_1, \ldots, \mathcal{P}_n)$ per n giocatori è un insieme Ω ed n partizioni $\mathcal{P}_1 \ldots \mathcal{P}_n$ su Ω.

Analogamente alla teoria della probabilità, gli oggetti della conoscenza in questo modello sono gli eventi: viene detto *evento* un sottoinsieme di Ω. Sullo spazio degli eventi (le parti di Ω) viene definita per ogni individuo una funzione conoscenza K_i, che ad ogni evento B associa l'insieme degli stati in cui il giocatore i *sa* che l'evento E accade.

$$\mathrm{K}_i : \mathscr{P}(\Omega) \longrightarrow \mathscr{P}(\Omega)$$

$$\mathrm{K}_i(B) = \{w \in \Omega \mid \mathcal{P}_i(w) \subseteq B\}$$

L'evento "i sa B" è dunque l'insieme degli stati il cui insieme di informazione è incluso in B. Si dimostra facilmente che K_i gode delle seguenti proprietà:

- è monotona: se $E \subseteq F$ allora $\mathrm{K}_i(E) \subseteq \mathrm{K}_i(F)$;
- $K(E \cap F) = K(E) \cap K(F)$;
- $K(E) \subseteq E$;
- $K(E) = K(K(E))$;
- $\Omega \setminus K(E) = \Omega \setminus K(K(E))$;

Se $\mathcal{P}$ è una partizione, diremo che la partizione $\mathcal{P}'$ è più fine se per ogni $w \in \Omega$ si ha che $\mathcal{P}'(w) \subseteq \mathcal{P}(w)$. Se $\mathcal{P}_1, \ldots, \mathcal{P}_n$ sono le partizioni informative, definiamo una nuova partizione $\mathcal{P}_C$ detta *partizione meet* come la partizione più fine tale che ogni $\mathcal{P}_i$ sia più fine di $\mathcal{P}_C$.
Essendo $\mathcal{P}_C$ una partizione, possiamo definire una funzione conoscenza che si riferisce ad essa:

$$\mathrm{C}(B) = \{w \in \Omega \mid \mathcal{P}_C(w) \subseteq B\}$$

Un evento si dice *conoscenza comune* in w se e solo se $w \in \mathrm{C}(B)$ (come definito in [Aum99]).

L'analogia con la logica epistemica definita alla sezione precedente è immediata. A livello di frame, le strutture di Aumann e le strutture di Kripke per $\mathbf{S5}_n^C$ sono esattamente la stessa cosa; se $\mathcal{A} = (\Omega, \mathcal{P}_1 \dots \mathcal{P}_n)$ è una struttura di Aumann, definiamo il frame $\mathcal{F}^A = (\Omega, R_1 \dots R_n)$ utilizzando lo stesso insieme Ω e le relazione di equivalenza associate alle partizioni $\mathcal{P}_i$: wR_iv se e solo se $\mathcal{P}_i(w) = \mathcal{P}_i(v)$. Viceversa ogni relazione di equivalenza R_i induce una partizione di Ω, associando ad ogni struttura di Kripke $\mathcal{F} = (\Omega, R_1 \dots R_n)$ la struttura di Aumann corrispondente $\mathcal{A}^{\mathcal{F}}$.

La corrispondenza è in realtà molto più profonda. Prendiamo un modello $\mathcal{M}$ di $\mathbf{S5}_n^C$, con la valutazione $V : Var \longrightarrow \mathscr{P}(\Omega)$ e sia $\mathcal{A}^{\mathcal{M}}$ la struttura di Aumann corrispondente: dimostriamo che la "semantica" delle due strutture è uguale, ciò che esprimono le formule in $\mathcal{L}$ può essere espresso allo stesso modo in termini di eventi e funzione di conoscenza, e viceversa.

Presa una formula ψ nel linguaggio $\mathcal{L} = (Var, \Box_i, C)$ definiamo l'insieme $\psi^{\mathcal{M}} = \{w \in \Omega \mid \mathcal{M}, w \models \psi\}$. La semantica in $\mathcal{A}^{\mathcal{M}}$ è definita invece in termini di eventi. Definiamo per ogni $p \in Var$ l'evento $B_p = V(p)$ e definiamo per induzione sulle formule gli eventi $ev(\varphi)$:

- $ev(p) = B_p$
- $ev(\varphi \wedge \psi) = ev(\varphi) \cap ev(\psi)$
- $ev(\neg\varphi) = \Omega \setminus ev(\varphi)$
- $ev(\Box_i\varphi) = \mathrm{K}_i(ev(\varphi))$
- $ev(C\varphi) = \mathrm{C}(ev(\varphi))$

Mostriamo che la formula φ vale in $w \in \Omega$ se e solo se $s \in ev(\varphi)$, ottenendo l'equivalenza cercata:

Proposizione 2.7. *Sia $\mathcal{M}$ un modello di $\mathbf{S5}_n^C$ e sia $\mathcal{A}^{\mathcal{M}}$ la struttura di Aumann corrispondente. Per ogni formula ψ si ha che $\psi^{\mathcal{M}} = ev(\psi)$.*

Dimostrazione. La dimostazione è immediata per i primi tre casi, il primo caso interessante è $(\Box_i\varphi)^{\mathcal{M}} = \mathrm{K}_i(ev(\varphi)$. Per ipotesi induttiva $\mathrm{K}_i(ev(\varphi)) = \mathrm{K}_i(\varphi^{\mathcal{M}})$; sia ora $w \in \mathrm{K}_i(\varphi^{\mathcal{M}})$: $\mathcal{P}_i(w) \subseteq (\varphi^{\mathcal{M}})$ dunque ogni vR_iw è tale che $(W, v) \models \varphi$ dunque $w \in (\Box_i\varphi)^{\mathcal{M}}$ e dunque $\mathrm{K}_i(ev(\varphi)) \subseteq (\Box_i\varphi)^{\mathcal{M}}$.
Viceversa se $w \in (\Box_i\varphi)^{\mathcal{M}}$ allora ogni vR_iw verifica φ, quindi $\mathcal{P}_i(w) \subseteq ev(\varphi)$ e così $w \in \mathrm{K}_i(ev(\varphi)$.
Per quanto riguarda l'ultimo caso, per ipotesi induttiva $C_i\varphi^{\mathcal{M}} = \mathrm{C}(\varphi^M)$ e per definizione $\mathrm{C}(\varphi^M) = \{w \in \Omega \mid \mathcal{P}_C(w) \subseteq \varphi^M\} = \{w \mid \mathcal{M}, v \models \varphi \forall v \in \mathcal{P}_C(w)\}$. Se dimostriamo allora che $v \in \mathcal{P}_C(w)$ se e solo se v è raggiungibile da w, allora si ottiene la tesi; infatti se φ è vera per ogni v raggiungibile da w si ha che $\mathcal{M}, w \models C\varphi$.

Supponiamo che v sia raggiungibile da w, dunque esistono $w_1 = w, \ldots, w_k = v$ tali che $wR_{i_1}w_1R_{i_2}\ldots w_{k-1}R_{i_k}v$; per induzione su k mostriamo che $v \in \mathcal{P}_C(w)$. Se $k = 1$ non cè nulla da dimostrare; per v è raggiungibile in k passi, per ipotesi induttiva $w_{k-1} \in \mathrm{C}(w)$. Ma $v \in \mathcal{P}_{i_k}(v) = \mathcal{P}_{i_k}(w_{k-1})$ che per proprietà del meet $P_{i_k}(w_{k-1} \subseteq \mathcal{P}_C(w)$ dunque $v \in \mathcal{P}_C(w)$. Per l'inclusione inversa supponiamo $v \in \mathcal{P}_C(w)$. Allora v deve essere raggiungibile da w per minimalità di $\mathcal{P}_C$. □

Il procedimento contrario, partendo invece da una struttura di Aumann $\mathcal{A}$, richiede che venga definito (a seconda della situazione che si sta rappresentando) un insieme di eventi base $\mathcal{B} \subseteq \mathcal{P}(\Omega)$.
Possiamo così definire $\mathcal{M}^{\mathcal{A}}$ come il modello di Kripke corrispondente associando ad ogni $B \in \mathcal{B}$ una variabile p_B, e definendo $V(p_B) = B$. Per la proposizione precedente le formule in questo linguaggio e gli eventi ad esse associati coincidono: $\mathcal{M}^{\mathcal{A}}, s \models \varphi \Leftrightarrow s \in ev(\varphi)$.

Capitolo 3

Interpretazione normativa

Nell'interpretazione normativa un concetto risolutivo è visto come una raccomandazione teorica che indica a giocatori razionali un insieme di esiti "ideali" del gioco.

La domanda fondazionale è: quali raccomandazioni possono essere accettate e seguite da giocatori razionali (nel nostro caso massimizzatori di utilità)? La proposta è la seguente nozione di consistenza: assumendo che tutti i giocatori seguano la raccomandazione, dunque ricadendo nel caso della teoria della decisione "semplice", a nessun giocatore deve essere raccomandata un'azione irrazionale.

La situazione è ben riassunta da Selten:

"The modern game theoretical interpretation of equilibrium points in the sense of Nash (1951)...is based on the idea that a rational theory should not be a self-destroying prophecy which creates an incentive to deviate for those who believe in it" ([Sel85])

sviluppando un'idea di consistenza già introdotta nell'ambito dei giochi cooperativi da von Neumann e Morgenstern in [vNM47].

3.1 Giochi in forma strategica

3.1.1 Dai giochi ai modelli di Kripke

Ad ogni gioco in forma strategica ad n giocatori G possiamo associare una struttura di Kripke $\mathcal{F}_G = (\Omega, R_1, \ldots, R_n)$ nella seguente maniera:

- come insieme degli stati Ω l'insieme degli esiti $\prod_{i \in N} A_i$;
- una relazione per ogni giocatore, con l'interpretazione che sR_it se il giocatore i-esimo ha un azione a sua disposizione per passare dallo

stato s allo stato t: $(s_1, \ldots, s_n) R_i (t_1, \ldots, t_n)$ se e solo se $s_j = t_j$ per ogni $j \neq i$.

In questo modo viene rappresentata la struttura strategica del gioco. Per poter rappresentare anche le utilità (equivalentemente le preferenze) è necessario passare al livello dei modelli, definendo una valutazione ed un opportuno insieme di variabili.
Si possono seguire due strade equivalenti:

I) la prima consiste nel considerare un linguaggio numerabile unico per tutti i giochi. Per il Teorema 1.6, possiamo considerare funzioni di utilità a valori in $\mathbb{Q}$, dato che l'insieme degli esiti è finito, e definire l'insieme Var composto dalle seguenti formule atomiche:

   i) $u_i = p$ per ogni $i \in N$ e per ogni $p \in \mathbb{Q}$;
   ii) $p \leqslant q$ per ogni $p, q \in \mathbb{Q}$.

   Dato un gioco in forma strategica $G = \langle N, A_i, u \rangle$ possiamo definire un modello basato sul frame $\mathcal{F}_G$, definendo una valutazione $V : Var \longrightarrow \mathscr{P}(\Omega)$ che rappresenti le utilità del gioco:

   Definizione 3.1. Dato un gioco finito in forma strategica $G = \langle N, A_i, u \rangle$ diciamo *modello di* G il seguente modello di Kripke $\mathcal{M}_G = (\Omega, R_1, \ldots, R_N, V)$ per il linguaggio $\mathcal{L} = \{Var, \Box_i\}$, dove:

   - $\Omega = \prod_{i \in N} A_i$;
   - $t R_i s$ se e solo se $t_j = s_j$ per ogni $j \neq i$;
   - $s \in V(u_i = p)$ se e solo se $u_i(s) = p$;
   - se $p \leqslant q$ in $\mathbb{Q}$ allora, per ogni $w \in \Omega$, $w \in V(p \leqslant q)$; altrimenti $V(p \leqslant q) = \emptyset$.

   Lo svantaggio di questo approccio è che soltanto un numero ristretto di formule atomiche sono rilevanti al fine dell'analisi del gioco; la maggior parte di esse è falsa in ogni nodo di Ω.

II) Il secondo approccio[1] ha il vantaggio di utilizzare un linguaggio finito, ma non unico per tutti i giochi. Dato il gioco finito $G = \langle N, A_i, u \rangle$, possiamo restringerci alle sole variabili rilevanti definendo il sottoinsieme finito $val \subseteq \mathbb{Q}$ come l'unione delle immagini delle u_i. L'insieme Var è composto da:

   i) $u_i = p$ per ogni $i \in N$ e $p \in val$;
   ii) $p \leqslant q$ per ogni $p, q \in val$.

   La valutazione V è definita esattamente come nel caso del linguaggio infinito.

[1] Suggerito in [vdHP06].

Indicheremo con $\mathcal{L}_s$ il linguaggio composto dall'insieme Var definito in uno dei due modi appena esposti e dal vocabolario $\{\Box_i \mid i \in N\}$. In generale verrà privilegiato il linguaggio finito.

Vediamo in un esempio come in questo linguaggio si possono esprimere alcuni concetti riguardanti la struttura strategica di un gioco, descrivendo il modello $\mathcal{M}_G$ associato al gioco *carta, forbice, sasso*:
$\Omega = \{C, F, S\} \times \{C, F, S\}$; le R_i e la V sono definite come alla Definizione 3.1, dunque $(a_1, a_2)R_1(b_1, b_2)$ sse $a_2 = b_2$, e allo stesso modo per R_2 con gli indici invertiti; l'insieme var è in questo caso $\{-1, 0, 1\}$. In ogni stato (a_1, a_2) di $\mathcal{M}_G$ la formula

$$\Box_1 \Diamond_2 (u_2 = 1)$$

è vera. Infatti qualsiasi azione faccia I in ogni stato (dunque $\Box_1$), II ha a disposizione un'azione che gli permette di vincere, dunque uno stato accessibile dove $u_2 = 1$ è vera. Visto che il gioco è simmetrico, analogamente la formula

$$\Box_2 \Diamond_1 (u_1 = 1)$$

è globalmente vera su $\mathcal{M}_G$.

Aumentiamo il modello inserendo i concetti risolutivi, che nell'interpretazione normativa indicano per ogni gioco un sottoinsieme di esiti ideali per tutti i giocatori.

Definizione 3.2. Dato un gioco G diciamo *modello normativo* di G il modello $(\mathcal{M}_G, R)$ dove $\mathcal{M}_G$ è il modello di G e R un sottoinsieme di Ω.

3.1.2 Consistenza interna ed equilibrio di Nash

In un modello normativo $(\mathcal{M}_G, R)$ R rappresenta il sottoinsieme degli esiti che possono essere raccomandati come esiti ideali: ogni $s \in R$ è una raccomandazione ad ogni giocatore su quale azione intraprendere. Con un abuso di notazione indicheremo il sottoinsieme R come una "raccomandazione per G", intendendo che ogni suo elemento è una raccomandazione possibile[2].

Come anticipato nell'introduzione al capitolo, una raccomandazione è consistente con l'ipotesi di razionalità se a nessun giocatore è raccomandato compiere scelte irrazionali, prevedendo il comportamento degli altri giocatori tramite la raccomandazione stessa. Questa nozione di consistenza è stata introdotta da von Neumann e Morgenstern in [vNM47], ed è l'argomento centrale di [Gre90] nello sviluppo della teoria delle situazioni sociali: spostare

[2]Non è possibile infatti raccomandare a tutti i giocatori più di un esito. Si prenda l'esempio BoS: raccomandando ai singoli giocatori di giocare uno dei due equilibri di Nash si può ottenere il profilo (B, S), fuori dalla raccomandazione.

le assunzioni di razionalità e coerenza dai giocatori (dove più ardua è la giustificazione, come accennato al capitolo precedente) alla teoria stessa.

Nel nostro caso i giocatori sono massimizzatori di utilità, dunque fissato il comportamento degli altri giocatori tramite la raccomandazione la scelta razionale è il massimo della loro utilità. Dunque un esito $s \in R$ è accettabile se nessun giocatore ha un'azione che gli permette di ottenere utilità maggiore, fissato il comportamento degli altri.
Questo fatto è esprimibile in una formula modale nel linguaggio esposto alla sezione precedente:

$$[IC] \qquad \bigwedge_{(p_1,\dots,p_n)\in val^n} {}_{\wedge}(u_i = p_i) \to \bigwedge_i \Box_i [\bigwedge_{q\in Val} (u_i = q \to q \leqslant p_i)] \qquad (3.1)$$

Indicando con ${}_{\wedge}(u_i = p_i)$ la congiunzione $(u_1 = p_1) \wedge \dots \wedge (u_n = p_n)$, per generici $p_i \in val$.

Definizione 3.3. Una raccomandazione R per G si dice *internamente consistente* se $\mathcal{M}_G, s \models IC$ per ogni $s \in R$.

Leggendo la formula si nota una strettissima somiglianza con la definizione di equilibrio di Nash. Definiamo infatti per ogni gioco in forma strategica G la raccomandazione $R^N = N(G)$ su M_G associata ad N. Ogni elemento di R^N raccomanda che venga raggiunto un equilibrio di Nash.

Proposizione 3.4. *Per ogni gioco G la raccomandazione R^N è internamente consistente.*

Dimostrazione. È sufficiente leggere la formula IC: la congiunzione ${}_{\wedge}(u_i = p_i)$ è verificata per una sola n-pla di valori, dunque basta verificare la seconda parte dell'implicazione solamente per questi valori. Ma se s è un equilibrio di Nash, per la definizione 1.9 ogni giocatore deviando dall'equilibrio ha utilità minore, dunque la formula è verificata. □

Se R è una generica raccomandazione è immediato anche il risultato inverso:

Proposizione 3.5. *Se R è internamente consistente allora $R \subseteq R^N$.*

Infatti se R è internamente consistente allora ogni $s \in R$ verifica la formula IC; ciò implica che $u(s)$, l'unica n-pla di valori per cui ${}_{\wedge}(u_i = p_i)$ è vera, verifica la seconda parte dell'implicazione in IC, che è equivalente alla definizione di equilibrio di Nash.

Ogni concetto risolutivo F associa ad ogni gioco G una raccomandazione $R^F = F(G)$ su $\mathcal{M}_G$. Diremo che un concetto risolutivo è internamente consistente se (M_G, R^F) è internamente consistente per ogni G.
Dalle proposizioni precedenti segue immediatamente che:

Proposizione 3.6. *Un concetto risolutivo in strategie pure è internamente consistente se e solo se per ogni G si ha che $F(G) \subseteq N(G)$, dove $N(G)$ sono gli equilibri di Nash di G.*

L'equilibrio di Nash è dunque il concetto risolutivo internamente consistente massimale.

3.2 Giochi in forma estesa

Un concetto risolutivo per giochi in forma estesa associa ad ogni gioco un insieme di profili di strategie, ossia mappe definite su ogni nodo dell'albero. Per l'interpretazione normativa questo profilo rappresenta una raccomandazione a ciascun giocatore su come giocare ad *ogni* nodo.

La consistenza di una raccomandazione si traduce in questo caso in una condizione su tutto l'albero di gioco: ad ogni nodo, prevedendo il comportamento degli altri giocatori tramite il concetto risolutivo, nessun giocatore deve compiere scelte irrazionali.

In questa sezione modellizziamo una raccomandazione per un gioco in forma estesa come una previsione sull'albero di gioco, sfruttando una teoria modale già utilizzata come modello per previsioni su strutture ad albero (come presentata in [Bon01b]). Verranno ottenuti risultati analoghi a quelli esposti alla sezione precedente, considerando come concetto risolutivo l'induzione a ritroso (ossia gli equilibri perfetti nei sottogiochi).

3.2.1 Logica temporale ad albero con giocatori

Per tradurre la struttura sequenziale dei giochi in forma estesa utilizziamo un'estensione della logica temporale detta logica temporale ad albero[3].
Lo scopo della logica temporale è quello di proporre un modello che rappresenti situazioni e concetti legati allo scorrimento del tempo. Un primo modello temporale si può ottenere considerando un insieme Ω e una relazione binaria $\prec$ su Ω, rappresentando il flusso temporale come un insieme di momenti legati da una relazione sucessore.
Alle diverse proprietà di $\prec$ è lasciato il compito di rappresentare diverse tipologie di situazione temporale: ad esempio un flusso temporale lineare con un ordine totale $\prec$, successioni temporali distinte se è un ordine parziale, fino a situazioni molto più complicate come avremo modo di vedere ai paragrafi successivi.

Si possono associare due diverse modalità temporali riferendosi alla stessa relazione di successione $\prec$: "sempre sarà che" se ci si riferisce al futuro, "è sempre stato che" se invece si guarda nel verso opposto e si considera $\prec^{-1}$.

[3] I risultati in questa sezione sono un riadattamento di risultati distribuiti in [BdRV01].

Il primo è indicato in letteratura con G[4] e scriveremo $[G]$ per ricordare che è un operatore di tipo Box, ossia che $[G]\varphi$ è vera se φ è definitivamente vera in ogni nodo accessibile tramite $\prec$. Il secondo operatore viene indicato con $[H]$[5], e dunque $[H]\psi$ indica che "è sempre stato vero che ψ".

Definiamo linguaggio temporale di base $\mathcal{L}_t = \{Var, [G], [H]\}$, e definiamo gli operatori di tipo diamond: $\langle F\rangle p \leftrightarrow \neg[G]\neg p$ e $\langle P\rangle p \leftrightarrow \neg[H]\neg p$, ad indicare "$p$ succederà in futuro" e "p è successo nel passato".

Osservazione: a priori una struttura di Kripke per $\mathcal{L}_t$ è fornita di due relazioni R_G ed R_H, ma dall'interpretazione di $[G]$ ed $[H]$ che abbiamo dato si richiede che $R_H = R_G^{-1}$. Seguendo la notazione di [BdRV01], chiameremo frame bidirezionale un frame $\mathcal{F}(\Omega, R_G, R_H)$ tale che $R_G = R_H^{-1}$, e scriveremo solamente $\mathcal{F}(\Omega, R_G)$, dato che R_H è definita a partire da R_G.

Definiamo la minima logica temporale $\mathbf{K}_t$ formata dai seguenti assiomi:

$\mathbf{K}_t$	
$A1$	$p \rightarrow [G]\langle P\rangle p$
$A2$	$p \rightarrow [H]\langle F\rangle p$

In cui non sono riportati gli assiomi logici per entrambi gli operatori (in questo caso K_G: $[G](p \rightarrow q) \rightarrow ([G]p \rightarrow [G]q)$ e K_H: $[H](p \rightarrow q) \rightarrow ([H]p \rightarrow [H]q)$).
Gli assiomi $A1$ e $A2$ esprimono il legame tra le due modalità temporali: se p è vera in un nodo allora da una parte sarà sempre vero che p è accaduto, dall'altra è sempre stato vero che p fosse possibile.

Definiamo ora una classe di strutture bidirezionali adeguate a rappresentare diverse possibilità future ad ogni nodo:

Definizione 3.7. Definiamo *struttura temporale ad albero* (che chiameremo BT[6]-struttura) un frame bidirezionale $\mathcal{F} = (\Omega, \prec)$ tale che la relazione binaria $\prec$ gode delle seguenti proprietà:
($P0$) è irriflessiva;
($P1$) è transitiva;
($P2$) è lineare a ritroso: se $t_1 \prec s$ e $t_2 \prec s$ allora o $t_1 = t_2$ o $t_1 \prec t_2$ o $t_2 \prec t_1$.

Il risultato è un albero transitivo $(\Omega, \prec)$: le ramificazioni indicano in ogni nodo i possibili svolgimenti lineari del futuro, e seguendo la relazione $\prec$

[4] "it is always Going to be the case".
[5] "It Has always been the case".
[6] Branching time frame.

lungo un ramo si ritrova lo scorrimento lineare del tempo. Infatti $P2$ implica che per collegare due punti di Ω esiste un solo cammino.

Aggiungiamo a questa definizione n relazioni $R_1, \ldots, R_n$ a rappresentare le azioni a disposizione di ogni giocatore nei vari istanti di tempo:

Definizione 3.8. Definiamo *struttura temporale ad albero multiagente* (BTA[7]-struttura) una struttura $\mathcal{F} = (\Omega, \prec, R_1, \ldots, R_n)$ tale che:

- $(\Omega, \prec)$ sia una struttura temporale ad albero;
- per ogni $t, s \in \Omega$, se tR_is allora $t \prec s$.

L'interpretazione di tR_is è che al nodo t l'agente i può compiere un azione per spostarsi in s; l'unica condizione imposta alle relazioni R_i è quella di essere sottorelazioni di $\prec$, ossia che le azioni di un giocatore possano influenzare soltanto il futuro. In particolare l'insieme $\{s/tR_is\}$ delle azioni disponibili ad i al nodo t può essere vuoto, esprimendo il fatto che il giocatore i non ha alcuna influenza al nodo t.

Estendiamo il linguaggio temporale con n operatori modali $\Box_i$, uno per ogni individuo, ad esprimere con $\Box_i\varphi$ "qualsiasi azione compia i si ha che φ", dunque $\mathcal{L}_{BTA} = \{Var, [G], [H], \Box_i\}$.

Definizione 3.9. Diciamo *logica temporale ad albero con giocatori*, in breve **BTA**[8], la logica temporale nel linguaggio $\mathcal{L}_{BTA}$ composta dai seguenti assiomi (al solito sottintendendo gli assiomi logici):

BTA	
$\mathbf{K}_t$	$p \rightarrow [G]\langle P\rangle p$
	$p \rightarrow [H]\langle F\rangle p$
4	$[G]p \rightarrow [G][G]p$
$\mathbf{.3_l}$	$(\langle P\rangle p \wedge \langle P\rangle q) \rightarrow [\langle P\rangle(p \wedge q) \vee \langle P\rangle(p \wedge \langle P\rangle q) \vee \langle P\rangle(\langle P\rangle p \wedge q)]$
$\subseteq$	$[G]p \rightarrow \Box_i p$

Innanzitutto dimostriamo che:

Proposizione 3.10. BTA *è corretta rispetto alla classe dei BTA-frame.*

Dimostrazione. Una BTA-struttura $\mathcal{F} = (\Omega, \prec, R_1, \ldots, R_n)$ è un frame bidirezionale, dato che i due operatori $[G]$ e $[H]$ si riferiscono alla stessa relazione $\prec$. Mostriamo che $A1$ ed $A2$ sono validi su ogni frame bidirezionale.
Sia $\mathcal{M} = (\Omega, R_G, R_H, V)$ un modello basato su un frame bidirezionale ($R_H =$

[7] Branching time with agents.
[8] Nella notazione di [BdRV01] sarebbe $\mathbf{K}_t 4.3_l A$.

R_G^{-1}), e sia t un punto del modello. Supponiamo che $\mathcal{M}, t \models p$: per ogni sR_Gt si ha che tR_Hs per bidirezionalità, dunque $\mathcal{M}, s \models \langle P\rangle p$; dunque $A1$ è valido. Allo stesso modo si dimostra la validità di $A2$.
4 è valido su frame transitivi per la dimostrazione della Proposizione 1.23.
Dimostriamo ora che $\mathbf{.3_l}$ è valido su ogni struttura che verifichi $(P2)$: prendiamo un modello che la verifichi e supponiamo che in un nodo s si abbia che $\mathcal{M}, s \models \langle P\rangle p \wedge \langle P\rangle q$. Allora esistono t_1 e t_2 tali che $t_1 \prec s$ e $t_2 \prec s$ tali che $\mathcal{M}, t_1 \models p$ e $\mathcal{M}, t_2 \models q$. Per $(P2)$, deve necessariamente essere $t_1 = t_2$ o $t_1 \prec t_2$ o $t_2 \prec t_1$. Nel primo caso $\mathcal{M}, s \models \langle P\rangle(p \wedge q)$, nel secondo $\mathcal{M}, t_2 \models \langle P\rangle p$ e dunque $\mathcal{M}, s \models \langle P\rangle(q \wedge \langle P\rangle p)$. Allo stesso modo nel terzo caso $\mathcal{M}, s \models \langle P\rangle(p \wedge \langle P\rangle q)$, dunque $\mathbf{.3_l}$ è valido.
La correttezza di **BTA** segue dal Lemma 1.22. □

Per ottenere anche la completezza utilizziamo il metodo del modello canonico come esposto alla Sezione 1.3. Dimostriamo innanzitutto che il modello canonico di ogni teoria contenente la minima logica temporale $\mathbf{K}_t$ è un frame bidirezionale, e che dunque i due assiomi $A1$ ed $A2$ sono canonici. Dimostrando la canonicità (Definizione 1.31) dei restanti assiomi di **BTA** otteniamo la tesi.

Proposizione 3.11. BTA *è canonica dunque completa rispetto alla classe delle BTA-strutture.*

Dimostrazione. Il modello canonico $\mathcal{M}^c = (\Omega, R_G, R_H, R_1, \ldots, R_n, V)$ di **BTA** è così composto:

- Ω i sottoinsiemi consistenti massimali di formule in $\mathcal{L}_{BTA}$;
- sR_Gt sse $[G]^{-1}s \subseteq t$;
- sR_Ht sse $[H]^{-1}s \subseteq t$;
- sR_it sse $\Box_i^{-1} \subseteq s$.
- $s \in V(p)$ sse $p \in s$.

Dimostriamo innanzitutto che $R_H = R_G^{-1}$. Supponiamo che sR_Ht, ossia $[H]^{-1}s \subseteq t$; ora se $\varphi \in s$ allora per $A2$ e chiusura per modus ponens anche $[H]\langle F\rangle\varphi \in s$. Ma allora $\langle F\rangle\varphi \in t$, e per generalità di φ otteniamo che tR_Gs. Utilizzando $A1$ si ottiene analogamente che se sR_Gt allora tR_Hs. Indichiamo dunque con $\prec$ la relazione R_G e $\prec^{-1} = R_H$.
Grazie alla Proposizione 1.33, sappiamo già che $[G]p \rightarrow [G][G]p$ è canonico. Mentre per quanto riguarda $.3_l$ è canonico rispetto a (P2) (cfr. ad esempio [BdRV01], pag.208).
Resta da dimostrare che $\subseteq$ è canonico, ma la dimostrazione è immediata. Infatti se tR_is allora $\Box_i^{-1}(t) \subseteq s$, ma dall'assioma $\subseteq$ sappiamo che $[G]^{-1}(t) \subseteq \Box_i^{-1}(s)$ e dunque si ottiene che $[G]_i^{-1}(t) \subseteq s$ ossia che $t \prec s$.
Non ci siamo curati sinora di richiedere che il frame canonico sia irriflessivo.

Non esiste una formula modale che sia vera su tutti i frame irriflessivi[9], ma è possibile restringersi a questa classe grazie al seguente fatto: ogni frame $\mathcal{F}$ riflessivo è *bisimulabile*[10] ad un frame $\mathcal{F}'$ irriflessivo. Abbiamo dimostrato la completezza di **BTA** rispetto a classe di frame $\mathscr{F}$, e **BTA** è corretta restringendosi ai frame irriflessivi di $\mathscr{F}$ (i BTA frames). Ma allora se φ è valida su ogni BTA-frame, allora per bisimulabilità φ è valida su ogni frame di $\mathscr{F}$, e per completezza ciò implica che φ è dimostrabile da **BTA**.
Dunque **BTA** è completa sulla classe delle BTA-strutture. □

3.2.2 Dai giochi in forma estesa ai modelli

Data una struttura di gioco in forma estesa[11] $G = \langle N, (\Omega, \rightarrowtail), \iota \rangle$ possiamo associare un BTA-frame $\mathcal{F}_G = (\Omega, \prec, R_1, \ldots, R_n)$ corrispondente definendo:

- Ω l'insieme dei nodi;
- $\prec = (\rightarrowtail)^*$ la chiusura transitiva di $\rightarrowtail$, che risulta essere transitiva, irriflessiva, e lineare a ritroso dato che $(\Omega, \rightarrowtail)$ è un albero;
- tR_is se e solo se $\iota(t) = i$ e $t \rightarrowtail s$.

Allo stesso modo che per i giochi in forma strategica utilizziamo una valutazione V per rappresentare le utilità dei diversi giocatori. Anche in questo caso si possono seguire due approcci:

I) definire un linguaggio numerabile unico per tutti i giochi. L'insieme dei nodi di un gioco in forma estesa è al più numerabile, dunque per il Teorema 1.6 possiamo considerare funzioni di utilità a valori in $\mathbb{Q}$. Definiamo l'insieme Var composto dalle seguenti formule atomiche:

   i) $u_i = p$ per ogni $i \in N$ e per ogni $p \in \mathbb{Q}$;
   ii) $p \leqslant q$ per ogni $p, q \in \mathbb{Q}$.

   Dato un gioco in forma estesa $G = \langle N, (\Omega, \rightarrowtail), \iota, u \rangle$ definiamo *modello di* G il modello $\mathcal{M}_G$ basato su $\mathcal{F}_G = (\Omega, \prec, R_1, \ldots, R_n)$ con valutazione $V : Var \longrightarrow (\Omega)$ che rappresenta le utilità del gioco:

   - $w \in V(u_i = p)$ se e solo se $w \in Z$ e $u_i(w) = p$;
   - $w \in V(p \leqslant q)$ se e solo se $p \leqslant q$ è vero in $\mathbb{Q}$.

   Rimane lo svantaggio la maggior parte delle formule è falsa in ogni nodo di $\mathcal{M}_G$.

II) Il caso del linguaggio finito. Dato il gioco finito $G = \langle N, (\Omega, \rightarrowtail), \iota, u \rangle$ definiamo il sottoinsieme finito $val \subseteq \mathbb{Q}$ come l'unione delle immagini delle u_i. L'insieme Var è composto da:

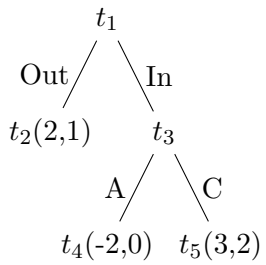
[9] Si veda ad esempio [HC96], Cap.10.
[10] Per la definizione, ad esempio, [BdRV01] Cap.2.
[11] Definizione 1.12, un gioco in forma estesa senza che siano espresse le utilità.

i) $u_i = p$ per ogni $i \in N$ e $p \in val$;
ii) $p \leqslant q$ per ogni $p, q \in val$.

La valutazione V è definita esattamente come nel caso del linguaggio infinito.

Nella maggior parte dei casi utilizzeremo il linguaggio finito. Avremo comunque modo di confrontare i due approcci nelle sezioni successive.

Mostriamo in un esempio la potenza espressiva di questo linguaggio analizzando la struttura sequenziale del *più semplice gioco di entrata* come esposto alla Sezione 1.1.4.
L'insieme Ω in questo caso è composto da 5 nodi che denoteremo con $\{t_1, t_2, t_3, t_4, t_5\}$, la relazione $\prec$ è la chiusura transitiva di $\rightarrowtail$, $R_1 = \{Out = (t_1, t_2), In = (t_1, t_3)\}$ mentre $R_2 = \{A = (t_3, t_4), C = (t_3, t_5)\}$. La valutazione è definita come sopra sull'insieme $Var = \{u_i = 2, u_i = -2, u_i = 0, u_i = 1, u_i = 3, p \leqslant q$ con $p, q \in \{-2, 0, 1, 2, 3\}\}$.



La seguente formula:

$$\mathcal{M}, t_1 \models \Diamond_1(u_1 = 2) \wedge \langle F \rangle \Diamond_2(u_1 = -2)$$

esprime il fatto che, in t_1, I ha un azione che gli permette di garantirsi utilità 2, e che in un nodo successivo II ha la possibilità di minacciarlo con un guagagno negativo di -2. D'altra parte:

$$\mathcal{M}, t_3 \models \Box_2 \langle P \rangle \Diamond_1(u_1 = 2)$$

dice che qualsiasi azione faccia II in t_3, I ha un azione in un nodo precedente che gli permette di garantirsi 2.

3.2.3 Una teoria per i giochi

I modelli di gioco (nel linguaggio infinito) sono modelli finiti di **BTA**, tali che le relazioni R_i provengano da una funzione di turni ι, e tali che le valutazioni

delle variabili sono ristrette ai soli nodi terminali (equivalentemente, che tutte le formule atomiche siano false al di fuori dei nodi terminali).

Nel tentativo di trovare una teoria che dimostri tutte e sole le validità di questi modelli, si possono aggiungere i seguenti assiomi:

$(R1)$: $\Diamond_i p \rightarrow \bigwedge_{i \neq j} \Diamond_j p$;
$(R2)$: $\Diamond_i[(\langle F \rangle q \wedge \neg \langle F \rangle \langle F \rangle q) \rightarrow \Diamond_i q]$;

per ogni i, che caratterizzano completamente le proprietà delle R_i nei modelli provenienti da giochi in forma estesa. È necessario anche aggiungere i seguenti assiomi per la valutazione V:

$(V1)$: se $p \leqslant q$ $(p \leqslant q) \leftrightarrow \top$;
$(V2)$: se $p \not\leqslant q$ $(p \leqslant q) \leftrightarrow \bot$;
$(V3)$: $(u_i = q) \rightarrow \neg(u_i = p)$ per ogni i per ogni $p \neq q$.

La classe dei giochi in forma estesa ad n giocatori si può caratterizzare utilizzando la nozione di "general frame" ([BdRV01], pag.28) come strutture $\mathcal{F}_G = (\Omega, \prec, R_1, \ldots, R_n, \mathscr{P}(Z))$ tali che $\prec$ è irriflessiva, transitiva e lineare a ritroso, le R_i provengono da una funzione ι come alla Sezione 3.2.2 e $\mathscr{P}(Z)$ (con Z i nodi terminali) sono gli insiemi ammissibili per la valutazione V. È ragionevole supporre che **BTA** con l'aggiunta degli assiomi sopra menzionati sia completa rispetto alla classe dei giochi in forma estesa.

3.2.4 Previsioni

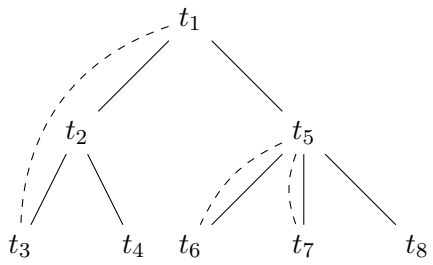
Le strutture definite finora permettono di rappresentare correttamente situazioni temporali in cui si vuole descrivere e valutare diversi possibili "futuri" (le ramificazioni) in un determinato momento. Presentiamo ora una teoria utilizzata come modello per previsioni su strutture ad albero, come definita in [Bon01b]. L'idea consiste nell'inserire una seconda relazione $\prec_p$ inclusa in $\prec$, che indichi in ogni punto quali tra i diversi nodi futuri vengono previsti. $\prec_p$ verifica tutte le proprietà di $\prec$ più una nozione di consistenza che caratterizza le previsioni:

Definizione 3.12 ([Bon01a])**.** Data una struttura temporale ad albero $\mathcal{F} = (\Omega, \prec, R_1, \ldots, R_n)$ una *previsione* è una relazione binaria $\prec_p$ su Ω tale che:

$(P4)$ sia una sottorelazione di $\prec$: se $t_1 \prec_p t_2$ allora $t_1 \prec t_2$;
$(P5)$ sia transitiva: $t_1 \prec_p t_2$ e $t_2 \prec_p t_3$ allora $t_1 \prec_p t_3$;
$(P6)$ sia definita ovunque è definita $\prec$: se $t_1 \prec t_2$ per qualche t_2 allora $t_1 \prec_p t_3$ per qualche t_3;
$(P7)$ verifichi la seguente condizione di consistenza temporale: se $t_1 \prec t_2$ e $t_2 \prec t_3$ e $t_1 \prec_p t_3$ allora $t_1 \prec_p t_2$ e $t_2 \prec_p t_3$.

La prima proprietà richiede che l'insieme dei nodi previsti da un nodo t sia un sottoinsieme dei nodi concepibili, dunque appartenenti ad un ramo

che parte da t. Non si richiede nulla sull'unicità della previsione: nè che l'insieme dei nodi previsti $\{s/t \prec_p s\}$ sia un unico ramo, nè che sia un sottoinsieme proprio dei nodi concepibili (la previsione vaga $\prec_p = \prec$ non viene scartata dalla definizione).
L'interpretazione di $\prec_p$ come previsione rende la transitività un requisito standard, mentre la proprietà $(P6)$, detta di serialità, richiede che la previsione sia completa: se l'insieme dei nodi futuri da w non è vuoto allora si richiede che su questo insieme sia fatta una previsione[12].
L'ultima proprietà viene introdotta in [Bon01b] come proprietà di consistenza temporale. Supponiamo infatti che al nodo t_1 un futuro concepibile sia $t_1 \prec t_2 \prec t_3$ (quindi t_1, t_2 e t_3 siano i vertici di un ramo che parte da t_1) e che dal nodo t_1 la previsione indichi il nodo t_3. La situazione è rappresentata nell'albero sottostante, dove le frecce continue indicano $\prec$, mentre le frecce tratteggiate la previsione $\prec_p$.



La proprietà $P7$ impone che:

i) dato che per raggiungere t_3 da t_1 è necessario passare per t_2 (il cammino è unico sull'albero), allora t_2 deve appartenere alla previsione fatta da t_1;

ii) il passaggio da t_1 a t_2 è una parziale realizzazione della previsione $t_1 \prec_p t_3$, e dunque la stessa previsione deve continuare a valere in t_2.

Definiamo BTA_p*-struttura* l'insieme di una BTA-struttura e di una previsione $\prec_p$:

$$\mathcal{F} = (\Omega, \prec, R_1, \ldots, R_n, \prec_p).$$

Dimostriamo i seguenti risultati nel caso senza giocatori, definendo BT_p-*struttura* una BT-struttura e una previsione $\prec_p$.
Il linguaggio $\mathcal{L}_{BT_p}$ estende il linguaggio temporale di base aggiungendo due nuovi operatori temporali basati sulla relazione di accessibilità $\prec_p$. Li denoteremo con $[G_p]$ ed $[H_p]$, con la stessa interpretazione di $[G]$ e di $[H]$ su $\prec$;

[12]Questa proprietà non caratterizza le previsioni: in [Bon01b] viene data la definizione minimale di previsione priva di questa proprietà.

i due operatori duali sono indicati con $\langle F_p \rangle$ e $\langle P_p \rangle$.
Definiamo la teoria $\mathbf{BT}_p$ nel linguaggio $\mathcal{L}_{BT_p}$ formata dai seguenti schemi di assiomi:

Mostriamo che:

$\mathbf{BT_p}$	
BT	...
$A1_p$	$p \rightarrow [G_p]\langle P_p \rangle p$
$A2_p$	$p \rightarrow [H_p]\langle F_p \rangle p$
$\subseteq_p$	$[G]p \rightarrow [G_p]p$
$\mathbf{4}_p$	$[G_p]p \rightarrow [G_p][G_p]p$
serialità	$[G_p]p \wedge \langle F \rangle p \rightarrow \langle F_p \rangle p$
$.\mathbf{3_l}$	$(\langle P_p \rangle p \wedge \langle P_p \rangle q) \rightarrow [\langle P_p \rangle (p \wedge q) \vee \langle P_p \rangle (p \wedge \langle P_p \rangle q) \vee \langle P_p \rangle (\langle P_p \rangle p \wedge q)]$
time cons	$(\langle P_p \rangle p \wedge \langle P \rangle q) \rightarrow \langle P_p \rangle (p \wedge q) \vee \langle P_p \rangle (p \wedge \langle P \rangle q) \vee \langle P_p \rangle (\langle P \rangle p \wedge q)$

Proposizione 3.13. $\mathbf{BT_p}$ *è corretta sulla classe dei* BT_p*-frames.*

Dimostrazione. Una BT_p struttura è un BT-frame con 2 relazioni $\prec$ e $\prec_p$ entrambe con interpretazione bidirezionale. Dunque gli assiomi di **BT** sono validi così come gli assiomi temporali per $\prec_p$. L'assioma di inclusione, l'assiom **4** di transitività e la serialità corrispondono chiaramente alle proprietà $(P4)$, $(P5)$ e $(P6)$ della Definizione 3.12.
Dimostriamo ora che $(P7)$ implica che valgano le due seguenti proprietà:

i)linearità a ritroso di $\prec_p$: se $t_2 \prec_p t_1$ e $t_3 \prec_p t_1$ allora o $t_2 = t_3$ o $t_2 \prec_p t_3$ o $t_3 \prec_p t_2$;

ii)se $t_1 \prec_p t_3$ e $t_2 \prec t_3$ allora o $t_1 = t_2$ o $t_2 \prec t_1$ oppure $t_1 \prec t_2$ e in questo caso $t_2 \prec_p t_3$.

Sappiamo già che $.\mathbf{3_l}$ è valido sui frame che verificano la prima di queste due proprietà. È immediato poi mostrare che la seconda corrisponde esattamente all'assioma di consistenza temporale.
Supponiamo dunque che esistano $t_2 \prec_p t_1$ e $t_3 \prec_p t_1$, per linearità a ritroso di $\prec$ sappiamo che o $t_2 = t_3$ o $t_2 \prec t_3$ o $t_3 = t_2$. Nel primo caso abbiamo finito, nel secondo caso abbiamo $t_2 \prec t_3$, $t_3 \prec t_1$, $t_2 \prec_p t_1$ dunque per $(P7)$ otteniamo $t_2 \prec_p t_3$. Allo stesso modo nel terzo caso.
La seconda proprietà si dimostra con lo stesso metodo. □

Inoltre si dimostra che:

Teorema 3.14 ([Bon01b])**.** $\mathbf{BT}_p$ *è canonica, dunque completa rispetto alla classe delle* BT_p*-strutture.*

In [Bon01b] le previsioni vengono introdotte in strutture temporali ad albero senza giocatori, ma i risultati ottenuti si generalizzano facilmente al caso della teoria $\mathbf{BTA}_p$ e alle BTA_p-strutture.

3.2.5 Consistenza interna ed induzione a ritroso

Nelle sezioni precedenti abbiamo associato ad ogni gioco in forma estesa un modello di **BTA**, che ne modellizza la struttura sequenziale. Vogliamo ora inserire i concetti risolutivi all'interno del modello, come previsioni sull'albero di gioco.

Definizione 3.15. Definiamo *modello normativo* di G un modello di $\mathbf{BTA}_p$ $\mathcal{M}_p = (\mathcal{M}_G, \prec_p)$, dato dal modello $\mathcal{M}_G$ di G e da una previsione $\prec_p$ su $\mathcal{M}_G$.

Profili di strategie e previsioni

Sia $\sigma = (\sigma_1, \ldots, \sigma_n)$ un profilo di strategie per il gioco G; σ è una funzione definita su tutti i nodi non terminali, scriviamo $s \rightarrowtail^\sigma t$ se $\sigma(s) = t$. Definiamo inoltre la funzione valore di σ, $v^\sigma : \Omega \longrightarrow \mathbb{R}^n$, definita per induzione a partire dai nodi terminali:

- per ogni $z \in Z$ $v^\sigma(z) = u(z)$
- per ogni $t \in \Omega \setminus Z$ e $t' \in \Omega$, se $t \rightarrowtail^\sigma t'$ allora $v^\sigma(t) = v^\sigma(t')$.

Osservazione: Un profilo di strategie σ^* è dato dall'induzione a ritroso, se e solo se $\rightarrowtail^*$ e v^* associati al profilo σ^* sono tali che

$$v^*_{\iota(t)}(t) = \max_{t' : t \rightarrowtail t'} v^*_{\iota(t)}(t').$$

Lemma 3.16. *Dato un gioco finito in forma estesa ad informazione perfetta* $\langle N, (\Omega, \rightarrowtail), \iota, u\rangle$ *ed un profilo di strategie* σ*, se indichiamo con* $\prec_p^\sigma$ *la chiusura transitiva di* $\rightarrowtail^\sigma$*, allora* $\prec_p^\sigma$ *è una previsione secondo la Definizione 3.12.*

Dimostrazione. Le proprietà $(P4)$ e $(P5)$ sono immediatamente verificate dato che $\prec_p^\sigma$ è la chiusura transitiva di $\rightarrowtail^\sigma$ per un qualche profilo di strategie σ, e $\prec$ è la chiusura transitiva di $\rightarrowtail$ con $\rightarrowtail^\sigma \subseteq \rightarrowtail$. $(P6)$ è valida poiché un profilo di strategie è una funzione definita su tutti i nodi non terminali, dunque $\prec_p^\sigma$ è definita in ogni nodo che ammette un successore.
Sia ora $t_1 \prec t_2$ e $t_2 \prec t_3$ e $t_1 \prec_p^\sigma t_3$, perché valga la proprietà $(P7)$ dobbiamo dimostrare che anche $t_1 \prec_p^\sigma t_2$ e $t_2 \prec_p^\sigma t_3$. Dato che Ω è un albero, il cammino

da t_1 a t_3 è unico e passa per t_2; $\prec_p^\sigma$ è la chiusura transitiva di una relazione ($\rightarrowtail^\sigma$) definita su ogni nodo, dunque da $t_1 \prec_p^\sigma t_3$ otteniamo che $t_i \prec_p^\sigma t_j$ per ogni $t_{i,j}$ sul cammino tra t_1 e t_3. □

Preso un gioco in forma estesa $G = \langle N, (\Omega, \rightarrowtail), \iota, u\rangle$, ad ogni profilo di strategie σ possiamo associare un modello normativo $\mathcal{M}_p = (\mathcal{M}_G, \prec_p^\sigma)$ di G. Tra di essi figurano in particolare i modelli dove $\prec_p$ è definita da un profilo di strategie proveniente dall'induzione a ritroso. Per il teorema di Kuhn esiste sempre un tale modello.

Consistenza interna

Come nel capitolo precedente per i giochi in forma strategica, anche in questo caso è possibile esprimere la consistenza interna di una raccomandazione in una formula del linguaggio $\mathcal{L}_{BTA_p}$.
Come già accennato all'inizio del capitolo, è necessario esprimere una formula che valga in ogni punto sull'albero, dunque se si prevede in un nodo che l'utilità di un giocatore i sia q, allora per ogni azione a disposizione di i se l'utilità risultante è r o si prevede che dopo aver compiuto l'azione la sua utilità sarà r, q deve essere maggiore di r.
Questo è traducibile nella seguente formula, come introdotta in [Bon01a]:

$$\langle F_p\rangle(u_i = q) \rightarrow \Box_i[((u_i = r) \vee \langle F_p\rangle(u_i = r)) \rightarrow (r \leqslant q)] \qquad (3.2)$$

Nel caso del linguaggio infinito la consistenza interna della raccomandazione $\prec_p$ si traduce nella richiesta che la formula 3.2 valga per ogni $i \in N$ e per ogni $p, q \in \mathbb{Q}$.
Nel caso del linguaggio finito si ottiene una rappresentazione compatta raccogliendo tutte le istanze della formula 3.2 in un'unica formula:

$$[IC] \quad \bigwedge_{i\in N}\bigwedge_{q\in val}\left[\langle F_p\rangle(u_i = q) \rightarrow \Box_i\left(\bigwedge_{r\in val}((u_i = r) \vee \langle F_p\rangle(u_i = r)) \rightarrow (q \leqslant r)\right)\right]$$

Indicheremo indifferentemente con "IC è vero in s", a seconda del linguaggio utilizzato, il fatto che la formula riassuntiva IC è vera in s, oppure equivalentemente che ogni istanza della formula 3.2 è vera per ogni i, p, q.

Definizione 3.17. Una previsione $\prec_p$ per G si dice *internamente consistente* se e solo se $(\mathcal{M}_G, \prec_p) \models IC$[13].

Le proposizioni che seguono permettono di caratterizzare completamente le previsioni internamente consistenti. Al solito supporremo che i giochi siano finiti e ad informazione perfetta.

[13] IC è globalmente vera su $(\mathcal{M}_G, \prec_p)$.

Proposizione 3.18 ([Bon01a])**.** *Per ogni gioco G, se σ^* è un profilo definito per induzione a ritroso, allora la previsione associata $\prec_p^*$ è internamente consistente.*

Dimostrazione. Indichiamo con $\mathcal{M}_p$ il modello normativo $(\mathcal{M}_G, \prec_p^*)$. Dimostriamo che se $\prec_p^*$ è la chiusura transitiva di $\rightarrowtail^*$ definita per induzione a ritroso, ogni istanza della formula 3.2 è vera in ogni nodo di $\mathcal{M}_p$.
Se $z \in Z$, ossia il nodo è terminale, allora $\mathcal{M}_p, z \models \neg\langle F_p\rangle\psi$ per ogni formula ψ, dunque ogni istanza di IC è verificata poiché $\langle F_p\rangle(u_i = q)$ è falso per ogni q .
Consideriamo il caso di $t \in \Omega \setminus Z$ e $\iota(t) = i$; se $j \neq i$ allora IC è vera perché $\mathcal{M}_p, t \models \Box_j\psi$ per ogni formula ψ. Rimane da dimostrare che

$$\mathcal{M}_p, t \models \langle F_p\rangle(u_i = q) \rightarrow \Box_i[((u_i = r) \vee \langle F_p\rangle(u_i = r)) \rightarrow (r \leqslant q)]$$

per $i = \iota(t)$.
Supponiamo per assurdo che sia falso, e dunque che esista un q tale che $\mathcal{M}_p, t \models \langle F_p\rangle(u_i = q)$, ed un nodo t' tale che tR_it' e tale che

$$\mathcal{M}_p, t' \models ((u_i = r) \vee \langle F_p\rangle(u_i = r)) \wedge \neg(r \leqslant q).$$

Dato che $\prec_p^*$ è associata al profilo di strategie σ^*, esiste un unico $z \in Z$ tale che $t' \prec_p^* z$ (può essere t' stesso, e in tal caso $\mathcal{M}_p, t' \models (u_i = r)$). Ricordando la definizione di v^σ, sui nodi terminali si ha che $v_i^\sigma(t') = u_i(z) = r$. Per l'osservazione alla Sezione 3.2.5 si ha che deve essere

$$q = v_i^\sigma(t) = \max_{t':t\rightarrowtail t'} v_i^\sigma(t')$$

Ma $t \rightarrowtail t'$ e $v_i^\sigma(t') = r > q$, contro il fatto che σ è definita per induzione a ritroso. □

Il Lemma che segue è utile per dimostrare l'inverso della proposizione precedente:

Lemma 3.19. *Sia $(\mathcal{M}_G, \prec_p)$ un modello normativo di G; se $\prec_p$ è internamente consistente ed esistono nodi t, z_1, z_2, con z_1 e z_2 terminali, tali che $t \prec_p z_1$ e $t \prec_p z_2$ allora $u_{\iota(t)}(z_1) = u_{\iota(t)}(z_2)$.*

Dimostrazione. Supponiamo per assurdo che sia $u_\iota(t)(z_1) = r > q = u_\iota(z_2)$. Esiste un unico cammino da t a z_2, e se t' è il nodo successivo a t su questo cammino, si ha che per la proprietà $(P7)$ della previsione $\prec_p$ anche $t' \prec_p z_2$. Ma allora si avrebbe che $\mathcal{M}_p, t \models \langle F_p\rangle(u_i = q)$ e, dato che $\iota(t) = i$ dunque tR_it', anche $\mathcal{M}_p, t' \models \langle F_p\rangle(u_i = r) \wedge \neg(r \leqslant q)$ contro l'ipotesi che IC sia vera in t. □

Proposizione 3.20 ([Bon01a])**.** *Sia* $(\mathcal{M}_G, \prec_p)$ *un modello normativo per il gioco* G*; se* $\prec_p$ *è internamente consistente allora:*

i) $\prec_p$ *è la chiusura transitiva di una sottorelazione* $\rightarrowtail^p$ *di* $\rightarrowtail$ *su* Ω*;*

ii) esiste un profilo di strategie σ *definito per induzione a ritroso tale che* $\rightarrowtail^\sigma$ *è incluso in* $\rightarrowtail^p$*;*

iii) se $s \rightarrowtail_p t$ *ma* $s \not\rightarrowtail^\sigma t$ *allora se* z_1 *è l'unico nodo terminale* σ*-raggiungibile da* s *e* z_2 *quello* σ*-raggiungibile da* t*, si ha che* $u_{\iota(s)}(z_1) = u_{\iota(s)}(z_2)$*.*

Dimostrazione. Il primo punto è ovvio: $\prec_p$ è transitiva, dunque basta definire $s \rightarrowtail^p t$ se $s \prec_p t$ e non esiste z tale che $t \prec_p z \prec_p t$.
Consideriamo ora una qualsiasi funzione $\rightarrowtail^\sigma$ inclusa in $\rightarrowtail^p$, definita su tutto $\Omega \setminus Z$ (per serialità di $\prec_p$, $\rightarrowtail^p$ è definita su tutto $\Omega \setminus Z$). Dimostriamo che il profilo di strategie σ indotto è definito per induzione a ritroso.
Per l'osservazione alla Sezione 3.2.5 è sufficiente mostrare che per ogni $s \in \Omega$

$$v^\sigma_{\iota(s)}(s) = \max_{t:s\rightarrowtail t} v^\sigma_{\iota(s)}(t).$$

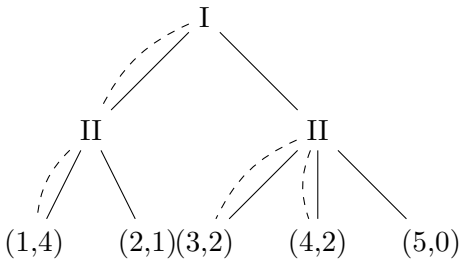
Supponiamo che non sia così: sia $\iota(s) = i$, e t tale che $v^\sigma_i(t) = r > q = v^\sigma_i(s)$. Si avrebbe allora che $\mathcal{M}_p, s \models \langle F_p\rangle(u_i = q)$ ed $\mathcal{M}_p, t \models \langle F_p\rangle(u_i = r) \wedge \neg(r \leqslant q)$ con sR_it. Se $s \not\rightarrowtail^p t$ allora si ha un assurdo poiché $\mathcal{M}_p, s \models \Diamond_i\langle F_p\rangle(u_i = r) \wedge \neg(r \leqslant q)$, contro l'ipotesi che IC fosse globalmente vero. Se invece $s \rightarrowtail^p t$ allora per il Lemma 3.19 si avrebbe che $v^\sigma_i(t) = v^\sigma_i(s)$ dato che $\rightarrowtail^\sigma \subseteq \rightarrowtail^p$, contro l'ipotesi che fossero diversi.
Per lo stesso motivo se $s \rightarrowtail_p t$ ma $s \not\rightarrowtail^\sigma t$ e z_1 e z_2 sono i due nodi terminali σ-raggiungibili rispettivamente da s e da t, allora dato che $\rightarrowtail^\sigma \subseteq \rightarrowtail^p$ si ha che entrambi z_1 e z_2 sono $\rightarrowtail^p$-raggiungibili da s, ossia $s \prec_p z_1$ ed $s \prec_p z_2$., e dal Lemma 3.19 si ottiene *iii*). □

In un gioco generico (nessun giocatore è indifferente rispetto a nessun esito) c'è un unico profilo di strategie dato dall'induzione a ritroso: il seguente corollario afferma che questo profilo è l'unica raccomandazione internamente consistente.

Corollario 3.21. *Se il gioco* G *è generico (Definizione 1.17) allora* $\prec_p$ *è internamente consistente per* G *se e solo se* $\prec_p = \prec^\sigma_p$ *dove* σ *è l'unico profilo di strategie definito per induzione a ritroso.*

Dimostrazione. Se G è generico nessun giocatore ha la stessa utilità in due nodi diversi, dunque il Lemma 3.19 implica che per ogni nodo s esiste un unico nodo terminale z tale che $s \prec_p z$. Definiamo il profilo di strategie σ come alla Proposizione 3.20 ed è immediato vedere che $\prec^\sigma = \prec_p$. □

Nel caso di giochi non generici è possibile invece definire previsioni che, pur soddisfando IC, non sono associate all'induzione a ritroso nè a nessun altro profilo di strategie. Un esempio è esposto nella figura sottostante, dove la previsione, indicata dalle frecce tratteggiate, è consistente ma non è associata a nessun profilo di strategie.



Dalla Proposizione 3.20 segue immediatamente il seguente corollario:

Corollario 3.22. *Se σ è un profilo di strategie per G, allora $\prec_p^\sigma$ è internamente consistente per G se e solo se σ è definito dall'induzione a ritroso.*

Si può dunque ottenere un risultato analogo a quello ottenuto alla sezione 3.1 per concetti risolutivi nel caso di giochi in forma strategica.
Ogni concetto risolutivo F associa ad ogni gioco in forma estesa un insieme di profili di strategie. Ad ogni profilo di strategie σ possiamo associare una previsione sul modello $\mathcal{M}_G$ data da $\prec_p^\sigma$.
Diciamo che un concetto risolutivo è internamente consistente se per ogni gioco G e per ogni $\sigma \in F(G)$, $\prec_p^\sigma$ è internamente consistente per G.

Teorema 3.23. *Un concetto risolutivo in strategie pure F è internamente consistente se e solo se per ogni G si ha che $F(G) \subseteq I(G)$, dove $I(G)$ sono gli equilibri perfetti nei sottogiochi.*

L'induzione a ritroso I è dunque il concetto risolutivo internamente consistente massimale.

Bibliografia

[AB95] R.J. Aumann and A. Brandenburger. Epistemic conditions for Nash equilibria. *Econometrica*, (63):1161–1180, 1995.

[All53] M. Allais. Le comportement de l'homme rationnel devant de risque: Critique des postulats de l'ecole americaine. *Econometrica*, 21:503–546, 1953.

[Aum76] R.J. Aumann. Agreeing to disagree. *The Annals of Statistic*, 4(6):1236–1239, 1976.

[Aum85] R.J. Aumann. What is game theory trying to accomplish? In K. Arrow and S. Honkapohja, editors, *Frontiers of Economics*, pages 77–87. Basil Blackwell, 1985.

[Aum99] R.J. Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, (28):263–300, 1999.

[BB99] P. Battigalli and G. Bonanno. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, (53):149–225, 1999.

[BdRV01] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

[Bon01a] G. Bonanno. Branching time, perfect information games, and backward induction. *Games and Economic Behaviour*, (36):57–73, 2001.

[Bon01b] G. Bonanno. Prediction in branching time logic. *Mathematical Logic Quarterly*, 47(2):239–247, 2001.

[Bon02] G. Bonanno. Modal logic and game theory: two alternative approaches. *Risk Decision and Policy*, 7(3):309–324, 2002.

[dF63a] B. de Finetti. La teoria dei giochi. *Civiltà delle macchine*, (4), 1963.

[dF63b] B. de Finetti. Riflessioni attuali sulla teoria dei giochi. *Civiltà delle macchine*, (5), 1963.

[FHMV95] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning About Knowledge*. MIT Press, 1995.

[Fis70] P.C. Fishburn. *Utility Theory for Decision Making*. John Wiley and sons, New york, 1970.

[Gre90] J. Greenberg. *The Theory of Social Situations*. Cambridge University Press, 1990.

[HC96] G.E. Huges and M.J. Cresswell. *A New Introduction to Modal Logic*. Routledge, 1996.

[Hin62] J. Hintikka. *Knowledge and Belief*. Cornell University Press, 1962.

[Kre88] D. Kreps. *Notes on the Theory of Choice*. Undergraduate Classics in Economics, 1988.

[Kre90] D. Kreps. *Game Theory and Economic Modelling*. Oxford University Press, 1990.

[Len78] W. Lenzen. Recent work in epistemic logic. *Acta Philosophica Fennica*, (30):1–219, 1978.

[Lew69] D. Lewis. *Convention, a Philosophical Study*. Cambridge, Mass.: Harvard University Press, 1969.

[Mer96] L. Mero. *Calcoli Morali*. Per l'edizione italiana, Edizioni Dedalo,, 1996.

[Nas50] J. Nash. Equilibrium points in n-persons games. In *Proceedings of the National Academy of Sciences of the United States of America*, number 36, pages 48–49, 1950.

[Nas54] J. Nash. Non cooperative games. *Annals of Mathematics*, 54:286–295, 1954.

[OR94] M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[Pat06] F. Patrone. *Decisori (razionali) intelligenti*. Plus: Pisa University Press, 2006.

[Sav54] J. Savage. *The Foundations of Statistics*. New York, Wiley, 1954.

[Sch60] T. Schelling. *The strategy of conflict*. Harvard University Press, 1960.

[Sel85] R. Selten. Comment. In K. Arrow and S. Honkapohja, editors, *Frontiers of Economics*, pages 77–87. Basil Blackwell, 1985.

[vdHP06] W. van der Hoek and M. Pauly. Modal logic for games and information. In P. Blackburn, J. van Benthem, and F.Wolter, editors, *Handbook of Modal Logic, 3*. North Holland, 2006.

[vNM47] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Priceton Univeristy Press, 1947.