Una teoria della decidibilità: entropia e scelte in condizioni di incertezza

Pietro Coretto*
Università di Salerno

Questo lavoro presenta un nuovo modello di scelta in condizioni di incertezza. Dopo aver introdotto una caratterizzazione del concetto di incertezza, si dimostra, su base assiomatica, come sia possibile interpretare la funzione di entropia come misura di incertezza debole. *I concetti di entropia e di utilità attesa sono impiegati per costruire su base assiomatica una nuova funzione: la funzione di decidibilità, che è in grado di ordinare le preferenze sullo spazio delle lotterie. Infine si mostra che questo modello è in grado di razionalizzare sia il paradosso di Allais che quello di Ellsberg.* [Cod. JEL: D81, D89, C91].

1. - Introduzione

Una prima sfida alla teoria dell'utilità attesa (EUT) di Von Neumann e Morgenstern [23] fu lanciata da Allais [1] con un fa-

---

* <*pietro.coretto@tin.it*>, ringrazia in modo particolare il Prof. Marco Pagano, per il ruolo centrale che ha svolto nella sua formazione, nella sua vita di studente, e per il grande aiuto ricevuto come relatore della tesi di laurea; ringrazia inoltre il Prof. Kenneth J. Arrow, il Prof. Mark Machina, il Prof. Aldo Rustichini ed il Prof. Itzhak Gilboa per i loro commenti. Inoltre un ringraziamento va alla *Rivista di Politica Economica* ed alla Confindustria per la grande opportunità che il premio *Angelo Costa* ci offre. La versione finale di questo lavoro ha beneficiato molto dei commenti e dei suggerimenti dei *referees* internazionali ed italiani. Infine ringrazia Vanessa Janowski per il suo aiuto nella stesura della versione inglese di questo articolo.

Avvertenza: i numeri nelle parentesi quadre si riferiscono alla Bibliografia alla fine del testo.

moso esperimento che condusse all'ormai celebre risultato conosciuto come *paradosso di Allais*; esso era un esperimento in cui le persone fanno scelte che non sono coerenti con il modello di utilità attesa. Successivamente un altro esperimento condotto da Ellsberg [11] mise in dubbio la possibilità di descrivere il comportamento individuale nelle scelte in condizioni di incertezza attraverso il modello di utilità attesa. La differenza principale fra i due esperimenti è che nell'esperimento di Allais la probabilità di un dato evento è data, cioè siamo in un contesto con probabilità oggettive, mentre l'esperimento di Ellsberg è condotto in un ambiente in cui le probabilità non sono note.

Successivamente molti modelli alternativi a quello della EUT di Von Neumann e Morgenstern sono stati elaborati per cercare di spiegare gli aspetti del comportamento umano che questo modello non riesce a catturare. Possiamo dividere questi modelli in due gruppi omogenei: *(i)* modelli basati su una concezione della probabilità oggettivista; *(ii)* modelli basati su probabilità soggettive. L'elemento comune ai modelli di ognuno dei due gruppi è la formulazione alternativa dell'assioma di indipendenza — elemento centrale nella EUT — e quindi nel modo in cui gli individui formulano le loro aspettative.

Il primo tentativo di risolvere il primo dei due paradossi, in un contesto di probabilità oggettive, venne proprio da Allais [1], il quale ipotizzò la violazione dell'assioma di indipendenza. L'assioma di indipendenza costituisce un'assunzione fondamentale per la EUT secondo la quale date due lotterie, se queste sono combinate con una terza l'ordine delle preferenze rimane invariato. Allais [1] tentò di spiegare il comportamento individuale in termini di effetto *fanning out* delle curve di indifferenza, ovvero le curve di indifferenza rappresentate sul simplesso sono lineari ma non parallele. Dopo il tentativo di Allais [1], gran parte dei ricercatori cercarono di modellare l'effetto *fanning out*. Questi studi, concentrati tra la fine degli anni '70 ed gli anni '80, sono accomunati dalla modificazione dell'assioma di indipendenza, ed in questa diversa assiomatizzazione l'ipotesi di Allais finiva per emergere come risultato della teoria. Tra queste proposte ci sono il modello *weighted expected utility* di Chew e MacCrimmon [9] ed il più noto modello della *non-linear expected utility* dovuto a Machina [18].

Il secondo gruppo di modelli si basa sulla formalizzazione del concetto di probabilità soggettive, inizialmente proposto da De Finetti [10] e Savage [20], e successivamente assiomatizzato da Ascombe ed Aumann [3]. Recentemente vi è un forte interesse di ricerca in questo campo, e le proposte teoriche più suggestive sono costituite dal modello della *Choquet expected utility* dovuto a Schmeidler [21], della *maxmin expected utility* di Gilboa e Schmeidler [13], e la *expected utility without completeness* di Bewley [5], [6]. Tuttavia la EUT resta ancora il modello più utilizzato in economia, dalla teoria della finanza alle microfondazioni della teoria macroeconomica. Perché vi è tale riluttanza ad abbandonare tale modello, nonostante i suoi noti limiti e la presenza di modelli alternativi? La risposta non è semplice, ma ci sono almeno due ragioni importanti da considerare. La prima ragione è che probabilmente sembra molto difficile rinunciare alla semplicità descrittiva della EUT, che sembra essere un'espressione naturale della logica individuale nelle decisioni. Un'altra ragione è che tutti i modelli alternativi citati sono analiticamente complessi e la loro applicazione è ardua. Il nostro tentativo sarà quello di elaborare una teoria di cui la EUT, fusa con una nuova visione dell'incertezza, sarà una parte rilevante.

2. - Incertezza

Nella costruzione della EUT le probabilità sono assunte come misura dell'incertezza. In contrasto con questa visione, la nostra idea è che le probabilità siano il grado di plausibilità di un evento, e non possano essere interpretate come una misura dell'incertezza. Una misura di incertezza dovrebbe essere legata all'intero fenomeno oggetto dell'analisi.

Per comprendere meglio quest'affermazione, introduciamo un esempio. Supponiamo che un annunciatore stia dando il notiziario meteorologico alla TV. Supponiamo di ascoltare due previsioni A e B. Nel caso A si prevede pioggia con probabilità 1/2, e sole con probabilità 1/2. Nella previsione B si annuncia che con probabilità 1/2 ci sarà pioggia, con probabilità 1/6 ci sarà sole, e che con probabilità 1/3 ci sarà neve. Dovendo stabilire quale delle due notizie genera in noi un grado di incertezza maggiore, probabilmente diremmo il notiziario *A* contiene un grado di incertezza maggiore. Infatti, sia nel caso *A* che *B* la probabilità che vi sia pioggia è di 1/2, ma, mentre nel caso *A* c'è un altro evento con lo stesso grado di plausibilità di quello relativo alla pioggia, nel caso *B* c'è un evento la cui probabilità domina quella di tutti gli altri. Sebbene una previsione sia sempre un'affermazione incerta, nel caso *B* le persone sono persuase del fatto che vi sia un evento (pioggia) più plausibile degli altri. Questo esempio mostra che quando le persone non considerano il singolo evento, ma l'insieme degli eventi o dei risultati, il grado di incertezza che esse percepiscono dipende da un giudizio che coinvolge le probabilità di tutti gli eventi. La nostra idea è che la probabilità è un concetto assoluto, nel senso che essa esprime il grado di plausibilità di una dato evento; mentre l'incertezza è un concetto relativo, nel senso che il suo livello dipende da una valutazione che coinvolge le probabilità di tutti gli eventi che si stanno considerando.

2.1 *Caratterizzazione dei contesti di incertezza*

Prima di tutto occorre definire esattamente l'esatto contesto in cui ci muoveremo. Nel paragrafo precedente si è sottolineato che l'incertezza non è un concetto univoco, cosicché ad ogni tipologia di incertezza corrisponderà una data misura di incertezza. Per semplicità supponiamo di essere interessati ad un fenomeno o esperimento E_X, laddove $X \subseteq \mathbb{R}$ e l'insieme di tutti i possibili risultati, $\boldsymbol{d} = \{x_1, x_2, \dots, x_n\}$ è il vettore dei dati sperimentali. Supponiamo che $f(x; \boldsymbol{v})$ è una funzione parametrica che descrive il vero modello generatore dei dati, dove $\boldsymbol{v}$ è un vettore di parametri. In un esperimento la funzione $f(x, \boldsymbol{v})$ potrebbe essere la distribuzione di probabilità, o una funzione di densità che descrive il meccanismo probabilistico di E_X. Distinguiamo cinque diversi contesti:

(i) assenza di conoscenza: non conosciamo nulla. Sappiamo soltanto che il fenomeno E_X esiste, ma non siamo in grado di osservarlo o di misurare i suoi effetti. Non abbiamo dati da analizzare;

(ii) incertezza forte o informazione debole: conosciamo parzialmente il vettore $\boldsymbol{d}$ ma non conosciamo il modello che lo genera. Vale a dire che $f(x, \boldsymbol{v})$, è ignota. Questa situazione è equivalente al caso dell'incertezza knightiana, ed in questo contesto gli individui formulano probabilità soggettive[1];

(iii) incertezza semi-debole o informazione semi-debole: conosciamo $\boldsymbol{d}$ e conosciamo parzialmente $f(x, \boldsymbol{v})$ perché non conosciamo $\boldsymbol{v}$. In questo caso dobbiamo inferire $\boldsymbol{v}$ attraverso un processo di stima basato su $\boldsymbol{d}$;

(iv) incertezza debole o informazione forte: si conosce $\boldsymbol{d}$ ed $f(x; \boldsymbol{v})$, cosicché il modello sottostante i dati è noto. In questo caso siamo in un contesto di probabilità oggettive;

(v) conoscenza completa (o certezza): conosciamo perfettamente tutti gli elementi di E_X, ed il fenomeno che osserviamo è deterministico.

Questa caratterizzazione fornisce la possibilità di sviluppare una teoria della misura dell'incertezza relativamente ai diversi contesti. Inoltre in questo modo introduciamo un nuovo legame tra informazione ed incertezza sostituendo ed arricchendo i classici ed opposti concetti di informazione completa ed informazione incompleta. Seguendo un'idea molto dibattuta in fisica (Brillouin [7]), in questo lavoro si propone una visione secondo la quale incertezza ed informazione sono due faccie di una stessa medaglia. La nozione di informazione è opposta a quella di incertezza nel senso che se dobbiamo assumere una decisione ci sentiamo incerti se non siamo abbastanza informati. Seguendo l'idea di Brillouin [7], si ritiene che «l'incertezza è uguale ad informazione negativa». In ognuno dei contesti descritti sopra abbiamo diverse misure di incertezza — vale dire diverse misure di informazione

---

[1] Secondo KNIGHT F.H. [17], un evento è incerto se non se ne conosce la probabilità.

negativa — le cui proprietà verranno fuori dal contesto specifico. Questo significa che se riusciamo a trovare una misura $S(\cdot)$ di incertezza forte, allora $-S(\cdot)$ sarà una misura di informazione debole. Questo non significa che una stessa misura di incertezza $S(\cdot)$ valida in un certo contesto possa essere validamente impiegata in un altro contesto. In questo lavoro tratteremo di misure di incertezza debole o informazione forte.

C'è un punto in comune tra le diverse situazioni proposte sopra: l'incertezza si manifesta nel momento in cui dobbiamo assumere decisioni. Quando qualcosa è probabile questo non significa che sia incerto, ma piuttosto che è plausibile con un certo grado che chiamiamo probabilità. In questo senso, citando Bernoulli, la probabilità può essere vista come un grado di approssimazione alla realtà. Invece, siamo incerti nel momento in cui dobbiamo assumere delle decisioni, e per questo motivo l'incertezza è legata alle attività mentali coinvolte nel processo decisionale.

Noi interpretiamo l'incertezza come una sensazione di imbarazzo psicologico, ovvero uno stato mentale di indeterminazione dovuto alla necessità di dover prendere una decisione. Il bisogno di assumere decisioni provoca un'attività mentale la cui intensità cresce al crescere del livello di incertezza. Se dobbiamo decidere su un'azione più incerta di un'altra, allora noi impieghiamo più tempo e risorse mentali per scegliere l'azione preferita tra quelle disponibili. Ispirandoci ad un'idea di Shannon [22] e De Finetti [10], e seguendo parzialmente il loro ragionamento, introduciamo l'idea secondo la quale una misura di incertezza è anche una misura del "grado di decisione" o della "quantità di scelta" occorrente per assumere una decisione. Quando una scelta è legata ad una situazione molto incerta, il processo decisionale diventa arduo, cosicché la quantità di scelta aumenta. Questa è la ragione per la quale i termini: grado di decisione, quantità di scelta o misura di incertezza saranno utilizzati equivalentemente.

Infine, una misura di incertezza debole è anche uguale ad una misura negativa di informazione forte. Questi argomenti saranno molto importanti per comprendere la logica dell'assiomatizzazione nel paragrafo 3.2.

3. - La funzione di entropia

3.1 *L'entropia nelle scienze naturali*

Durante il 20° secolo il concetto di entropia è stato uno dei più discussi nella fisica e nella teoria dell'informazione. La nozione di entropia fu introdotta per la prima volta da R.J.E. Clasius in termodinamica, e successivamente Boltzman ne diede un'interpretazione in termini di probabilità. Il secondo principio della termodinamica è basato sul concetto di entropia, che in generale in fisica è assunto come misura del disordine di un sistema fisico. Nel 1948 Shannon [22] impiegò la nozione di entropia nell'ambito della teoria delle comunicazioni dimostrando che è possibile utilizzare l'entropia come misura dell'informazione trasmessa attraverso un canale di comunicazione fisico.

Cercheremo di introdurre brevemente la nozione di entropia nel contesto esplorato da Shannon. Sia $X = \{x_i;\, i = 1, 2, \dots, n\}$ un alfabeto, cioè un insieme finito di simboli che vengono trasmessi da una sorgente S. Supponiamo che il simbolo $x_i \in X$, sia trasmesso con probabilità p_i, e sia $\mathcal{P}_X = \{p_i;\, i = 1, 2, \dots, n\}$ una distribuzione di probabilità su X. Shannon [22] dimostrò, sotto un certo numero di ipotesi (assiomi), che è possibile misurare l'ammontare di informazione trasmessa attraverso un canale fisico attraverso la funzione:

$$H(p_1, p_2, \dots, p_n) = -K \sum_{i=1}^{n} p_i \log p_i \tag{1}$$

chiamata *entropia di Shannon*, dove K è una costante reale e positiva che dipende dall'unità misura adottata[2]. Esiste una vasta letteratura su questo argomento, ma in questo lavoro non siamo interessati a tracciare un quadro delle diverse interpretazioni del concetto di entropia nei diversi contesti scientifici. Piuttosto il nostro scopo principale è quello di comprendere sotto quali ipotesi coerenti l'entropia di Shannon possa essere assunta come misura di incertezza.

[2] In teoria dell'informazione viene generalmente impiegata l'entropia binaria, vale a dire si fissa $K = 2$. Il ruolo centrale dell'entropia binaria è determinato del massiccio uso dell'alfabeto binario $B = \{0, 1\}$.

Sebbene in molti lavori — nella fisica e nella teoria dell'informazione — si ritrovi l'idea che il tasso di informazione emessa da una sorgente possa essere misurato attraverso la riduzione di incertezza provocata dall'emissione, dove la riduzione di incertezza è misurata attraverso l'entropia, questo punto di vista non è accettato dall'intera comunità scientifica. In letteratura si trovano diversi tentativi di rendere l'entropia un concetto universale (Brillouin [7]), ma nessuno di essi sembra chiarire la questione. L'osservazione che l'estensione dell'entropia alle sorgenti continue non ha senso, ha sollevato dubbi sull'universalità della nozione in questione. Inoltre, l'assetto teorico proposto da Shannon [22] è strettamente legato ai canali di comunicazione, e le sue ipotesi sono fortemente vincolate al fatto che gli eventi abbiano una dimensione temporale. Le ipotesi su cui si basa il lavoro di Shannon [22] sono strettamente legate al fatto che la trasmissione dell'informazione si articola nel tempo, cosicché diventa centrale l'ipotesi che gli eventi[3] seguano un processo di Markov. Qui non si vuole discutere della validità di queste ipotesi, ma si vuole studiare se sia possibile, basandosi su minimo numero di assunzioni coerenti, costruire una misura di incertezza che sia valida in un contesto più generale. Per persuadere il lettore del fatto che l'entropia possa essere assunta come valida misura di incertezza, baseremo questo lavoro su un'assiomatizzazione alternativa dove l'ipotesi che gli eventi siano governati da un certo processo stocastico sarà rimossa[4].

3.2 *L'entropia ed i suoi assiomi*

Ora presenteremo l'assiomatizzazione necessaria per costruire una misura di incertezza debole in un contesto generale. In-

[3] Nel caso della trasmissione, un evento è un simbolo trasmesso da una certa sorgente.

[4] Si osservi che esiste una letteratura economica riguardante il controllo robusto ed i modelli con errori di specificazione (HANSEN L.P. - SARGENT T.J. [14]), in cui si utilizza spesso il cosiddetto principio di entropia (BERNARDO J.M. - SMITH F.M.A. [4]), dove il termine entropia viene impiegato in luogo di entropia relativa o discrepanza. In questo lavoro sarà utilizzata la funzione di entropia che è del tutto differente da quella di entropia relativa.

fatti, come già si è sottolineato, in questa assiomatizzazione non supporremo che gli eventi siano governati da un certo processo stocastico, ed in questo senso questa assiomatizzazione[5] differisce da quella di Shannon [22]. Procederemo introducendo prima gli assiomi, e poi ne daremo una spiegazione.

Siano $I = \{1, 2, ..., n\}$ e $J = \{1, 2, ... , m\}$ due insiemi di indici, siano essi insiemi finiti, e sia $m \leq n$. Descriviamo un fenomeno o un esperimento dove ogni possibile risultato è dato da $x_i \in X$, dove $X = \{x_i;\ i \in I\}$ ed X è finito, cosicché $|X| = n$.[6] Sia $S = \{S_j;\ j = \in J\}$ una collezione di sottoinsiemi di X, tale che S è una partizione su X, ovvero:

$$\bigcup_{j\in J} S_j = X \text{ e } S_t \cap S_k = \varnothing; \qquad \forall t,k \in J \ \text{ e } \ t \neq k$$

e si definisca il sottoinsieme $S_j = \{x_i^j;\ i \in I,\ \mathrm{j} \in J\}$, dove x_i^j è il risultato $x_i \in X$ simultaneamente appartenente a S_j. Dato X, definiamo una misura di probabilità $\mathcal{P} = \{p_i,\ i \in I\}$ tale che $\Sigma_{i\in I}\, p_i = 1$ e $p_i \geq 0$ per tutti gli $i \in I$, dove p_i la probabilità del risultato $x_i \in X$. Dati X, $\mathcal{P}$, S ed il sottoinsieme $S_j \in S$, sia q_j la probabilità di un qualsiasi risultato $x^j \in S_j$. Otteniamo così una nuova distribuzione di probabilità $\mathcal{Q}$ definita come: $\mathcal{Q} = \{q_j;\ j \in J\}$,

$$\text{dove:} \quad q_j = \sum_{\substack{i\in I,\, j\in J \\ i,j | x_i^j \in S_j}} p_i \geq 0 \quad \text{, cosicché:} \quad \sum_{i\in J} q_j = 1$$

Sia π_i^j la probabilità di ottenere un certo risultato x_i^j dal sottoinsieme S_j,

cosicché $\qquad \Pi_j = \{\pi_i^j,\ i \in I,\ j \in J\}$,

---

[5] Un *referee* ha suggerito che forse questo tipo di assiomatizzazione potrebbe essere già stata proposta da un matematico russo in un libro sulla teoria dell'informazione. Poiché il *referee* non ha indicato il riferimento, non è stato possibile verificarlo.

[6] $|X|$ indica il numero di elementi contenuti in X.

dove:
$$\sum_{i \in I | x_i^j \in S_j} \pi_i^j = 1 \quad \text{e} \quad \pi_i^j \geq 0 \ \ \forall i \in I, j \in J$$

è una nuova distribuzione di probabilità. Sia $\Delta^n = \{(p_1, p_2, \ldots, p_i, \ldots, p_n) \in [0, 1]^n : \sum_{i \in I} p_i = 1\}$; una funzione $\mathcal{H}$, chiamata entropia, è considerata una misura di incertezza debole; viceversa $-\mathcal{H}$ è una misura di informazione forte, se e solo se essa soddisfa i seguenti assiomi:

ASSIOMA 1. $\mathcal{H} : \Delta^n \to \mathbb{R}_0^+$, dove $\mathcal{H}$ è una funzione continua.

ASSIOMA 2. $\mathcal{H}(p_1, p_2, \ldots, p_n) = \mathcal{H}(p_1, p_2, \ldots, p_{n-t}) + \mathcal{H}(p_{n-t+1}, \ldots, p_n)$ per ogni $t = \{1, 2, \ldots, n-1\}$.

ASSIOMA 3. $\mathcal{H}(p_1, p_2, \ldots, p_n) = \max[\mathcal{H}(p_1, p_2, \ldots, p_n)]$ se e solo se $p_i = 1/n$ per ogni $i \in I$. $\mathcal{H}(p_1, p_2, \ldots, p_n) = \min[\mathcal{H}(p_1, p_2, \ldots, p_n)]$ se e solo se $\mathcal{P}$ è una distribuzione di probabilità degenere.

ASSIOMA 4. $\mathcal{H}(p_1, p_2, \ldots, p_n) = \mathcal{H}(q_1, q_2, \ldots, q_m) + \sum_{j \in J} q_j \, \mathcal{H}(\Pi_j)$, dove $\mathcal{H}(\Pi_j)$ è l'entropia calcolata su tutte le probabilità $\pi_i^j \in \Pi_j$, per ogni $i \in I$, $j \in J$.

Prima di spiegare gli assiomi appena enunciati sopra, vogliamo fare alcune osservazioni sulla notazione. Se abbiamo una distribuzione di probabilità $\mathcal{P}$ su X, $\mathcal{H}(\mathcal{P})$ e $\mathcal{H}(X)$ sono due modi molto comuni per indicare $\mathcal{H}(p_1, p_2, \ldots, p_n)$. Questa simbologia viene usata sia in fisica che in teoria dell'informazione per permettere una presentazione più leggera.

L'assioma 1 si riferisce all'osservazione fatta nel paragrafo 2 in cui si è stabilito che una misura di informazione deve dipendere da tutte la probabilità definite sull'insieme dei risultati. Questo assioma stabilisce anche che una siffatta misura è non negativa e continua. L'assioma 2 stabilisce una proprietà molto comune alle misure: la separabilità additiva della funzione rispetto ai suoi argomenti. L'assioma 3 stabilisce che si ha il grado più basso di incertezza quando la distribuzione di probabilità è degenere, cioè esiste un evento certo. Al contrario, si ha il più alto livello di incertezza quando tutti gli attributi hanno la stessa probabilità. Infatti, in una tale situazione gli individui sono turbati perché ogni evento ha lo stesso grado di plausibilità. L'equiprobabilità causa uno stato di indeterminatezza, un'inabilità nello scegliere che rende la decisione tormentata.

Infine l'assioma 4 cattura due effetti: *(i) ceteris paribus*, c'è una riduzione di incertezza quando il numero di alternative si riduce, cioè la misura di incertezza è crescente nel numero dei possibili risultati; *(ii)* il grado di incertezza resta invariato se modifichiamo la distribuzione di probabilità di partenza in una nuova distribuzione. Questo assioma rappresenta la differenza con l'assiomatizzazione proposta da Shannon [22]. In un certo senso questo assioma può essere visto come una generalizzazione di quella che in teoria dell'informazione viene detta proprietà di ramificazione.

Il ruolo centrale di questo assioma ed il suo potere descrittivo possono essere compresi attraverso il prossimo esempio. Ispirandoci alla notazione usata sopra, supponiamo di dover scegliere sull'insieme dei risultati $X = \{x_1, x_2, \dots, x_{11}\}$. La distribuzione di probabilità definita su:

$$X \text{ è } \mathcal{P} = \{\tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{10}, \tfrac{1}{10}, \tfrac{1}{10}, \tfrac{1}{10}, \tfrac{1}{10}\}$$

Sulla base degli assiomi, una misura di incertezza su X è data da:

$$\mathcal{H}(X) = \mathcal{H}(\tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{10}, \tfrac{1}{10}, \tfrac{1}{10}, \tfrac{1}{10}, \tfrac{1}{10})$$

Supponiamo di non dovere scegliere più su X ma su S, ottenuto alterando X. Ciò significa che non si è più interessati al risultato $x_i \in X$, ma ad uno qualsiasi dei risultati presi in S_1 o S_2. Supponiamo che $S = \{S_1, S_2\}$, dove esso è una collezione di sottoinsiemi di X, i quali sono $S_1 = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, ed $S_2 = \{x_7, x_8, x_9, x_{10}, x_{11}\}$. Dobbiamo assumere una decisione su S conoscendo $Q = \{q_1, q_2\}$, dove $q_1 = 6 \times 1/12$ e $q_2 = 5 \times 1/10$. Infine dobbiamo definire la distribuzione:

$$\Pi_1 = \{\tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}\}$$

e:

$$\Pi_2 = \{\tfrac{1}{5}, \tfrac{1}{5}, \tfrac{1}{5}, \tfrac{1}{5}, \tfrac{1}{5}\}$$

Ricordando che possiamo sempre scrivere $\mathcal{H}(\Pi_1) = \mathcal{H}(S_1)$ e $\mathcal{H}(\Pi_2) = \mathcal{H}(S_2)$, ed applicando l'assioma 4 otteniamo:

(2) $$\mathcal{H}(X) = \mathcal{H}(S) + \frac{1}{2}\mathcal{H}(S_1) + \frac{1}{2}\mathcal{H}(S_2)$$

oppure:

(3) $$\mathcal{H}(S) - \mathcal{H}(X) = -\frac{1}{2}(\mathcal{H}(S_1) + \mathcal{H}(S_2))$$

La *(2)* stabilisce che l'ammontare totale di incertezza di un certo insieme di risultati è data dalla somma ponderata dell'incertezza misurata sui singoli sottoinsiemi di risultati. Ma più interessante è l'osservazione sulla *(3)*: essa stabilisce che se sostituiamo una scelta con numero maggiore di risultati (nell'esempio è X) con una scelta con numero minore di risultati (nell'esempio è S), otteniamo una riduzione di incertezza. Questa riduzione è data dalla somma ponderata dell'entropia calcolata sui sottoinsiemi di risultati (nell'esempio sono $\mathcal{H}(S_1)$ e $\mathcal{H}(S_2)$) ottenuti raggruppando quelli dell'insieme originario (nell'esempio era X).

Sulla base degli assiomi 1, 2, 3 e 4, abbiamo il seguente:

TEOREMA 1. Esiste una funzione $\mathcal{H}\colon \Delta^n \to \mathbb{R}_0^+$ che soddisfa gli assiomi 1, 2, 3 e 4, ed essa è l'entropia di Shannon: $\mathcal{H}(p_1, p_2, \dots, p_n) = -\sum_{i\in I} p_i \log_k(p_i)$, $k \in \mathbb{R}+$, $k > 1$. $\mathcal{H}(p_1, p_2, \dots, p_n)$ è unica a meno di una costante reale e positiva $k > 1$.

DIMOSTRAZIONE. Prima dimostriamo che gli assiomi 1, 2, 3 e 4, sono soddisfatti dall'entropia di Shannon. Supponiamo che $\mathcal{P} = \{p_i;\ i \in I\}$ sia una distribuzione di probabilità uniforme definita sull'insieme $X = \{x_i;\ i \in I\}$, cosicché $P(X = x_i) = 1/n$ per ogni $i \in I$. Supponiamo che $S = \{S_i;\ i \in I\}$ sia la partizione più fine su X tale che $S_i = \{x_i\}$. Questo implica che $\cup_{i\in I} S_i = X$, $S_k \cap S_t = \emptyset$ per ogni t, $k \in I$ e $t \neq k$. Chiaramente $|S_i| = n_i = 1$ per ogni $i \in I$. Applicando gli assiomi 1 e 4, otteniamo:

(4) $$\mathcal{H}(p_1, p_2, \dots, p_n) = \mathcal{H}(q_1, q_2, \dots, q_n) + \sum_{i\in I} q_i\, \mathcal{H}(\Pi_i)$$

Per l'assioma 1 possiamo scrivere $\mathcal{H}(X) = \mathcal{H}(p_1, p_2, \dots, p_n)$, ed essendo $p_i = 1/n$ per ogni $i \in I$, ne segue che $\mathcal{H}$ è una funzione di n. Ora possiamo scrivere $\mathcal{H}(X) = \mathcal{H}(n)$. Poiché ogni sottoinsieme $S_i \in S$ contiene solo $x_i \in X$, allora $q_i = \sum_{i|x_i\in S_i, i\in I} p_i = p_i = 1/n$, e quindi $\mathcal{H}(q_1, q_2, \dots, q_n) = \mathcal{H}(S) = \mathcal{H}(n)$. Sulla base di queste osservazioni, la *(4)* diventa:

$$\mathcal{H}(S) = \sum_{i\in I} p_i\, \mathcal{H}(X) - \sum_{i\in I} p_i\, \mathcal{H}(\Pi_i) \tag{5}$$

Ma $S_i = \{x_i;\ i \in I\}$, e quindi $\Pi_i = \{1/n_i;\ i \in I\}$, e per la ragione menzionata sopra $\mathcal{H}\,(\Pi_i)\ (i) = \mathcal{H}\,(n_i)$. L'equazione *(5)* diventa:

$$\mathcal{H}\,(p_1,\ p_2,\ \dots\ ,\ p_n) = -\sum_{i\in I} p_i\,[\mathcal{H}(n_i) - \mathcal{H}\,(n)] \tag{6}$$

Ora dobbiamo trovare una funzione con valori reali e continua che per l'assioma 1 è definita in Δ^n, e per l'assioma 2 è additivamente separabile nei suoi argomenti. Inoltre questa funzione deve essere unica a meno di una costante reale positiva. L'unica funzione reale che ci assicura tutte le proprietà menzionate e che può rappresentare la *(6)*, è il logaritmo. Infatti, prendendo i logaritmi la *(6)* diventa

$$\mathcal{H}\,(p_1,\ p_2, \dots , p_n) = -\sum_{i\in I} p_i\,[\log_k(n_i) - \log_k(n)] = -\sum_{i\in I} p_i \log_k \frac{n_i}{n} \tag{7}$$

e ricordando che $n_i = 1$ e $1/n = p_i$, otteniamo

$$\mathcal{H}\,(p_1,\ p_2, \dots , p_n) = -\sum_{i\in I} p_i \log_k(p_i) \tag{8}$$

che è una funzione non negativa, con valori reali in Δ^n, additivamente separabile rispetto ai suoi argomenti. Notiamo che dalla *(8)* $\mathcal{H}\,(p_1,\ p_2, \dots , p_n)$ dipende da una costante reale k presa come base dei logaritmi, cosicché $\mathcal{H}\,(p_1,\ p_2, \dots , p_n)$ è unica a meno di una costante reale positiva k. Per caratterizzare la costante k diamo il seguente:

Lemma 1. Se $\mathcal{H}\,(X) = -\sum_{i\in I} p_i \log_k p_i$ è una funzione di entropia, allora $k > 1$.

Dimostrazione: supponiamo di decidere su $X = \{x_i;\ i \in I\}$, dove $|X| = n$. Applicando l'equazione *(8)*, l'entropia associata alla scelta su X è: $\mathcal{H}\,(X) = -\log_k \frac{1}{n}$. Supponiamo di decidere su un nuovo insieme $X' = \{x_i;\ i = 1,\ 2,\ \dots\ ,\ n+t\}$ con $|X'| = n+t$, e $t > 0$. Sia $\mathcal{P}' = \{p'_i\ ;\ i = 1, 2, \dots , n+t\}$ una distribuzione uniforme su X'. L'entropia associata alla seconda scelta è: $\mathcal{H}\,(X') = -\log_k \frac{1}{n+t}$. Poiché $|X'| > |X|$, allora per l'assioma 4 deve essere $\mathcal{H}(X) < \mathcal{H}(X') \Rightarrow k > 1$.

Sulla base dell'ultimo risultato abbiamo completato la prima parte della dimostrazione del teorema. Ora dimostriamo che l'entropia di Shannon soddisfa gli assiomi 1, 2, 3 e 4. È semplice verificare che l'entropia di Shannon $\mathcal{H}(p_1, p_2, \dots, p_n) \geq 0$ per ogni $p_i \in [0, 1]$. Inoltre essa è continua ed additivamente separabile, e da ciò segue che gli assiomi 1 e 2 sono soddisfatti. In Shannon [22] si dimostra che $0 \leq \mathcal{H} \leq \log_k(n)$. Poiché $\lim_{p_i \to 0+} (-p_i \log_k p_i) = 0$, per la *(8)* risulta $\mathcal{H}(X) = \min[\mathcal{H}(X)] = 0$ e quindi la distribuzione di probabilità su X è degenere. Inoltre $\mathcal{H}(X) = \max[\mathcal{H}(X)] = \log_k(n)$, quando i risultati di X hanno uguale probabilità, e quindi anche l'assioma 3 è verificato. Verifichiamo che l'entropia di Shannon soddisfa l'assioma 4. Tornando all'inizio della dimostrazione, consideriamo nuovamente l'equazione *(4)*, che diventa

$$\text{(9)} \qquad \mathcal{H}(p_1, p_2, \dots, p_n) = -\sum_{i \in I} q_i \log_k(q_i) - \sum_{i \in I} q_i \left(-\sum_{i \in I} \pi_i^i \log_k(\pi_i^i)\right)$$

Poiché $\pi_i^i \in \Pi_i = 1/n_i = 1$ per ogni $i \in I$, $-\sum_{i \in I} \pi_i^i \log_k(\pi_i^i) = 0$, inoltre $q_i = p_i$ per ogni $i \in I$ e la *(9)* diventa

$$\text{(10)} \qquad \mathcal{H}(p_1, p_2, \dots, p_n) = -\sum_{i \in I} p_i \log_k(p_i)$$

La dimostrazione del teorema è ora completata.

Nel contesto della teoria dell'informazione, diversamente dal nostro, l'entropia è misurata impiegando un'unità di misura. Anche se l'unità di misura più utilizzata è il Bit (*binary digit units*), ve ne sono anche altre, ed ognuna di esse si caratterizza per la base del logaritmo impiegata nella *(8)*. Nel contesto della teoria delle decisioni, l'entropia ha il ruolo di ordinare le distribuzioni di probabilità in base al grado di incertezza. Nel nostro lavoro l'entropia, in relazione allo spazio delle distribuzioni, assume un ruolo simile a quello svolto da una funzione di utilità sullo spazio delle lotterie. Sebbene non abbiamo bisogno di un'unità di misura, possiamo introdurre un piccolo legame con la teoria dell'informazione. Sia X variabile casuale che assume due valori con probabilità p ed $(1 - p)$; fissando la base dei logaritmi pari a 2, consideriamo un caso *standard* in cui $p = 1/2$ e definiamo $\mathcal{H}(1/2) = 1\text{Bit}$. Il Bit, come qualsiasi unità di misura, si riferisce ad un caso *stan-*

dard. Ciò vuol dire che se noi abbiamo una distribuzione Y, e risulta $\mathcal{H}(Y) = 2\text{Bit}$, diremo che Y è due volte più incerta rispetto al caso *standard*. In questo senso diciamo che la quantità di scelta necessaria per decidere su Y è doppia rispetto al caso *standard*.

4. - L'utilità attesa ed i suoi paradossi

Lo scopo di questo paragrafo è quello di riportare alcuni risultati relativi alla teoria di Von Neumann e Morgenstern [23] ed i suoi risultati paradossali. Poiché vogliamo impiegare un'esposizione omogenea, abbiamo bisogno di fissare il significato dei simboli utilizzati d'ora innanzi.

Dato un insieme di indici $I = \{1, 2, \dots, n\}$, utilizzeremo le lettere maiuscole per indicare gli insiemi di premi, ad esempio $X = \{x_i; i \in I\}$, mentre in lettere minuscole, ad esempio x_i, sarà indicato il singolo risultato. Gli insiemi di risultati devono essere intesi come una lista di premi esaustivi e mutuamente esclusivi. $\mathcal{P}$ è una misura di probabilità su X, ovvero $\mathcal{P} = \{p(x_i); i \in I\}$, dove $p(x_i)$ è la probabilità di ottenere il risultato $x_i \in X$, ovvero $p(x_i) \geq 0$ e $\Sigma_{i\in I}\, p(x_i) = 1$. Indicheremo una lotteria in lettere minuscole ed in grassetto, ad esempio $\boldsymbol{x} = \{x_i, p(x_i); i \in I\}$. Generalmente in letteratura una lotteria è rappresentata come l'insieme delle probabilità dei premi, ma tenendo conto che un risultato può essere visto esso stesso come una lotteria, la notazione sopra indicata è preferita. L'insieme di tutte le lotterie sarà indicato con $\mathcal{L}$: esso rappresenta lo spazio delle lotterie. Ora introduciamo una relazione di preferenza binaria $\succsim$ su $\mathcal{L}$ la quale costituisce un ordinamento completo su $\mathcal{L}$, se per ogni $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{L}$, risulta $\boldsymbol{x} \succsim \boldsymbol{y}$ oppure $\boldsymbol{x} \precsim \boldsymbol{y}$, e se per $\boldsymbol{x} \succsim \boldsymbol{y}$ e $\boldsymbol{y} \succsim \boldsymbol{z}$ risulta $\boldsymbol{x} \succsim \boldsymbol{z}$. Inoltre se entrambe $\boldsymbol{x} \succsim \boldsymbol{y}$ e $\boldsymbol{x} \precsim \boldsymbol{y}$ sono vere, allora $\boldsymbol{x} \backsim \boldsymbol{y}$, dove $\backsim$ è una relazione di indifferenza transitiva e riflessiva. D'ora innanzi $U(\cdot)$ denota una funzione di utilità sulle lotterie, mentre $u(\cdot)$ rappresenta una funzione sui premi.

La teoria di Von Neumann e Morgenstern asserisce che, sotto un certo numero di assiomi su $\mathcal{L}$ e $\succsim$, esiste una funzione reale $U : \mathcal{L} \to \mathbb{R}$ che rappresenta $\succsim$ su $\mathcal{L}$. La funzione U è unica meno che per una trasformazione affine, ed ha la forma dell'utilità attesa. Questo risultato è ben noto come teorema di Von Neumann-Morgenstern (v. il teorema 3 in in appendice). Secondo questo modello un individuo che assume una decisione in un contesto di incertezza debole, massimizzerebbe il valore atteso della sua utilità.

4.1 *Risultati paradossali*

Come abbiamo già sottolineato, negli ultimi anni la EUT è uno degli argomenti più dibattuti. Sebbene essa non sembri descrivere il comportamento degli individui nelle decisioni, resta ancora largamente applicata. Il problema proposto da Allais [1] e poi da Ellsberg [11] è che le persone si comportano in un modo che non è spiegato dalla EUT. Ora presenteremo i due noti esperimenti che hanno condotto a queste conclusioni.

Allais condusse un esperimento mentale successivamente provato sulle persone, e molti di questi risultati si possono trovare in Kahneman e Tversky [16]. Nell'esperimento di Allais [1] ad un individuo si chiede di compiere due scelte fra quattro lotterie con premi monetari[7]. Viene chiesto prima di scegliere fra $\boldsymbol{a}_1 = \{0, 0; 100, 1\}$ e $\boldsymbol{b}_1 = \{0, 0.01; 100, 0.89; 500, 0.10\}$, e poi tra $\boldsymbol{c}_1 = \{0, 0.89; 100, 0.11\}$ e $\boldsymbol{d}_1 = \{0, 0.90; 500, 0.10\}$. È facile verificare come secondo la EUT se $\boldsymbol{a}_1 \succ \boldsymbol{b}_1$ nella prima scelta, poi dovrebbe essere $\boldsymbol{c}_1 \succ \boldsymbol{d}_1$ nella seconda; o se $\boldsymbol{a}_1 \prec \boldsymbol{b}_1$ allora deve essere $\boldsymbol{c}_1 \prec \boldsymbol{d}_1$. Assumiamo il caso di un individuo rappresentativo, se $u(\cdot)$ è una funzione di utilità sui premi $u(\cdot)$, allora se $\boldsymbol{a}_1 \succ \boldsymbol{b}_1$ segue che:

$$U(\boldsymbol{a}_1) > U(\boldsymbol{b}_1) \Rightarrow u(100) > 0{,}1u(500) + 0{,}89u(100) + 0{,}01u(0)$$

Sostituendo nella precedente disuguaglianza $u(100) = 0{,}1u(100) + 0{,}89u(100) + 0{,}01u(100)$, e sottraendo da ogni membro $0{,}89u(100)$, otteniamo:

$$0{,}11u(100)+0{,}89u(0)>0{,}1u(500)+0{,}90u(0)\Rightarrow U(\boldsymbol{c}_1)>U(\boldsymbol{d}_1)\Rightarrow\boldsymbol{c}_1\succ\boldsymbol{d}_1$$

---

[7] In questi esperimenti abbiamo premi monetari, ma per convenienza non indicheremo la valuta.

Similmente, se avessimo supposto che $\boldsymbol{a}_1 < \boldsymbol{b}_1$, avremmo trovato che $\boldsymbol{c}_1 < \boldsymbol{d}_1$. Negli esperimenti condotti sulle persone molti scelgono a_1 nella prima decisione e $\boldsymbol{d}_1$ nella seconda[8]. Questa scelta anomala è conosciuta come "paradosso di Allais". Molti esperimenti hanno confermato questo risultato, e ciò ci consente di concludere che la EUT rappresenta, al limite, una parte del vero modello che guida le decisioni individuali, sempre che un vero modello esista.

Un esperimento diverso fu proposto da Ellsberg [11]. Nell'esperimento di Ellsberg, diversamente da quello di Allais, le persone non conoscono le probabilità di ogni evento. Un'urna contiene 90 palline sistemate come segue: 30 palline sono rosse, 60 sono blu o gialle (dove l'esatta proporzione è ignota). Una pallina viene estratta casualmente dall'urna, e legando ogni estrazione ad un premio monetario, si chiede di scegliere fra due possibilità. Prima si sceglie fra:

$$A: = \begin{cases} 100, & \text{se rossa} \\ 0, & \text{se blu o gialla} \end{cases} \quad \text{e} \quad B: = \begin{cases} 100, & \text{se blu} \\ 0, & \text{se rossa o gialla} \end{cases}$$

e poi fra:

$$C: = \begin{cases} 100, & \text{se rossa o gialla} \\ 0, & \text{se blu} \end{cases} \quad \text{e} \quad D: = \begin{cases} 100, & \text{se blu o gialla} \\ 0, & \text{se rossa} \end{cases}$$

Se $p(r)$, $p(b)$, $p(y)$ sono le probabilità che la pallina rossa, la blu e la gialla vengano estratte, un individuo dotato di comune razionalità dovrebbe stimare $p(r) = 1/3$ e $p(b) + p(y) = 2/3$, da cui $p(r) + p(b) + p(r) = 1$. Possiamo ricondurre le due scommesse a due scelte su quattro lotterie. Infatti la prima scelta è fra le lotterie $\boldsymbol{a}_2 = \{100,\ 1/3;\ 0,\ 2/3\}$ e $\boldsymbol{b}_2 = \{100,\ p(b);\ 0,\ 1/3 + p(y)\}$, e la seconda è fra le lotterie $\boldsymbol{c}_2 = \{100,\ 1/3 + p(y);\ 0,\ p(b)\}$ e $\boldsymbol{d}_2 = \{100,\ 2/3;\ 0,\ 1/3\}$. Negli esperimenti, molte persone preferiscono A a B, e D a C. Se

[8] Nella versione di questo esperimento condotta da KAHNEMAN D. - TVERSKY A. [16] la proporzione di persone che assume questo tipo di decisione è dell'82% per la prima scelta e l'83% per la seconda (*Problem* 1 e 2).

gli individui fossero ispirati dal modello dell'utilità attesa, allora dovrebbe risultare $U(\boldsymbol{a}_2) > U(\boldsymbol{b}_2)$ e $U(\boldsymbol{d}_2) > U(\boldsymbol{c}_2)$, ma questo è possibile se e solo se la stima soggettiva della probabilità $p(b)$ fosse $p(b) < 1/3$ nella prima scelta e $p(b) > 1/3$ nella seconda. Questo risultato è ben noto come paradosso di Ellsberg. Questo tipo di esperimenti vengono solitamente discussi in lavori in cui si prendono in considerazione le probabilità soggettive, considerando questo problema differente rispetto a quello posto da Allais. Come vedremo più avanti, si può sostenere che l'esperimento di Allais e quello di Ellsberg abbiano qualcosa in comune, e quindi le implicazioni del nostro modello saranno analizzate con riferimento sia all'esperimento di Allais che a quello di Ellsberg.

5. - Scelte in condizioni di incertezza e decidibilità

Come si è già affermato, il modello di utilità attesa descrive un modo molto semplice di pensare alla razionalità umana. Tuttavia occorre chiedersi fino a che punto la EUT riesca ad interpretare la logica delle scelte umane in condizioni di incertezza.

Il primo problema con il quale dobbiamo confrontarci è quello di definire l'oggetto della scelta in condizioni di incertezza. Nella letteratura spesso troviamo come oggetto di decisione la lotteria, ma non è chiaro cosa sia realmente una lotteria. Nella teoria di Von Neumann e Morgenstern il problema di un agente è quello di scegliere fra lotterie, dove una lotteria è definita come l'insieme delle probabilità sui possibili risultati. Naturalmente se una lotteria è semplicemente una distribuzione di probabilità, potrebbe sembrare illogico dire che una persona preferisce una distribuzione di probabilità a un'altra. Noi pensiamo che le persone traggano la loro utilità dai premi: in altre parole, le persone mangiano mele e non la probabilità di ottenere mele. Tuttavia, l'idea è che le persone effettuano le loro scelte basandosi sulle lotterie intese come combinazioni di premi e delle relative probabilità. Questo concetto è molto importante perché implica che il processo decisionale è basato sulle lotterie considerate come un singolo oggetto di decisione comprendente sia probabilità che premi.

Un'altra questione da affrontare riguarda gli effetti prodotti dalla decisione su un individuo. Quando noi dobbiamo assumere una decisione su qualcosa di incerto avvertiamo due distinte sensazioni: *(i)* una sensazione di benessere legata a ciò che ci aspettiamo di ottenere dalla decisione. Ogni possibile risultato ci soddisfa con un'intensità che dipende dal suo grado di utilità e dalla sua plausibilità, e questo è un effetto positivo; *(ii)* uno *stress* psicologico legato al fatto di essere indecisi. La necessità di dover assumere una decisione è stressante, e questo è un effetto negativo la cui intensità dipende dal livello di incertezza che caratterizza il processo decisionale.

Riteniamo che la EUT catturi soltanto il primo dei due effetti. Facciamo un esempio: se abbiamo la lotteria di premi monetari $\boldsymbol{x} = \{10,\ 1/2;\ 7,\ 1/4;\ 5,\ 1/4\}$, l'utilità attesa della lotteria è data da $U(\boldsymbol{x}) = 1/2u(10)+1/4u(7)+1/4u(5)$. È chiaro che la funzione $U(\boldsymbol{x})$ incorpora soltanto l'effetto positivo — in termini di soddisfazione attesa dal consumo dei premi – misurato dalle utilità $u(\cdot)$ sui premi ponderate con le probabilità di questi ultimi. Questo è il motivo per cui la EUT non cattura tutti gli elementi caratterizzanti il processo di decisione. In altri termini, la funzione di utilità attesa ignora l'effetto negativo determinato dall'esistenza di incertezza debole. Infatti si è sottolineato che sebbene nella funzione di utilità attesa vengano prese in considerazione le probabilità dei singoli risultati, esse non possono essere assunte come misure di incertezza.

Ma c'è un ultimo problema da prendere in considerazione. Supponiamo di considerare ancora la lotteria $\boldsymbol{x}$, e di confrontarla con una nuova lotteria:

$$\boldsymbol{y} = \{10,\ 1/2;\ 7,\ 1/6;\ 5,\ 1/3\}$$

per cui $U(\boldsymbol{y}) = 1/2u(10)+1/6u(7)+1/3u(5)$. Secondo la EUT, il peso 1/2 attribuito al risultato 10 è uguale in $U(\boldsymbol{x})$ ed $U(\boldsymbol{y})$. Infatti abbiamo $1/2u(10)$ sia in $U(\boldsymbol{x})$ che $U(\boldsymbol{y})$. Tuttavia, se si volesse tener conto del fatto che la seconda lotteria è caratterizzata da maggiore incertezza della prima, il peso attribuito a $u(10)$ in $U(\boldsymbol{x})$ dovrebbe essere maggiore che in $U(\boldsymbol{y})$.

Sulla base di queste osservazioni concludiamo che un modello di scelta dovrebbe avere le seguenti caratteristiche: *(i)* gli individui scelgono sulla base dell'utilità della decisione, ed essa è concepita come diversa dall'utilità della lotteria. Infatti l'utilità della decisione è composta da due componenti: il primo è costituito dal vantaggio atteso che si otterrà dopo che la decisione è stata assunta, e questo è un effetto positivo; l'altra componente è legata al grado di *stress* psicologico, o allo sforzo mentale determinato dall'incertezza, e questo è un effetto negativo; *(ii)* il peso attribuito all'utilità di un premio non dovrebbe essere semplicemente uguale alla probabilità di quest'ultimo. L'utilità di ogni risultato deve assumere un peso che dipende sia dalla sua probabilità, sia dall'incertezza legata alla lotteria. In tal modo, ogni risultato di una lotteria darebbe un contributo — in termini di utilità della decisione — che è rapportato all'incertezza che caratterizza la lotteria.

Dopo aver tracciato le proprietà che un modello di scelta dovrebbe avere, introdurremo un nuovo concetto: la "decidibilità". Sembra che nell'ambito della teoria delle decisioni, questo concetto non sia mai stato utilizzato prima. Perciò occorre chiarire che il significato che si attribuisce al termine decidibilità in questo lavoro è differente da quello inteso nella teoria degli algoritmi o nell'ambito della logica matematica. In base alle osservazione fatte sin qui diamo la seguente

DEFINIZIONE 1 (decidibilità). Data $x \in \mathcal{L}$, siano $U : \mathcal{L} \to \mathbb{R}$ e $\mathcal{H} : \mathcal{L} \to \mathbb{R}$ rispettivamente una funzione di utilità ed una misura di incertezza debole su $\boldsymbol{x} \in \mathcal{L}$. Sia $\boldsymbol{x}' \in \mathcal{L}$, e siano $\Delta U(\boldsymbol{x}) = U(\boldsymbol{x}) - U(\boldsymbol{x}')$ e $\Delta\mathcal{H}(\boldsymbol{x}) = \mathcal{H}(\boldsymbol{x}) - \mathcal{H}(\boldsymbol{x}')$ per ogni $\boldsymbol{x} \in \mathcal{L}$, $\boldsymbol{x} \neq \boldsymbol{x}'$. Una funzione $\mathcal{D} : \mathcal{L} \to \mathbb{R}$, che ha le seguenti proprietà: *(i)* $\Delta\mathcal{D}(\boldsymbol{x}) \cdot \Delta U(\boldsymbol{x}) > 0$; *(ii)* $\Delta\mathcal{D}(\boldsymbol{x}) \cdot \Delta\mathcal{H}(\boldsymbol{x}) < 0$, dove $\Delta\mathcal{D}(\boldsymbol{x}) = \mathcal{D}(\boldsymbol{x}) - \mathcal{D}(\boldsymbol{x}')$, è detta decidibilità della lotteria $\boldsymbol{x} \in \mathcal{L}$.

La decidibilità di una lotteria $\boldsymbol{x}$ ci dà l'utilità della decisione di scegliere $\boldsymbol{x}$ e, come già si è osservato, essa è completamente differente dall'utilità di $\boldsymbol{x}$. La funzione di decidibilità cattura i due effetti menzionati attraverso le due proprietà: *(i)* cattura l'effetto positivo di ottenere un risultato, dove $U(\boldsymbol{x})$ è l'utilità della lotteria $\boldsymbol{x}$; *(ii)* cattura la disutilità del processo di decisione. Infatti, quando il grado di incertezza cresce, il processo di decisione diventa arduo e faticoso, cosicché l'utilità della decisione — che abbiamo chiamato decidibilità — decresce.

Sulla base delle osservazioni fatte in questo paragrafo, l'idea proposta è che l'individuo, nell'assumere una decisione, non guarda all'utilità attesa della lotteria, bensì alla decidibilità della lotteria. L'obiettivo da massimizzare sarebbe la funzione di decidibilità e non quella di utilità attesa.

6. - Rappresentazione delle preferenze attraverso una funzione di decidibilità

L'obiettivo di questo paragrafo è quello di stabilire, a partire da un certo numero di assiomi su $\mathcal{L}$ e $\succsim$, che è possibile rappresentare $\succsim$ attraverso una funzione di decidibilità. L'assiomatizzazione proposta qui è molto simile a quella proposta a sostegno della EUT. Introdurremo quattro assiomi e preferiremo un'assiomatizzazione semplificata finalizzata alla dimostrazione del teorema 2. Di conseguenza definiremo gli assiomi in modo tale da non dover ricorrere alla dimostrazione di teoremi intermedi.

Assioma 5 (completezza). $\mathcal{L}$ è completamente ordinato da $\succsim$.

Assioma 6 (continuità ed unicità). Per ogni $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{L}$ tali che $\boldsymbol{x} \succ \boldsymbol{y} \succsim \boldsymbol{z} \vee \boldsymbol{x} \succsim \boldsymbol{y} \succ \boldsymbol{z}$ è vera, esiste un unico $\beta \in [0, 1]$ tale che $\boldsymbol{y} \backsim \beta\boldsymbol{x} + (1 - \beta)\boldsymbol{z}$.

Assioma 7 (indipendenza). Per ogni $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{L}$, dove $\boldsymbol{x} \sim \boldsymbol{y}$, e $\beta \in (0, 1)$, allora $\beta\boldsymbol{x} + (1 - \beta)\boldsymbol{z} \backsim \beta\boldsymbol{y} + (1 - \beta)\boldsymbol{z}$.

Assioma 8 (confronto). Per ogni $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w} \in \mathcal{L}$ e $\boldsymbol{x} \succsim \boldsymbol{z}$, se esistono $\alpha, \beta \in [0, 1]$ tali che $\boldsymbol{y} \backsim \alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{z}$ e $\boldsymbol{w} \backsim \beta\boldsymbol{x} + (1 - \beta)\boldsymbol{z}$, allora $\boldsymbol{y} \succ \boldsymbol{w}$ se e solo se $\alpha > \beta$ e $\boldsymbol{y} \backsim \boldsymbol{w}$ se e solo se $\alpha = \beta$.

Il primo assioma dovrebbe essere familiare dalla teoria tradizionale. Alcune volte esso viene identificato col nome di assioma di "ordinamento debole". L'assioma 6 incorpora sia la proprietà di continuità di $\succsim$, sia la cosiddetta proprietà dell'"unicità delle soluzioni". Questa proprietà potrebbe essere ottenuta partendo da un'assunzione più semplice e naturale sulla continuità di $\succsim$ (v. Herstein e Milnor [15]). Questo modo di procedere sarebbe stato più rigoroso, ma avrebbe determinato una presentazione eccessivamente laboriosa se consideriamo che il nostro unico obiettivo è la dimostrazione del teorema 2. L'assioma 7 è l'assioma di indipendenza nella sua versione debole basata sulla relazione $\backsim$; questa formulazione viene impiegata in Herstein e Milnor [15]. L'assioma 8 ci permette di confrontare due lotterie basandoci sul confronto fra α e β. Chiaramente l'esistenza di $\alpha, \beta \in [0, 1]$ nell'assioma 8 è legata alle condizioni imposte dall'assioma 6. Diamo ora il seguente:

TEOREMA 2. Sulla base degli assiomi 5, 7, 6 e 8, per ogni $x, y \in \mathcal{L}$ esiste una funzione di decidibilità $\mathcal{D} : \mathcal{L} \rightarrow \mathbb{R}$ su $(\mathcal{L}, \succsim)$ che rappresenta $\succsim$, ovvero $\boldsymbol{x} \succsim \boldsymbol{y} \Leftrightarrow \mathcal{D}(\boldsymbol{x}) \geq \mathcal{D}(\boldsymbol{y})$. Inoltre, $\mathcal{D}$ non è unica.

DIMOSTRAZIONE. Procederemo in due passaggi: prima proveremo che gli assiomi sulle preferenze implicano una rappresentazione in termini di decidibilità, e successivamente mostreremo che la funzione di decidibilità ottenuta soddisfa gli assiomi. Supponiamo di avere due lotterie $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{L}$. Per gli assiomi precedenti e per il teorema 3 (v. l'appendice) esiste una funzione di utilità attesa $U : \mathcal{L} \rightarrow \mathbb{R}$ tale che $\boldsymbol{x} \succsim \boldsymbol{y} \Leftrightarrow U(\boldsymbol{x}) \geq U(\boldsymbol{y})$. Consideriamo ancora due lotterie $\boldsymbol{l}_1, \boldsymbol{l}_2 \in \mathcal{L}$ tali che entrambe $\boldsymbol{l}_1 \succ \boldsymbol{x} \succsim \boldsymbol{l}_2 \vee \boldsymbol{l}_1 \succsim \boldsymbol{x} \succ \boldsymbol{l}_2$ e $\boldsymbol{l}_1 \succ \boldsymbol{y} \succsim \boldsymbol{l}_2 \vee \boldsymbol{l}_1 \succsim \boldsymbol{y} \succ \boldsymbol{l}_2$ siano vere, dove $\boldsymbol{l}_1, \boldsymbol{l}_2$ sono rispettivamente la migliore e la peggiore lotteria in $\mathcal{L}$. Applicando l'assioma 6, troviamo $\alpha, \beta \in [0, 1]$ unici, e tali che possiamo scrivere $\boldsymbol{x}, \boldsymbol{y}$ come $\boldsymbol{x} \backsim \alpha\boldsymbol{l}_1 + (1 - \alpha)\boldsymbol{l}_2$ e $\boldsymbol{y} \backsim \beta\boldsymbol{l}_1 + (1 - \beta)\boldsymbol{l}_2$. L'assioma 8 implica che $\boldsymbol{x} \succsim \boldsymbol{y} \Leftrightarrow \alpha \geq \beta$, e per il teorema 3 (in appendice) otteniamo la condizione:

$$\begin{aligned} (11) \qquad & \boldsymbol{x} \succsim \boldsymbol{y} \Leftrightarrow \alpha \geq \beta \Leftrightarrow U(\boldsymbol{x}) \geq U(\boldsymbol{y}) \Leftrightarrow \\ & \Leftrightarrow \alpha U(\boldsymbol{l}_1) + (1 - \alpha)U(\boldsymbol{l}_2) \geq \beta U(\boldsymbol{l}_1) + (1 - \beta)U(\boldsymbol{l}_2) \end{aligned}$$

Data una lotteria $\boldsymbol{x} = \{\boldsymbol{l}_1, \alpha; \boldsymbol{l}_2, (1 - \alpha)\}$ possiamo calcolarne l'entropia[9] come $\mathcal{H}(\boldsymbol{x}) = - \alpha \ln(\alpha) - (1 - \alpha) \ln(1 - \alpha)$, e lo stesso vale per $\boldsymbol{y}$. Ma $\mathcal{H}(\boldsymbol{x})$ può essere scritta come funzione di α, ed essendo $0 \leq \mathcal{H}(\alpha) \leq \ln(2)$, ne segue che $0 \leq \mathcal{H}(\alpha)/\ln(2) \leq 1$.

[9] La scelta della base dei logaritmi non è importante, così scegliamo *e*.

Fissiamo $\hbar(\alpha) = (\mathcal{H}(\alpha)/\ln(2)) + 1$, così $1 \leq \hbar(\alpha) \leq 2$. Costruiamo una nuova funzione definita come $\omega(\alpha) = \alpha / \hbar(\alpha)$, dove $0 \leq \omega(\alpha) \leq 1$. Lo stesso vale per la lotteria $\boldsymbol{y}$. Abbiamo così ottenuto due funzioni che chiamiamo "funzioni peso", definite nel modo seguente:

(12) $$\omega(\alpha) = \frac{\alpha}{\hbar(\alpha)} \text{ e } \omega(\beta) = \frac{\beta}{\hbar(\beta)} \quad \text{dove: } 0 \leq \omega(\alpha),\ \omega(\beta) \leq 1.$$

Le funzioni $\omega(\alpha)$ e $\omega(\beta)$ sono funzioni continue, e crescenti ω: $[0, 1] \rightarrow [0, 1]$, tali che se $\alpha \geq \beta \Leftrightarrow \omega(\alpha) \geq \omega(\beta)$. Prendiamo in considerazione la seguente:

(13) $$\omega(\alpha)U(\boldsymbol{l}_1)+\omega(1-\alpha)U(\boldsymbol{l}_2) \geq \omega(\beta)U(\boldsymbol{l}_1)+\omega(1-\beta)U(\boldsymbol{l}_2)$$

dove $\omega(\alpha)U(\boldsymbol{l}_1)+\omega(1-\alpha)U(\boldsymbol{l}_2)$ e $\omega(\beta)U(\boldsymbol{l}_1)+\omega(1-\beta)U(\boldsymbol{l}_2)$ sono due funzioni di decidibilità secondo la definizione 1. Infatti possiamo scrivere:

(14) $$\mathcal{D}(\boldsymbol{x}) = \omega(\alpha)U(\boldsymbol{l}_1) + \omega(1-\alpha)U(\boldsymbol{l}_2) = \frac{\alpha}{1+\dfrac{\mathcal{H}(\boldsymbol{x})}{\ln(2)}}U(\boldsymbol{l}_1)+\frac{1-\alpha}{1+\dfrac{\mathcal{H}(\boldsymbol{x})}{\ln(2)}}U(\boldsymbol{l}_2)$$

(15) $$\mathcal{D}(\boldsymbol{y}) = \omega(\beta)U(\boldsymbol{l}_1) + \omega(1-\beta)U(\boldsymbol{l}_2) = \frac{\beta}{1+\dfrac{\mathcal{H}(\boldsymbol{y})}{\ln(2)}}U(\boldsymbol{l}_1)+\frac{1-\beta}{1+\dfrac{\mathcal{H}(\boldsymbol{y})}{\ln(2)}}U(\boldsymbol{l}_2)$$

Supponiamo che $\boldsymbol{x}' \in \mathcal{L}$ dove $\boldsymbol{x} \neq \boldsymbol{x}' \in \mathcal{L}$, sia $\boldsymbol{x}' \in \mathcal{L}$, e $\Delta U(\boldsymbol{x}) = U(\boldsymbol{x}) - U(\boldsymbol{x}')$, $\Delta\mathcal{H}(\boldsymbol{x}) = \mathcal{H}(\boldsymbol{x}) - \mathcal{H}(\boldsymbol{x}')$ e $\Delta\mathcal{D}(\boldsymbol{x}) = \mathcal{D}(\boldsymbol{x}) - \mathcal{D}(\boldsymbol{x}')$, per ogni $\boldsymbol{x} \in \mathcal{L}$, $\boldsymbol{x} \neq \boldsymbol{x}'$. Supponiamo ancora che $\Delta U(\boldsymbol{x}) > 0$ e $\Delta\mathcal{H}(\boldsymbol{x}) > 0$. Dalla *(14)* risulta che $\Delta\mathcal{D}(\boldsymbol{x}) \cdot \Delta U(\boldsymbol{x}) > 0$ e $\Delta\mathcal{D}(\boldsymbol{x}) \cdot \Delta\mathcal{H}(\boldsymbol{x}) < 0$ come nella definizione 1. Lo stesso vale per la lotteria $\boldsymbol{y}$. Ora ritorniamo alla *(13)* e notiamo che $\hbar\,(\alpha) = \hbar\,(1-\alpha)$ e $\hbar\,(\beta) = \hbar\,(1-\beta)$. Inoltre, per $\alpha \geq \beta$ e prendendo $\alpha = 1-\beta$, la *(13)* diventa:

$$\begin{aligned}(16)\qquad \alpha \geq \beta &\Leftrightarrow \mathcal{D}\,(\boldsymbol{x}) \geq \mathcal{D}\,(\boldsymbol{y})\\ &\Leftrightarrow \alpha U\,(\boldsymbol{l}_1) + (1-\alpha)\,U\,(\boldsymbol{l}_2) \geq \beta U\,(\boldsymbol{l}_1) + (1-\beta)\,U\,(\boldsymbol{l}_2)\end{aligned}$$

Questo significa che abbiamo trovato una funzione peso $\omega : [0, 1] \rightarrow [0, 1]$, tale che esistono $\alpha, \beta \in [0, 1]$ che per l'assioma 6 sono unici, e per l'assioma 8 entrambe la *(11)* e la *(16)* sono verificate.

Per completare la prima parte della dimostrazione, occorre provare che $\mathcal{D}$ non è unica. Per il teorema 3 U è unica a meno di una trasformazione affine. Sia V una funzione affine tale che $V : \mathcal{L} \rightarrow \mathbb{R}$, dove $V\,(\boldsymbol{l}_1) = \phi U(\boldsymbol{l}_1) + \mu$ e $V\,(\boldsymbol{l}_2) = \phi U\,(\boldsymbol{l}_2) + \mu$, $\phi, \mu \in \mathbb{R}$ e $\phi > 0$. L'utilità attesa di $\boldsymbol{x}$, $\boldsymbol{y}$ diventa $V\,(\boldsymbol{x}) = \alpha V\,(\boldsymbol{l}_1) + (1-\alpha)\,V\,(\boldsymbol{l}_2)$ e $V\,(\boldsymbol{y}) = \beta V\,(\boldsymbol{l}_1) + (1-\beta)\,V\,(\boldsymbol{l}_2)$. Ora abbiamo due funzioni di decidibilità $\tilde{\mathcal{D}}\,(\boldsymbol{x}) = \omega(\alpha)\,V\,(\boldsymbol{l}_1) + \omega(1-\alpha)\,V\,(\boldsymbol{l}_2)$ e $\tilde{\mathcal{D}}\,(\boldsymbol{y}) = \omega(\beta)\,V\,(\boldsymbol{l}_1) + \omega(1-\beta)\,V\,(\boldsymbol{l}_2)$. Per il teorema 3 abbiamo le condizioni

$$\begin{aligned}(17)\qquad &\boldsymbol{x} \succsim \boldsymbol{y} \Leftrightarrow \alpha \geq \beta \Leftrightarrow V\,(\boldsymbol{x}) \geq V\,(\boldsymbol{y}) \Leftrightarrow\\ &\Leftrightarrow \alpha V\,(\boldsymbol{l}_1) + (1-\alpha)\,V\,(\boldsymbol{l}_2) \geq \beta V\,(\boldsymbol{l}_1) + (1-\beta)\,V\,(\boldsymbol{l}_2)\end{aligned}$$

Come abbiamo già notato, fissando $\alpha = 1 - \beta$ otteniamo:

$$\begin{aligned}(18)\qquad \alpha \geq \beta &\Leftrightarrow \tilde{\mathcal{D}}(\boldsymbol{x}) \geq \tilde{\mathcal{D}}(\boldsymbol{y})\\ &\Leftrightarrow \alpha V\,(\boldsymbol{l}_1) + (1-\alpha)\,V\,(\boldsymbol{l}_2) \geq \beta V\,(\boldsymbol{l}_1) + (1-\beta)\,V\,(\boldsymbol{l}_2)\end{aligned}$$

Così data $\omega : [0, 1] \rightarrow [0, 1]$, tale che esistono $\alpha, \beta \in [0, 1]$ che per l'assioma 6 sono unici, e per l'assioma 8, entrambe la *(17)* e la *(18)* sono soddisfatte, come risultato esiste un'altra funzione di decidibilità e $\tilde{\mathcal{D}}$ che soddisfa gli assiomi 5, 6, 7 e 8. La prima parte della dimostrazione è ora completa.

Ora proveremo che la funzione di decidibilità $\mathcal{D} : \mathcal{L} \rightarrow \mathbb{R}$ soddisfa gli assiomi 5, 6, 7 e 8. Poiché $\mathcal{D}$ è una funzione a valori reali, il primo assioma è soddisfatto. Guardiamo all'assioma 6 e sia $\boldsymbol{y} = \{\boldsymbol{y}_1, \gamma; \boldsymbol{y}_2, (1-\gamma)\}$, $\gamma \in [0, 1]$. Per verificare che $\mathcal{D}\,(\cdot)$ soddisfa l'assioma in questione dobbiamo mostrare che esiste un unico β ∈ [0, 1] tale che $\mathcal{D}(\boldsymbol{y}) = \mathcal{D}(\beta\boldsymbol{x} + (1-\beta)\boldsymbol{z})$. Ma:

$$\mathcal{D}(\boldsymbol{y}) = \mathcal{D}(\beta\boldsymbol{x} + (1-\beta)\boldsymbol{z}) \Leftrightarrow \frac{\gamma U(\boldsymbol{y}_1) + (1-\gamma)U(\boldsymbol{y}_2)}{\hbar(\gamma)} = \tag{19}$$

$$= \frac{\beta U(\boldsymbol{x}) + (1-\beta)\, U(\boldsymbol{y})}{\hbar(\beta)}$$

Fissando $\gamma = 1 - \beta$, la *(19)* diventa:

$$(1-\beta)\, U(\boldsymbol{y}_1) + \beta U(\boldsymbol{y}_2) = \beta U(\boldsymbol{x}) + (1-\beta)\, U(\boldsymbol{z}) \tag{20}$$

e risolvendo la *(20)* per β, troviamo:

$$\beta = \frac{U(\boldsymbol{z}) - U(\boldsymbol{y}_1)}{(U(\boldsymbol{z}) - U(\boldsymbol{y}_1)) + (U(\boldsymbol{x}) - U(\boldsymbol{y}_2))}$$

Poiché $U : \mathcal{L} \to \mathbb{R}$, e $U(\boldsymbol{x}) \geq U(\boldsymbol{y}_2)$, abbiamo ottenuto uno scalare $\beta \in [0, 1]$ tale che l'assioma 6 è soddisfatto dalla $\mathcal{D}(\cdot)$. Consideriamo ora l'assioma 7. Mostreremo che per ogni $\beta \in (0, 1)$ e $\mathcal{D}(\boldsymbol{x}) = \mathcal{D}(\boldsymbol{y})$ segue che:

$$\frac{U(\beta\boldsymbol{x} + (1-\beta)\, \boldsymbol{z})}{\hbar(\beta)} = \frac{U(\beta\boldsymbol{y} + (1-\beta)\, \boldsymbol{z})}{\hbar(\beta)} \tag{21}$$

A causa della linearità della funzione di utilità di Von Neumman-Morgenstern, la *(21)* implica $\beta U(\boldsymbol{x}) = \beta U(\boldsymbol{y})$ che è soddisfatta per ogni $U(\boldsymbol{x}) = U(\boldsymbol{y})$ e $\beta \in (0, 1)$, cosicché l'assioma 7 è verificato. Infine ci resta da verificare l'assioma 8. Sia $\mathcal{D}(\boldsymbol{y}) = \omega(\alpha)\, U(\boldsymbol{x}) + \omega(1-\alpha)\, U(\boldsymbol{z})$ e $\mathcal{D}(\boldsymbol{w}) = \omega(\beta)\, U(\boldsymbol{x}) + \omega(1-\beta)\, U(\boldsymbol{z})$, mostriamo che esiste un unico $\alpha > \beta$ tale che $\mathcal{D}(\boldsymbol{y}) \geq \mathcal{D}(\boldsymbol{w})$. Prendendo $\alpha \geq \beta$ e $\alpha = 1 - \beta$ otteniamo:

$$\mathcal{D}(\boldsymbol{y}) \geq \mathcal{D}(\boldsymbol{w}) \Leftrightarrow (1-\beta)U(\boldsymbol{x}) + \beta U(\boldsymbol{z}) \geq \beta U(\boldsymbol{y}) + (1-\beta)\, U(\boldsymbol{z}) \tag{22}$$

Risolvendo, otteniamo $U(\boldsymbol{x}) \geq U(\boldsymbol{z})$ che è soddisfatta perché $\boldsymbol{x} \succsim \boldsymbol{z}$. L'assioma 8 è soddisfatto e la dimostrazione del teorema è completa.

COROLLARIO 1. Sia Φ: $\mathbb{R} \rightarrow \mathbb{R}$ una funzione continua, monotona e crescente. Se $\mathcal{D} : \mathcal{L} \rightarrow \mathbb{R}$ rappresenta $\succsim$ su $\mathcal{L}$, allora $\Phi(\mathcal{D}(\cdot))$ rappresenta $\succsim$ su $\mathcal{L}$.

DIMOSTRAZIONE. Sia $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{L}$, e supponiamo $\boldsymbol{x} \succsim \boldsymbol{y}$. Per il teorema 2 $\boldsymbol{x} \succsim \boldsymbol{y} \Rightarrow \mathcal{D}(\boldsymbol{x}) \geq \mathcal{D}(\boldsymbol{y})$. Poiché $\Phi(\cdot)$ è una funzione reale continua e monotona, allora $\Phi(\mathcal{D}(\boldsymbol{x})) \geq \Phi(\mathcal{D}(\boldsymbol{y}))$. Per converso, supponiamo che $\Phi(\mathcal{D}(\boldsymbol{x})) \geq \Phi(\mathcal{D}(\boldsymbol{y}))$, cioè $\mathcal{D}(\boldsymbol{x}) \geq \mathcal{D}(\boldsymbol{y})$ e per il teorema 2 segue $\boldsymbol{x} \succsim \boldsymbol{y}$.

In generale se abbiamo una lotteria $\boldsymbol{x} \in \mathcal{L}$, definita sull'insieme dei risultati $X = \{x_i; i = 1, 2, \ldots, n\}$, sia $\mathcal{P} = \{p_i; i = 1, 2, \ldots, n\}$ una distribuzione di probabilità su X, allora la decidibilità di $\boldsymbol{x}$ è data da:

$$\mathcal{D}(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} p_i u(x_i)}{1 - \dfrac{\sum_{i=1}^{n} p_i \ln(p_i)}{\ln(n)}} \tag{23}$$

dove il numeratore è uguale all'utilità attesa di $\boldsymbol{x}$, ed il denominatore è uguale ad una trasformazione lineare crescente dell'entropia di $\boldsymbol{x}$. Inoltre possiamo scrivere $\mathcal{D}(\boldsymbol{x})$ in termini di funzioni peso $\omega(\cdot)$:

$$\mathcal{D}(\boldsymbol{x}) = \sum_{i=1}^{n} \omega(p_i) u(x_i) \tag{24}$$

dove: $$0 \leq \omega(p_i) = \frac{p_i}{\hbar(X)} \leq 1, \ \hbar(X) = 1 + \frac{\mathcal{H}(X)}{\ln(n)}$$

Ciò chiarisce che il peso $\omega(p_i)$ dipende non solo dalla probabilità di x_i, ma da tutte le probabilità della lotteria. La funzione peso $\omega(p_i)$ rappresenta il peso attribuito all'utilità del premio x_i, tenendo conto sia della probabilità del risultato x_i sia dell'incertezza legata all'esito della lotteria. Inoltre, è interessante notare che in questa teoria otteniamo una regola di decisione che dipende dal contenuto informativo della lotteria. Infatti, accettando l'idea, largamente diffusa tra i fisici, che $-\mathcal{H}(\cdot)$ è una misura del contenuto informativo in una distribuzione di probabilità, questa teoria spiega che l'utilità della decisione cresce al crescere del contenuto informativo della lotteria.

7. - Una spiegazione dei paradossi

Cercheremo ora di spiegare il comportamento individuale utilizzando il concetto di decidibilità. L'idea qui proposta è che le persone siano guidate dalla decidibilità e non dall'utilità attesa. Procederemo alla verifica di questo assunto sulla base dei risultati degli esperimenti di Allais e di Ellsberg. A causa della diversità dei due contesti sperimentali, le tecniche utilizzate per questa verifica saranno molto diverse tra loro.

7.1 *Il paradosso di Allais*

Applicheremo il nostro modello ai dati sperimentali forniti dal paradosso di Allais e descritti nel paragrafo 4.1. Per fare questo, dovremo comparare la decidibilità delle lotterie $\boldsymbol{a}_1$, $\boldsymbol{b}_1$, $\boldsymbol{c}_1$ e $\boldsymbol{d}_1$. Ricordando che nell'esperimento di Allais le persone si comportavano in modo tale che $\boldsymbol{a}_1 \succ \boldsymbol{b}_1$ e $\boldsymbol{d}_1 \succ \boldsymbol{c}_1$, il nostro obiettivo è di provare che esiste almeno una funzione di utilità sui premi $u(x_i)$, tale che applicando la *(23)* su $\boldsymbol{a}_1$, $\boldsymbol{b}_1$, $\boldsymbol{c}_1$ e $\boldsymbol{d}_1$, entrambe $\mathcal{D}(\boldsymbol{a}_1) > \mathcal{D}(\boldsymbol{b}_1)$ e $\mathcal{D}(\boldsymbol{d}_1) > \mathcal{D}(\boldsymbol{c}_1)$ siano vere. In altre parole, dobbiamo trovare una funzione di utilità dei premi $u(x_i)$ tale che il seguente sistema di disequazioni:

$$\begin{cases} \mathcal{D}(\boldsymbol{a}_1) - \mathcal{D}(\boldsymbol{b}_1) > 0 \\ \mathcal{D}(\boldsymbol{d}_1) - \mathcal{D}(\boldsymbol{c}_1) > 0 \end{cases} \tag{25}$$

sia verificato. Tuttavia, se utilizzassimo una funzione molto specifica (ad esempio $u(x_i) = \sqrt{x_i}$, $u(x_i) = \log(x_i)$ etc.) il risultato fornito sarebbe limitativo. Per superare questo problema, ci serviremo di una funzione di utilità appartenente alla classe CRRA (*constant relative risk aversion*). Sia x_i un possibile risultato, allora la funzione utilizzata sarà:

$$u_\gamma(x_i) = \frac{x_i^\gamma}{\gamma} \quad \text{per ogni } \gamma \in \mathbb{R} - \{0\}. \tag{26}$$

Questa funzione è largamente utilizzata nella teoria della finanza, ed ha la proprietà che la sua forma varia al variare del parametro γ. Per $\gamma < 1$ $u_\gamma(x_i)$ è una funzione concava, se $\gamma = 1$ allora $u_\gamma(x_i)$ è lineare, e per $\gamma > 1$ $u_\gamma(x_i)$ è convessa. La *(26)* ha un coefficiente di avversione relativa al rischio costante. Se ϱ e ϱ_R rappresentano gli indici di Arrow-Pratt (Arrow [2] e Pratt [19]), che sono rispettivamente il coefficiente di avversione assoluta al rischio ed il coefficiente di avversione relativa al rischio, allora

$$\varrho(x_i) = -\frac{\dfrac{\partial^2 u(x_i)}{\partial x_i^2}}{\dfrac{\partial u(x_i)}{\partial x_i}} = \frac{1-\gamma}{x_i} \qquad \text{cosicché:} \qquad \frac{d\varrho(x_i)}{dx_i} = -\frac{1-\gamma}{x_i^2}$$

e: $$\varrho_R(x_i) = x_i \varrho(x_i) = 1-\gamma \quad \text{cosicché:} \qquad \frac{d\varrho_R(x_i)}{dx_i} = 0$$

Data la lotteria $\boldsymbol{x} = \{x_i, p_i;\ i = 1, 2, \ldots, n\}$, allora per la *(24)* e la *(23)* abbiamo:

$$\mathcal{D}_\gamma(\boldsymbol{x}) = \sum_{i=1}^{n} \omega(p_i)\frac{x_i^\gamma}{\gamma} = \frac{\dfrac{1}{\gamma}\displaystyle\sum_{i=1}^{n} p_i x_i^\gamma}{1 - \dfrac{1}{\ln(n)}\displaystyle\sum_{i=1}^{n} p_i \ln p_i} \tag{27}$$

Le lotterie $\boldsymbol{a}_1$, $\boldsymbol{b}_1$, $\boldsymbol{c}_1$ e $\boldsymbol{d}_1$, sono completamente note ed applicando ad esse la *(27)*, le due disequazioni del sistema *(25)* diventano funzioni del parametro y, ovvero il sistema *(25)* diventa

$$(28) \qquad \begin{cases} \mathcal{D}_{\boldsymbol{a}_1}(\gamma) - \mathcal{D}_{\boldsymbol{b}_1}(\gamma) > 0 \\ \mathcal{D}_{\boldsymbol{d}_1}(\gamma) - \mathcal{D}_{\boldsymbol{c}_1}(\gamma) > 0 \end{cases}$$

Risulta molto difficile trovare soluzioni analitiche per il sistema *(28)*, ma non è difficile trovare soluzioni numeriche. È risultato che per $0{,}005 < \gamma < 0{,}998$ il sistema *(28)* è verificato. Poiché il risultato $\gamma \in (0{,}005,\ 0{,}998)$ è la stesso che dire $0{,}002 < \varrho_R(\gamma) < 0{,}95$, significa che:

(i) esiste un insieme di funzioni della classe CRRA:

$$\mathcal{U} = \left\{ \frac{x^\gamma}{\gamma};\ 0{,}05 < \gamma < 0{,}998 \right\}$$

tale che è possibile spiegare i risultati del paradosso di Allais secondo la teoria della decidibilità;

(ii) il paradosso di Allais è spiegato dalla nostra teoria per un basso livello del coefficiente di avversione relativa al rischio.

In Kahneman e Tversky [16] sono proposti molti esperimenti relativi al paradosso di Allais. Questa sperimentazione comprende anche il cosiddetto paradosso di Allais inverso, che si ottiene quando si cambia il segno dei premi nella formulazione originaria[10] . Finora la letteratura non è stata in grado di risolvere tale paradosso. Abbiamo applicato la teoria della decidibilità — seguendo la stessa tecnica proposta in questo paragrafo — a tutti gli esperimenti che in Kahneman e Tversky [16] sono battezzati come Problems 1, 2, 3, 4, 5, 6, 1', 2'[11] , che sono formulazioni alternative del paradosso di Allais. Attraverso questo *test* abbiamo scoperto risultati molti simili a quelli spiegati nei punti *(i)* e *(ii)* sopra, cosicché possiamo concludere che la teoria della decidibilità può essere assunta come una buona approssimazione del modello che guida le persone nelle scelte in condizioni di incertezza debole.

---

[10] Per esempio quando la lotteria $a_1 = \{0, 0; 100, 1\}$ diventa $a'_1 = \{0, 0; -100, 1\}$.

[11] In Kahneman D. - Tversky A. [16] *Problems* 1' e 2', si riferiscono al paradosso di Allais inverso.

7.2 *Il paradosso di Ellsberg*

In questo lavoro abbiamo sviluppato una teoria strettamente connessa ad una situazione di incertezza debole, così come è stato definito nel paragrafo 2.1. L'esperimento di Ellsberg [11] descritto nel paragrafo 4.1 riguarda invece una situazione che abbiamo chiamato incertezza forte, che corrisponde alla cosiddetta incertezza knightiana o ambiguità come viene definita in Ellsberg [11] e successivamente in Camerer [8]. Perché applichiamo la teoria della decidibilità al paradosso di Ellsberg? La risposta è che se le probabilità dei possibili risultati sono ignote, allora una volta che un individuo abbia formato le sue stime soggettive, egli dovrebbe ispirarsi ad un dato modello di decisione. In altri termini, il momento in cui si formulano le probabilità soggettive è distinto dal momento in cui viene effettuata la scelta. Pertanto, quando una teoria coerente ha spiegato il modo in cui si formulano le probabilità soggettive, il modo in cui si effettuano le scelte viene spiegato da un modello di decisione. Nel nostro modo di vedere, la logica della predizione (come viene chiamata in De Finetti [10]) e la logica della scelta sono due elementi distinti del processo decisionale. In un contesto di incertezza debole gli individui non devono effettuare stime soggettive delle probabilità, perché le probabilità fanno parte dell'insieme delle informazioni. La nostra idea è che, una volta compiuta la stima soggettiva delle probabilità ignote, il modello di decisione che guida le persone deve essere simile a quello utilizzato nel caso in cui le probabilità sono note.

Ora ritorniamo all'esperimento di Ellsberg. Perché il nostro modello possa essere assunto come il modello a cui si ispirano le persone, esso deve essere in grado di spiegare il loro comportamento e al contempo essere compatibile con una stima coerente delle probabilità incognite. Utilizzeremo $\hat{p}(b)$, $\hat{p}(g)$ per indicare la stima soggettiva di $p(b)$, $p(g)$ nel paragrafo 4.1. Ricordando che nel caso dell'esperimento di Ellsberg $\boldsymbol{a}_2 \succ \boldsymbol{b}_2$ e $\boldsymbol{d}_2 \succ \boldsymbol{c}_2$, il nostro obiettivo è verificare se esiste una stima coerente delle probabilità incognite tale che il sistema:

$$(29)\qquad \begin{cases} \mathcal{D}(\boldsymbol{a}_2) > \mathcal{D}(\boldsymbol{b}_2) \\ \mathcal{D}(\boldsymbol{d}_2) > \mathcal{D}(\boldsymbol{c}_2) \end{cases}$$

sia verificato. Ci aspettiamo che una stima coerente delle probabilità incognite sia tale che $1/3 + \hat{p}(b) + \hat{p}(g) = 1$, e $1/3 + \hat{p}(g) = 1 - \hat{p}(b)$, dove 1/3 è la probabilità di estrarre la pallina rossa dall'urna. Si noti che $\hbar\ (1/3,\ 2/3) \cong 1{,}92$. Così, possiamo scrivere il sistema *(29)* come

$$(30)\qquad \begin{cases} \dfrac{0{,}\bar{3}u(100)+0{,}\bar{6}u(0)}{1{,}92} > \dfrac{\hat{p}(b)u(100)+(1-\hat{p}(b))u(0)}{\hbar(\hat{p}(b))} \\[2ex] \dfrac{0{,}\bar{6}u(100)+0{,}\bar{3}u(0)}{1{,}92} > \dfrac{(1-\hat{p}(b))u(100)+\hat{p}(b)u(0)}{\hbar(\hat{p}(b))} \end{cases}$$

che diventa:

$$\begin{cases} U(100)(0{,}\bar{3}\hbar(\hat{p}(b)) - 1{,}92\hat{p}(b)) > U(0)(1{,}92(1-\hat{p}(b)) - 0{,}\bar{6}\hbar(\hat{p}(b))) \\ U(100)(0{,}\bar{6}\hbar(\hat{p}(b)) - 1{,}92(1-\hat{p}(b))) > U(0)(1{,}92\hat{p}(b) - 0{,}\bar{3}\hbar(\hat{p}(b))) \end{cases}$$

Fissando $\phi(\hat{p}(b)) = (0{,}\bar{3}\hbar(\hat{p}(b)) - 1{,}92\hat{p}(b))$, e $\psi(\hat{p}(b)) = (1{,}92(1 - \hat{p}(b)) - 0{,}\bar{6}\hbar(\hat{p}(b)))$, il sistema *(29)* diventa

$$(31)\qquad \begin{cases} \phi(\hat{p}(b))u(100) > \psi(\hat{p}(b))u(0) \\ -\psi(\hat{p}(b))u(100) > -\phi(\hat{p}(b))u(0) \end{cases}$$

Data una qualunque funzione di utilità sui premi, se esiste una probabilità $0 \leq \hat{p}(b) \leq 2/3$ tale che entrambe le disequazioni del sistema *(31)* sono verificate, allora siamo riusciti a spiegare il paradosso di Ellsberg attraverso il modello di decidibilità. Si noti che si potrebbe ripetere lo stesso ragionamento partendo dalla probabilità incognita $\hat{p}(g)$ invece che $\hat{p}(b)$. Supponiamo che $0 \leq u(0) \leq u(100)$. Per quanto riguarda le funzioni $\phi(\hat{p}(b))$ e $\psi(\hat{p}(b))$, si noti che $\phi(\hat{p}(b)) = \psi(\hat{p}(b)) = 0$ per $\hat{p}(b) = 1/3$ e $\hat{p}(b) = 2/3$. Quindi, per $\hat{p}(b) = 1/3$ e $\hat{p}(b) = 2/3$ il sistema *(31)* è privo di significato. Ora consideriamo l'intervallo $0 \leq \hat{p}(b) < 1/3$: risulta $\phi(\hat{p}(b)) > 0$, $\psi(\hat{p}(b))$

> 0 e $\psi(\hat{p}(b)) > \phi(\hat{p}(b))$. In quest'intervallo la seconda disuguaglianza del sistema *(31)* non è mai verificata qualunque sia $u(\cdot)$, per cui possiamo concludere che secondo la teoria della decidibilità gli individui non stimano una probabilità $\hat{p}(b)$ tale che $0 \leq \hat{p}(b) \leq 1/3$. Limitando ora l'attenzione all'intervallo $1/3 < \hat{p}(b) < 2/3$, qui $\phi(\hat{p}(b)) < 0$ e $\psi(\hat{p}(b)) < 0$, e inoltre $|\psi(\hat{p}(b))| > |\phi(\hat{p}(b))|$. La seconda disuguaglianza del sistema *(31)* è verificata per qualunque funzione $u(\cdot)$. Per quanto concerne invece la prima disuguaglianza del sistema, essa dipende dal valore di $u(\cdot)$ e $\hat{p}(b)$; ma si può osservare che prendendo $\hat{p}(b)$ molto vicino ad 1/3 (da destra), è abbastanza plausibile che qualunque sia $u(\cdot)$ la disuguaglianza è verificata. Infine nell'intervallo $2/3 < \hat{p}(b) \leq 1$, risulta $\phi(\hat{p}(b)) < 0$, $\psi(\hat{p}(b)) < 0$ e $|\psi(\hat{p}(b))| < |\phi(\hat{p}(b))|$. La prima disequazione del sistema *(31)* non è mai soddisfatta, così le persone non stimano una probabilità $2/3 < \hat{p}(b) < 1$. A questo punto possiamo concludere che le persone stimano verosimilmente $1/3 < \hat{p}(b) < 2/3$, che è una stima soggettiva coerente tale che il sistema *(31)* sia soddisfatto.

Concludendo, abbiamo trovato che il risultato dell'esperimento di Ellsberg non è paradossale se supponiamo che le persone siano ispirate dal modello di decidibilità. Secondo questo modello, gli individui sono razionali perché effettuano una stima soggettiva delle probabilità incognite che è coerente. Inoltre, questo risultato è più forte di quanto sembri. A causa della prima disuguaglianza del sistema *(31)*, nell'intervallo $1/3 < \hat{p}(b) < 2/3$, la probabilità soggettiva è verosimilmente tale che $\hat{p}(b) - 1/3 < \varepsilon$, dove ε sia un numero positivo arbitrariamente piccolo. Questo significa che se $p(r) = 1/3$ e $\hat{p}(b) \cong 1/3$, allora $\hat{p}(g) \cong 1/3$. Questo punto è veramente importante perché, se un individuo non conosce l'esatto numero di palline blu e gialle nell'urna, è naturale che egli assuma che il numero di palline blu e gialle sia più o meno uguale. Inoltre, quando le probabilità soggettive sono vicine all'equiprobabilità, abbiamo il massimo livello di incertezza, e l'entropia dei possibili risultati è massima mentre il contenuto informativo è al minimo. In una tale situazione la stima soggettiva dovrebbe essere pessimistica, e quest'ultima evidenza è compatibile con la filosofia che ha ispirato il modello *maxmin expected utility*, suggerito da Ellsberg ed assiomatizzato da Gilboa e Schmeidler [13].

Infatti, il modello *maxmin* assume che le persone si comportino in modo pessimistico, effettuando la scelta che massimizza la loro utilità immaginando che si realizzi lo scenario peggiore. Nel modello di decidibilità lo scenario peggiore – in termini di incertezza – si realizza quando tutti i risultati sono equiprobabili.

8. - Conclusioni

In conclusione, abbiamo sviluppato una teoria delle scelte che spiega il paradosso di Allais, compresa la sua versione inversa, razionalizza il comportamento nell'esperimento di Ellsberg, ed è molto semplice da applicare. Questa teoria si ispira all'idea che il processo decisionale sia governato da due fattori fondamentali: l'utilità attesa del guadagno proveniente dalla decisione e l'incertezza contenuta nella lotteria. Entrambi questi fattori sono catturati dalla funzione di decidibilità. Inoltre, poiché l'entropia è interpretata anche come misura del contenuto informativo di una distribuzione di probabilità, questa teoria riesce a spiegare come gli individui traggono beneficio dal contenuto informativo della lotteria. Infatti, *ceteris paribus*, essi scelgono la lotteria sulla quale sono più informati. In ogni caso questo lavoro non ha la presunzione di rivendicare che questa teoria sia immune da possibili esiti paradossali. Sebbene la teoria delle decisioni abbia lo scopo di formalizzare matematicamente il comportamento umano nei processi decisionali, siamo persuasi del fatto che tali processi sono molto complessi, e pertanto difficili da formalizzare.

APPENDICE

Teorema dell'utilità attesa

Teorema 3 (Von Neumman-Morgenstern). Sulla base degli assiomi 5, 6, e 7, esiste una funzione $U : \mathcal{L} \to \mathbb{R}$ tale che: *(i)* rappresenta la relazione $\succsim$ su $\mathcal{L}$, ovvero per ogni $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{L}\ \boldsymbol{x} \succsim \boldsymbol{y} \Leftrightarrow U(x) \geq U(y)$; *(ii)* ha la forma dell'utilità attesa, cioè per ogni $\boldsymbol{x} = \{x_i, p_i; i = 1, 2, \dots, n\}$, $U(\boldsymbol{x}) = \sum_{i=1}^{n} p_i u(\boldsymbol{x}_i)$; *(iii)* è unica a meno di una trasformazione affine, ovvero per ogni $\phi, \mu \in \mathbb{R}$ e $\phi > 0$, $U'(\boldsymbol{x}) = \phi U(\boldsymbol{x}) + \mu$ è un'altra funzione di utilità attesa.

Per un'ottima presentazione di questo teorema, vedi Fischburn [12].

BIBLIOGRAFIA

[1] ALLAIS M., «Le comportement de l'homme rationel devant le risque: critique des postulats et axiomes de l'ecole americaine», *Econometrica*, vol. 21, 1953, pp. 503–46.

[2] ARROW K.J., *Essays in the Theory of Risk Bearing*, Amsterdam, North Holland, 1974.

[3] ASCOMBE F.J. - AUMANN R.J., «A Definition of Subjective Probability», *Annals of Mathematical Statistics*, vol. 34, 1963, pp. 199–205.

[4] BERNARDO J.M. - SMITH F.M.A., *Bayesian Theory*, New York, John Wiley and Sons, 2000.

[5] BEWLEY T.F., «Knightian Decision Theory: Part I», Yale, Cowles Foundation, *Discussion Paper* n. 807, 1986.

[6] ——, «Knightian Decision Theory: Part II», Yale, Cowles Foundation, *Discussion Paper* n. 835, 1987.

[7] BRILLOUIN L., *Science and Information Theory*, New York, Academic Press, 1960.

[8] CAMERER C.F., *Ambiguity-aversion and Non-additive Probability: Experimental Evidence, Model and Applications*, Scuola estiva, Università di Siena, 1995.

[9] CHEW S.H. - MACCRIMMON K., «Alpha-nu Choice Theory: A Generalisation of Expected Utility Theory», University of British Columbia, *Working Paper,* n. 669, 1979.

[10] DE FINETTI B., *Teoria delle probabilità*, vol. I e II, Torino, Einaudi, 1970.

[11] ELLSBERG D., «Risk, Ambiguity, and the Savage Axioms», *Quarterly Journal of Economics*, vol. 75, 1961, pp. 643–69.

[12] FISCHBURN P.C., *Utility Theory for Decision Making*, New York, John Wiley and Sons, 1970.

[13] GILBOA I. - SCHMEIDLER D., «Maxmin Expected Utility with Non-unique Prior», *Journal of Mathematical Economics*, vol. 18, 1989, pp. 141–53.

[14] HANSEN L.P. - SARGENT T.J., «Acknowledging Misspecification in Macroeconomics», URL: www.stanford.edu/~sargent/research.html, *Discussion Paper*, gen. 2001.

[15] HERSTEIN I.N. - MILNOR J., «An Axiomatic Approach to Measurable Utility», *Econometrica*, vol. 21, 1953, pp. 291–97.

[16] KAHNEMAN D. - TVERSKY A., «Prospect Theory: An Analysis of Decision Under Risk», *Econometrica*, vol. 47, 1979, pp. 263–91.

[17] KNIGHT F.H., *Uncertainty and Profit,* Boston, Houghton Mifflin, 1921.

[18] MACHINA M.J., «Expected Utility Theory without the Indipendence Axiom», *Econometrica*, vol. 50, 1982, pp. 277–323.

[19] PRATT J., «Risk Aversion in the Small an the Large», *Econometrica*, vol. 32, 1964, pp. 122–36.

[20] SAVAGE L.J., *The Foundations of Statistics*, New York, John Wiley and Sons, 1954.

[21] SCHMEIDLER D., «Subjective Probability and Expected Utility Without Additivity», *Econometrica*, vol. 57, 1989, pp. 571–87.

[22] SHANNON C.E., «A Mathematical Theory of Telecommunication», *The Bell System Technical Journal*, vol. 27, 1948, pp. 379-423, pp. 623-56.

[23] VON NEUMANN J. - MORGENSTERN O., *Theory of Games and Economic Behavior*, Princeton (N.J), Princeton University Press, 1944.