11. Tre logiche per il ragionamento epistemico limitato

Fagin e Halpern (1985, 1988) prendono le mosse dai lavori che abbiamo esaminato nei capitoli precedenti per generalizzarli, sviluppando tecniche idonee ad affrontare i diversi aspetti connessi al problema dell'onniscienza logica. Secondo Fagin e Halpern, infatti, i tentativi precedenti si sono rivelati limitati in quanto non riconoscevano che la mancanza di onniscienza logica nei soggetti cognitivi reali dipende dall'interagire di cause differenti. Per questa ragione, ogni soluzione proposta affrontava di fatto soltanto alcuni aspetti del problema. Secondo Fagin e Halpern, la mancanza di onniscienza logica può essere ricondotta a quattro fonti differenti.

1 - *Mancanza di consapevolezza.* Può accadere che un soggetto non abbia opinioni circa la verità o la falsità di un enunciato che consegue dall'insieme delle sue credenze per la semplice ragione che nella formulazione di quell'enunciato compaiono concetti che non conosce, di modo che egli non è assolutamente consapevole della verità o falsità dell'enunciato stesso. Ad esempio, non ha molto senso chiedersi se un Bantu creda o meno nella verità dell'enunciato "i prezzi dei calcolatori stanno scendendo" qualora egli non ha la più pallida idea di cosa sia un calcolatore.

2 - *Risorse limitate.* Un soggetto razionale può ignorare certe verità logiche, o non conoscere alcune delle conseguenze logiche delle sue credenze perché non dispone del tempo o delle risorse di memoria per dedurle. Oppure perché sono formulate in maniera troppo complessa da poter essere comprese.

3 - *Ignoranza di regole di derivazione.* Spesso i ragionatori reali non conoscono o non sanno applicare alcune regole di ragionamento. Secondo l'esempio di Konolige (1984) già citato, uno studente può non essere in grado di risolvere un'equazione del tipo $x + a = b$ perché non conosce la regola in base alla quale si possono sottrarre quantità uguali su entrambi i lati dell'equazione. Le ricerche della psicologia cognitiva hanno messo in luce le difficoltà di molti soggetti nell'utilizzare la regola di contrapposizione[1].

4 - *Molteplicità dei contesti mentali.* Nel ragionamento le persone non utilizzano contemporaneamente tutte le informazioni di cui dispongono. In particolare, sembra che gli esseri umani abbiano difficoltà nell'utilizzare contemporaneamente informazioni che provengono da ambiti diversi. Sembra ragionevole pensare la memoria umana come strutturata in contesti diversi, in diversi "quadri mentali" (*frame of mind*), che difficilmente comunicano fra loro. Potrebbe accadere che, pur essendo ogni quadro mentale coerente, le informazioni in quadri mentali diversi siano fra loro incoerenti.

Fagin e Halpern propongono tre sistemi logici, con ciascuno dei quali si propongono di modellare in maniera formale alcuni aspetti del problema dell'onniscienza logica. La prima di tali logiche, la *logica della consapevolezza*, è stata sviluppata a partire dalla logica della credenza implicita ed esplicita di Levesque. Rispetto alla logica di Levesque, la differenza principale consiste nel fatto che la logica proposta da Fagin e Halpern evita l'uso di situazioni incomplete e/o inconsistenti nella semantica, introducendo in maniera esplicita nelle interpretazioni una funzione di *consapevolezza* (*awareness*). La seconda logica proposta, la *logica della consapevolezza generalizzata*, si basa su di un approccio ibrido, in base al quale nella semantica vengono combinati elementi model teoretici tradizionali con elementi di tipo sintattico. Ciò viene realizzato per mezzo di una funzione di consapevolezza generalizzata che è in grado di discriminare formule sulla base della loro struttura sintattica. La terza logica infine, la *logica del ragionamento locale*, viene proposta con l'intento di modellare le limitazioni inferenziali che dipendono dall'esistenza nella mente dei soggetti epistemici di più contesti o quadri mentali distinti, che non comunicano fra loro. La semantica di questa logica è basata sulla metafora della mente individuale vista come una "società delle menti". Queste logiche possono essere variamente combinate fra loro, in modo da fornire modelli di aspetti più complessi del ragionamento epistemico di soggetti razionali limitati. Tutte le logiche di Fagin e Halpern mantengono la distinzione di Levesque fra credenza implicita ed esplicita. La semantica di tali logiche viene ottenuta modificando le strutture di Kripke basate su mondi possibili classici, senza che siano introdotte strutture semantiche non classiche quali situazioni incoerenti o incomplete.

11.1 Una logica della consapevolezza

Alla fine del capitolo precedente abbiamo visto che Fagin e Halpern hanno individuato nella mancanza di consapevolezza di alcune formule primitive la ragione per cui nella logica di Levesque i soggetti epistemici non sono onniscienti rispetto alla conoscenza delle formule valide, e per cui (almeno in certi casi) non vale la chiusura rispetto alle implicazioni valide. Nella logica della consapevolezza Fagin e Halpern propongono di rendere esplicito mediante l'introduzione di una *funzione di consapevolezza* nella semantica il fatto che un soggetto epistemico possa non essere consapevole del significato di alcune formule primitive . Allo stesso tempo, il modello di Levesque viene esteso in maniera da rendere conto di operatori di credenza iterati, e in maniera da permettere di prendere in considerazione più soggetti epistemici. L'inserimento di funzioni di consapevolezza esplicite nella semantica rende inoltre possibile utilizzare mondi possibili classici, evitando l'uso di situazioni incomplete.

---

[1]Si vedano ad esempio (Wason e Johnson-Laird 1972) e (Johnson-Laird 1983).

Il linguaggio della logica della consapevolezza comprende gli operatori $E_1,...,E_n$ e $B_1,...,B_n$ rispettivamente per la credenza esplicita e implicita di *n* attori. Sono consentite iterazioni arbitrarie di tutti gli operatori E_i e B_j.

Una *struttura di Kripke per la consapevolezza* è una *(2n+2)*-pla $M = (W, \varphi, A_1,...,A_n, R_1,...,R_n)$. *W* è un insieme di mondi possibili. φ è un'assegnazione di valori di verità alle formule primitive del linguaggio che si comporta esattamente come la funzione interpretazione delle strutture di Kripke classiche: φ assegna uno ed un solo valore di verità a ciascuna formula primitiva rispetto a ciascun $w \in W$. I mondi sono quindi classicamente coerenti e completi. Per ogni *i*, R_i è la funzione di accessibilità fra mondi relativa all'*i*-esimo soggetto di credenza. Fagin e Halpern assumono che le relazioni di accessibilità R_i siano seriali, transitive ed euclidee. Le A_i sono le *funzioni di consapevolezza* relative a ciascun soggetto *i*. Esse associano a ciascun mondo $w \in W$ l'insieme $A_i(w)$ delle formule primitive di cui il soggetto *i* è consapevole in *w*.

Intuitivamente, lo scopo delle A_i è quello di "ritagliare" delle situazioni parziali all'interno dei mondi possibili, situazioni che corrispondono al punto di vista dei diversi attori. Così, oltre alla relazione $\models$ di soddisfacimento vengono definite le relazioni di sostegno $\models_{\mathbf{v}}^{\Psi}$ e $\models_{\mathbf{f}}^{\Psi}$ parametrizzate rispetto ad un insieme Ψ di formule primitive. Il loro significato intuitivo è quello di restringere la valutazione di una formula ad una situazione parziale in cui soltanto le formule di Ψ sono definite. Tali relazioni di sostegno, di fatto, vengono utilizzate per valutare le formule che compaiono nei contesti retti dagli operatori E_i per la credenza esplicita. L'insieme Ψ delle formule primitive rispetto al quale viene valutata un formula che compare nell'ambito di operatori E_i è determinato dalle funzioni di consapevolezza A_i loro associate. Come nella logica di Levesque, sono necessarie due relazioni di sostegno, una per la verità e una per la falsità, poiché, essendo le situazioni incomplete, non è detto che una situazione che non sostiene la verità di una formula ne sostenga la falsità, e viceversa.

La definizione formale della relazione di soddisfacimento $\models$ e delle relazioni di sostegno $\models_{\mathbf{v}}^{\Psi}$ e $\models_{\mathbf{f}}^{\Psi}$ è la seguente:

$M,w \models_{\mathbf{v}}^{\Psi} p$ dove p è una formula primitiva, sse $\varphi[w,p] = \mathbf{v}$ e $p \in \Psi$,
$M,w \models_{\mathbf{f}}^{\Psi} p$ dove p è una formula primitiva, sse $\varphi[w,p] = \mathbf{f}$ e $p \in \Psi$,
$M,w \models p$ dove p è una formula primitiva, sse $\varphi[w,p] = \mathbf{v}$;

$M,w \models_{\mathbf{v}}^{\Psi} \neg\alpha$ sse $M,w \models_{\mathbf{f}}^{\Psi} \alpha$,
$M,w \models_{\mathbf{f}}^{\Psi} \neg\alpha$ sse $M,w \models_{\mathbf{v}}^{\Psi} \alpha$,
$M,w \models \neg\alpha$ sse $M,w \not\models \alpha$;

$M,w \models_{\mathbf{v}}^{\Psi} \alpha \wedge \beta$ sse $M,w \models_{\mathbf{v}}^{\Psi} \alpha$ e $M,w \models_{\mathbf{v}}^{\Psi} \beta$,
$M,w \models_{\mathbf{f}}^{\Psi} \alpha \wedge \beta$ sse $M,w \models_{\mathbf{f}}^{\Psi} \alpha$ oppure $M,w \models_{\mathbf{f}}^{\Psi} \beta$,
$M,w \models \alpha \wedge \beta$ sse $M,w \models \alpha$ e $M,w \models \beta$;

$M,w \models_{\mathbf{v}}^{\Psi} E_i\alpha$ sse $M,w' \models_{\mathbf{v}}^{\Psi\cap A_i(w)} \alpha$ per tutti i w' tali che $R_i(w,w')$,
$M,w \models_{\mathbf{f}}^{\Psi} E_i\alpha$ sse $M,w' \models_{\mathbf{f}}^{\Psi\cap A_i(w)} \alpha$ per qualche w' tale che $R_i(w,w')$,
$M,w \models E_i\alpha$ sse $M,w \models_{\mathbf{v}}^{\Phi} E_i\alpha$ dove Φ è l'insieme di tutte le formule primitive;

$M,w \models_{\mathbf{v}}^{\Psi} B_i\alpha$ sse $M,w' \models_{\mathbf{v}}^{\Psi} \alpha$ per tutti i w' tali che $R_i(w,w')$,
$M,w \models_{\mathbf{f}}^{\Psi} B_i\alpha$ sse $M,w' \models_{\mathbf{f}}^{\Psi} \alpha$ per qualche w' tale che $R_i(w,w')$,
$M,w \models B_i\alpha$ sse $M,w' \models_{\mathbf{v}}^{\Phi} \alpha$ per tutti i w' tali che $R_i(w,w')$.

Anche qui, diremo che una formula α è valida (in simboli, $\models\alpha$) se e soltanto se vale $M,w \models \alpha$ in tutte le interpretazioni *M* e per tutti i mondi *w* in *M*.

Si noti che affinché un mondo *w* sostenga la verità di una formula del tipo $\neg\alpha$ rispetto ad un insieme Ψ di formule primitive, *w* deve sostenere esplicitamente la falsità di α rispetto a Ψ (non basta che non sostenga la verità di α). Analogamente, perché un mondo *w* sostenga la falsità di una formula $\alpha \wedge \beta$ rispetto a Ψ, la falsità di α o di β deve essere sostenuta esplicitamente.

Abbiamo detto che le relazioni di sostegno $\models_{\mathbf{v}}^{\Psi}$ e $\models_{\mathbf{f}}^{\Psi}$ entrano in gioco per la valutazione delle formule in cui compaiono gli operatori di credenza esplicita E_i. Infatti, nella definizione della relazione di soddisfacimento $\models$ le relazioni di sostegno compaiono solo per le formule di tipo $E_i\alpha$. Dalla definizione consegue che una formula $E_i\alpha$ è vera rispetto a un mondo w se e soltanto se $w' \models_{\mathbf{v}}^{A_i(w)} \alpha$ per tutti i w' tali che $R_i(w, w')$. Nella definizione delle relazioni di sostegno per le formule $E_i\alpha$ l'insieme Ψ viene ulteriormente ristretto a $\Psi \cap A_i(w)$, se ne considera cioè l'intersezione con l'insieme delle formule primitive di cui i è consapevole rispetto al mondo w. Questo risolve il problema delle modalità epistemiche iterate. Per formule del tipo $E_i E_j \alpha$ si ha che:

$$M,w \models E_i E_j \alpha \text{ sse } M,w' \models_{\mathbf{v}}^{A_i(w)} E_j \alpha \text{ per tutti i } w' \text{ tali che } R_i(w,w')$$

e

$$M,w' \models_{\mathbf{v}}^{A_i(w)} E_j \alpha \text{ sse } M,w'' \models_{\mathbf{v}}^{A_i(w) \cap A_j(w')} \alpha \text{ per tutti i } w'' \text{ tali che } R_j(w',w'').$$

Per la valutazione della verità di formule rette dagli operatori di credenza implicita B_i le funzioni di consapevolezza non vengono prese in considerazione. Di conseguenza, nella definizione delle relazioni di sostegno per formule del tipo $B_i\alpha$ l'insieme Ψ non viene ulteriormente ristretto. La credenza implicita soddisfa gli assiomi di **S5** debole. Quindi, ogni soggetto di credenza crede tutte le formule valide e tutte le conseguenze logiche delle sue credenze.

Possono essere dimostrate le seguenti proposizioni:

1) $\models$ è completa, vale a dire, per ogni M, w, α, vale $M,w \models \alpha$ oppure $M,w \models \neg\alpha$.
2.a) Se $\Psi \subseteq \Psi'$ e se $M,w \models_{\mathbf{v}}^{\Psi} \alpha$, allora $M,w \models_{\mathbf{v}}^{\Psi'} \alpha$
2.b) Se $\Psi \subseteq \Psi'$ e se $M,w \models_{\mathbf{f}}^{\Psi} \alpha$, allora $M,w \models_{\mathbf{f}}^{\Psi'} \alpha$
3.a) Per ogni insieme Ψ di proposizioni primitive, se $M,w \models_{\mathbf{v}}^{\Psi} \alpha$, allora $M,w \models \alpha$.
3.b) Per ogni insieme Ψ di proposizioni primitive, se $M,w \models_{\mathbf{f}}^{\Psi} \alpha$, allora $M,w \models \neg\alpha$.
4) $\models E_i\alpha \rightarrow B_i\alpha$.

La 4), che è una conseguenza immediata della 3.a), comporta che, come nella logica di Levesque, la credenza esplicita implichi la credenza implicita. Come in Levesque, non tutte le formule valide devono necessariamente essere credute esplicitamente. Ad esempio, $\neg E_i(\alpha \vee \neg \alpha)$ può essere soddisfatta (nel caso che α sia atomica, questo accade quando il soggetto i non è consapevole di α nel mondo preso in considerazione). Inoltre, anche qui la credenza esplicita non è chiusa rispetto all'implicazione valida. Così, $E_i p \wedge \neg E_i(p \wedge (q \vee \neg q))$ è una formula soddisfacibile. Tuttavia, il fatto di non utilizzare situazioni incoerenti, comporta le seguenti differenze rispetto alla logica di Levesque: i) l'insieme delle credenze esplicite di un agente è chiuso rispetto all'implicazione; ii) un agente non può avere credenze inconsistenti, ad esempio una formula come $E_i(p \wedge \neg p)$ non è soddisfacibile.

Per quanto concerne le credenze iterate, vale ad esempio quanto segue: $\models E_i B_i \alpha \leftrightarrow E_i\alpha$, vale a dire: un agente i crede esplicitamente di credere implicitamente α se e soltanto se egli crede esplicitamente α.

Per quanto concerne i rapporti fra credenza e consapevolezza, si noti che, per ogni formula primitiva p, vale che $M,w \models E_i(p \vee \neg p)$ sse $p \in A_i(w)$. E' così possibile introdurre nel linguaggio una classe di operatori di consapevolezza A_i nel modo seguente. Per ogni formula primitiva p, si abbia

$$A_i p =_{def} E_i(p \vee \neg p).$$

Per ogni formula α non primitiva, $A_i\alpha$ equivale per definizione alla congiunzione delle $A_i p$ per tutte le formule primitive p che compaiono in α. Si può dimostrare che, analogamente a quanto accade nella logica di Levesque, si ha che $\models A_i\alpha \rightarrow E_i\alpha$ per ogni formula valida α. Questo consente di mettere in luce alcune interessanti relazioni fra credenza implicita ed esplicita e consapevolezza. Ad esempio, si può dimostrare che:

$$\models E_i(p \vee q) \leftrightarrow (A_i p \wedge B_i p) \vee (A_i q \wedge B_i q) \vee (A_i p \wedge A_i q \wedge B_i(p \vee q))$$

e che

$$\models E_i E_j p \leftrightarrow (A_i p \wedge B_i(A_j p \wedge B_j p)).$$

In entrambi i casi, le occorrenze dell'operatore di credenza esplicita vengono eliminate a favore degli operatori di credenza implicita e di consapevolezza. Questo fatto può essere generalizzato. Vale infatti che:

Proposizione: Data una formula φ, si può sempre determinare in modo effettivo una formula φ^* tale che $\models \varphi \leftrightarrow \varphi^*$ e tale che gli operatori E_i compaiano in φ^* solo nel contesto $E_i(p \vee \neg p)$, dove p è una proposizione primitiva. (Fagin e Halpern 1988)

Fagin e Halpern assiomatizzano la logica della consapevolezza aggiungendo all'apparato deduttivo di **S5** debole per la credenza implicita il seguente schema di assiomi per la credenza esplicita:

$$(*)\ \ \varphi \leftrightarrow \varphi^*\ ,$$

dove φ^* è la formula citata nella proposizione precedente. In questo modo, è possibile dimostrare che:

Teorema: gli assiomi di $\mathbf{S5_n}$ debole (formulati per gli operatori di credenza implicita B_i) e l'assioma (*) forniscono un'assiomatizzazione corretta e completa della logica della consapevolezza generalizzata. (Fagin e Halpern 1988)

Tale assiomatizzazione risulta tuttavia molto poco intuitiva. Fagin e Halpern auspicano che in futuro si possa individuare un insieme di assiomi più "naturale".

11.2 Una logica della consapevolezza generalizzata

La seconda logica proposta da Fagin e Halpern si basa sulla generalizzazione delle funzioni di consapevolezza A_i, in maniera tale che esse possano assumere come valori insiemi di formule qualsiasi anziché esclusivamente formule primitive. Ciò comporta l'introduzione nel linguaggio di un insieme di operatori di consapevolezza A_i, uno associato a ciascun soggetto epistemico, a fianco degli usuali operatori E_i e B_i per la credenza esplicita ed implicita. In generale, una formula del tipo $A_i\alpha$ può essere letta intuitivamente come: "il soggetto i è consapevole di α", "i può figurarsi cosa significa che α sia vera". Ciò consente di incorporare in questa logica aspetti tipici del trattamento sintattico della credenza.

Una *struttura di Kripke per la consapevolezza generalizzata* è una *(2n+2)*-pla $M = (W, \varphi, A_1,...,A_n, R_1,...,R_n)$, dove, come di consueto, W è l'insieme dei mondi possibili, φ è la funzione interpretazione che assegna un valore di verità per ogni mondo $w \in W$ alle formule primitive del linguaggio, e le varie R_i sono le relazioni di accessibilità relative ad ogni soggetto epistemico. Fagin e Halpern assumono che tali relazioni siano seriali transitive ed euclidee. Come nella logica del paragrafo precedente, le $A_i(w)$ sono le funzioni di consapevolezza associate a ciascun attore. Come abbiamo detto, a differenza di quanto accadeva nella logica descritta nel paragrafo precedente, l'insieme di formule che tali funzioni associano ai soggetti epistemici nei vari mondi possibili è composto da formule arbitrarie anziché esclusivamente da formule primitive. Nel caso generale Fagin e Halpern non pongono alcuna restrizione sul comportamento delle A_i.

Per la logica della consapevolezza generalizzata non sono necessarie relazioni di sostegno, ma è sufficiente una relazione di verità classica a due valori definita come segue:

$M,w \models p$ (per p appartenente all'insieme delle formule primitive) sse $\varphi[w,p]=\mathbf{v}$
$M,w \models \alpha \wedge \beta$ sse $M,w \models \alpha$ e $M,w \models \beta$
$M,w \models \neg\alpha$ sse $M,w \not\models \alpha$
$M,w \models A_i\alpha$ sse $\alpha \in A_i(w)$
$M,w \models B_i\alpha$ sse $M,w' \models \alpha$ in tutti i w' t.c. $R_i(w,w')$.
$M,w \models E_i\alpha$ sse $\alpha \in A_i(w)$ e $M,w' \models \alpha$ in tutti i w' t.c. $R_i(w,w')$.

Anche qui, gli operatori di credenza implicita B_i si comportano come in **S5** debole. Vale a dire, gli assiomi di **S5** debole, formulati per gli operatori di credenza implicita, sono validi nella logica della consapevolezza generalizzata. Per quanto concerne la credenza esplicita, come nella logica precedente, un soggetto epistemico non è tenuto a credere tutte le formule valide. Ad esempio, $\neg E_i(p \wedge \neg p)$ è soddisfacibile. Inoltre, in questo caso, la credenza esplicita non è neppure chiusa rispetto all'implicazione. Una formula come $E_ip \wedge E_i(p \rightarrow q) \wedge \neg E_iq$ può essere soddisfatta, in quanto il soggetto i può essere, in linea di principio, consapevole di p e di $p \rightarrow q$, ma non essere allo stesso tempo consapevole di q. La credenza esplicita, se relativizzata rispetto alla consapevolezza, presenta proprietà analoghe alla credenza implicita. Ad esempio, in corrispondenza dell'assioma di distribuzione, per la credenza esplicita vale il seguente schema di formule:

$$E_i\alpha \wedge E_i(\alpha \rightarrow \beta) \wedge A_i\beta \rightarrow E_i\beta,$$

vale a dire, se sono credute esplicitamente α e $\alpha \rightarrow \beta$, allora è creduta esplicitamente β, *a patto di essere consapevoli di* β. In corrispondenza della regola di necessitazione, per la credenza esplicita è derivabile la seguente regola:

$$\text{da } \alpha \text{ segue } A_i\alpha \rightarrow E_i\alpha.$$

In base alla definizione di $\models$, un agente *i* crede esplicitamente una formula α se e soltanto se *i* crede implicitamente α e al tempo stesso *i* è consapevole di α. In altri termini, nella logica della consapevolezza generalizzata il rapporto fra credenza implicita, credenza esplicita e consapevolezza è schematizzato dalla fig. 11.1 (da Konolige 1986b). Le formule credute esplicitamente da un soggetto epistemico sono l'intersezione delle formule credute implicitamente con le formule di cui il soggetto è consapevole[2]. Vale quindi il seguente schema di formule:

$$(^{**})\ E_i\alpha \leftrightarrow B_i\alpha \wedge A_i\alpha.$$

In questa logica (**) caratterizza completamente il comportamento degli operatori di credenza esplicita E_i. Fagin e Halpern utilizzano quindi (**) come assioma per la logica della consapevolezza generalizzata. E' possibile dimostrare che:

Teorema: gli assiomi di $\mathbf{S5_n}$ debole (formulati per gli operatori di credenza implicita B_i) e l'assioma (**) forniscono un'assiomatizzazione corretta e completa della logica della consapevolezza generalizzata. (Fagin e Halpern 1988)

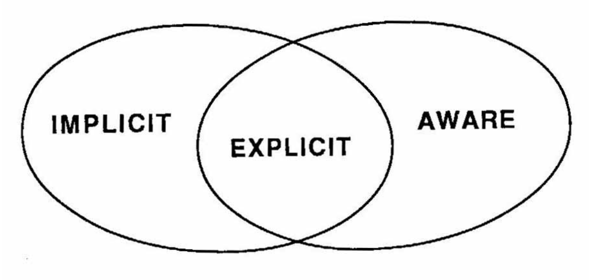

fig. 11.1

Sin qui, si è assunto che gli insiemi di formule associati alle funzioni di consapevolezza fossero del tutto arbitrari. Ad esempio, un soggetto *i* in un mondo *w* potrebbe essere consapevole di una certa formula α senza essere al contempo consapevole della sua negazione, potrebbe cioè accadere che $\alpha \in A_i(w)$, senza che $\neg\alpha \in A_i(w)$. Oppure, si potrebbe avere che $\alpha \wedge \beta \in A_i(w)$, senza che $\beta \wedge \alpha \in A_i(w)$ (questo, fanno notare Fagin e Halpern, può essere effettivamente il caso in molti esempi di prestazioni umane o di applicazioni di tipo computazionale, in cui l'ordine di presentazione dei congiunti in una congiunzione può essere rilevante). Volendo modellare tipi specifici di ragionamento epistemico limitato, restrizioni sulle funzioni di consapevolezza diventano necessarie. Ulteriori assiomi che regolino il comportamento degli operatori A_i possono essere aggiunti qualora si intendano porre restrizioni sull'insieme delle formule di cui un agente può essere consapevole, precisando in questo modo il comportamento e il significato della relazione di consapevolezza. Alcuni esempi tipici elencati da Fagin e Halpern sono i seguenti.

1) Si può rendere irrilevante l'ordine dei congiunti ponendo $\alpha \wedge \beta \in A_i(w)$ sse $\beta \wedge \alpha \in A_i(w)$. A livello di teoria della dimostrazione, ciò corrisponde all'assioma:

---

[2]Si noti che nella logica della consapevolezza esposta nel paragrafo precedente, questo vale solo per le formule primitive, mentre non vale in generale per le formule complesse. Abbiamo visto infatti che nella logica della consapevolezza del paragrafo 11.1, dato un soggetto epistemico *i* e una formula α non primitiva, $A_i\alpha$ equivale alla congiunzione delle formule $A_i\, p_k$ per tutte le formule p_k primitive che compaiono in α. Quindi, si ha ad esempio che:

$$A_i(\alpha \vee \beta) \leftrightarrow A_i\alpha \wedge A_i\beta.$$

D'altra parte, si può constatare che $E_i\alpha \rightarrow E_i(\alpha \vee \beta)$ risulta valida, a prescindere dal fatto che *i* sia consapevole di β. Quindi, può accadere che si abbia $E_i(\alpha \vee \beta)$ senza che si abbia $A_i(\alpha \vee \beta)$.

$$A_i(\alpha \wedge \beta) \leftrightarrow A_i(\beta \wedge \alpha).$$

Analogamente, si può imporre che un soggetto sia consapevole di una formula se e soltanto se è consapevole della sua negazione, vale a dire $\alpha \in A_i(w)$ sse $\neg\alpha \in A_i(w)$. Questo equivale all'assioma:

$$A_i\alpha \leftrightarrow A_i\neg\alpha.$$

Nelle teorie che soddisfano quest'ultimo vincolo vale che: $E_i\alpha \leftrightarrow E_i\neg\neg\alpha$.

2) La consapevolezza può essere chiusa rispetto alle sottoformule: se $\alpha \in A_i(w)$, e se β è una sottoformula di α, allora $\beta \in A_i(w)$. Questo può essere espresso mediante i seguenti assiomi:

$$A_i(\neg\alpha) \rightarrow A_i\alpha$$
$$A_i(\alpha \wedge \beta) \rightarrow A_i\alpha \wedge A_i\beta \text{ (**)}$$
$$A_i(E_j\alpha) \rightarrow A_i\alpha$$
$$A_i(B_j\alpha) \rightarrow A_i\alpha$$
$$A_i(A_j\alpha) \rightarrow A_i\alpha.$$

Se la consapevolezza è chiusa rispetto alle sottoformule, benché i soggetti epistemici non credano esplicitamente tutte le formule valide, tuttavia le loro credenze risultano chiuse rispetto all'implicazione, cioè lo schema $E_i\alpha \wedge E_i(\alpha \rightarrow \beta) \rightarrow E_i\beta$ risulta valido. L'assunzione della chiusura della consapevolezza rispetto alle sottoformule non è quindi adeguata per modellare il comportamento di soggetti epistemici con risorse computazionali limitate. Inoltre, questa assiomatizzazione corrisponde a un sistema che per computare la verità di una formula deve necessariamente computare la verità di tutte le sue sottoformule. E' tuttavia sensato immaginare che un soggetto epistemico possa credere esplicitamente $\alpha \vee \neg\alpha$ senza aver computato il valore di verità di α. Vi sono quindi casi in cui sembra sensato adottare assunzioni più deboli, ad esempio limitandosi a richiedere che, qualora $\alpha \wedge \beta \in A_i(w)$, allora si abbia $\alpha, \beta \in A_i(w)$. Sintatticamente, ciò corrisponde ad adottare il solo assioma (**) dei precedenti.

3) La consapevolezza degli agenti epistemici può essere circoscritta ad un certo sottoinsieme Φ dell'insieme delle formule primitive del linguaggio; in questo caso, $A_i(w)$ deve consistere di tutte e sole quelle formule in cui compaiano come formule primitive soltanto formule di Φ. Sintatticamente, questo si ottiene sostituendo con bicondizionali i segni di implicazione degli assiomi formulati al punto precedente. In questo modo si ottiene una logica per molti versi analoga alla logica della consapevolezza descritta nel paragrafo precedente, ma con alcune differenze. Ad esempio, nella logica della consapevolezza del paragrafo precedente la formula $E_i\alpha \rightarrow E_i(\alpha \vee \beta)$ è valida, a prescindere dal fatto che i sia consapevole o meno di β, mentre essa non è valida nella logica della consapevolezza generalizzata.

4) Si può fare in modo che un agente epistemico "non conosca" certi altri agenti, cioè si può fare in modo che un agente i non sia consapevole di alcuna formula in cui viene menzionato un altro agente j.

5) Un soggetto epistemico può essere consapevole di ciò di cui è consapevole, si può fare in modo cioè che valga che, ogni qual volta $\alpha \in A_i(w)$, allora si ha $A_i\alpha \in A_i(w)$. Sintatticamente, questo corrisponde all'assioma: $A_i\alpha \rightarrow A_i A_i \alpha$.

6) Un soggetto epistemico può sapere di quali formule è consapevole. Nell'interpretazione, questo comporta che se $R_i(w,w')$, allora $A_i(w) = A_i(w')$. Ciò è espresso dagli assiomi:

$$A_i\alpha \rightarrow B_i A_i\alpha$$
$$\neg A_i\alpha \rightarrow B_i\neg A_i\alpha.$$

Aggiungendo questi assiomi, diventano valide le formule corrispondenti agli assiomi di introspezione (positiva e negativa) formulate mediante gli operatori di credenza esplicita, relativizzati rispetto alla consapevolezza, cioè:

$$E_i\alpha \wedge A_i E_i\alpha \rightarrow E_i E_i\alpha$$

e

$$\neg E_i\alpha \wedge A_i(\neg E_i\alpha) \rightarrow E_i\neg E_i\alpha.$$

7) Gli elementi di $A_i(w)$ possono essere quelle formule che l'agente i può dedurre in un determinato lasso di tempo, o che non superano un certo grado di complessità strutturale.

11.3 Una logica per il ragionamento locale

La logica presentata in questa sezione è stata proposta da Fagin e Halpern per sviluppare un modello formale dell'intuizione secondo cui uno dei motivi della mancanza di onniscienza logica dipende dal fatto che gli esseri umani (e, in generale, i soggetti cognitivi limitati) non riescono a focalizzare contemporaneamente la loro attenzione su tutte le informazioni di cui dispongono, per cui le deduzioni effettivamente effettuate dipendono in realtà soltanto da una parte delle conoscenze (o delle credenze) loro disponibili. In questa prospettiva, le credenze degli attori epistemici vengono considerate come organizzate in una serie di *cornici mentali* (*frame of mind*) separate e non comunicanti fra loro, e i soggetti epistemici vengono modellati come una sorta di "società delle menti"[3]. Soluzioni simili al problema dell'onniscienza logica sono state proposte (in maniera più o meno formalizzata) da diversi autori in ambito filosofico, fra cui ad esempio Lewis (1982), Stalnaker (1984) e Rescher e Brandom (1980). Poiché i vari quadri mentali possono contenere informazioni fra loro in conflitto, questo tipo di logica è particolarmente adatto a spiegare il fatto che i soggetti epistemici possono avere delle credenze (globalmente) incoerenti. David Lewis cita, ad esempio, il caso seguente:

> Ero solito pensare che Nassau Street corresse grosso modo da est a ovest, che la vicina ferrovia corresse grosso modo da nord a sud, e che le due fossero grosso modo parallele. [...] Così, ogni enunciato di una terna inconsistente era vero secondo le mie credenze, senza che per questo ogni cosa risultasse vera secondo le mie credenze. E a proposito della congiunzione, palesemente inconsistente, dei tre enunciati? Io dico che essa non era vera secondo le mie credenze. Il mio sistema di credenze era spezzato in frammenti sovrapposti. Frammenti diversi entravano in azione in situazioni differenti, e l'intero sistema di credenze non si era mai manifestato nella sua completezza. Il primo e il secondo enunciato della terna inconsistente appartenevano a frammenti diversi - erano veri in base a frammenti diversi. Il terzo apparteneva ad entrambi. La congiunzione inconsistente di tutti e tre non apparteneva, non era in nessun modo implicata, né risultava vera in base ad alcun frammento. Perciò essa non era vera secondo il mio sistema di credenze preso nel suo insieme. (Lewis 1982, p. 436).

Incorporando questa intuizione, la logica di Fagin e Halpern riesce ad ammettere credenze contraddittorie senza inserire nelle interpretazioni situazioni inconsistenti. Mentre nelle classiche strutture a mondi possibili ad ogni soggetto epistemico viene associato un insieme di mondi (i mondi compatibili con ciò che egli crede), nelle interpretazioni per la logica del ragionamento locale ad ogni soggetto vengono associati più insiemi di mondi; ciascun insieme corrisponde a un "grappolo di credenze", ad un quadro mentale.

Dal punto di vista formale, una *struttura di Kripke per il ragionamento locale* è una $n+2$-pla $M = (W, \varphi, C_1,...,C_n)$, dove, come di consueto, W è un insieme di mondi possibili, e φ è una funzione interpretazione che assegna un valore di verità alle formule primitive del linguaggio per ogni mondo $w \in W$. Per ogni $w \in W$, $C_i(w)$ è un insieme non vuoto di sottoinsiemi non vuoti di W. Intuitivamente, se $C_i(w) = \{T_1,...,T_k\}$, allora gli insiemi $T_1,...,T_k$ sono gli insiemi di mondi possibili compatibili con le credenze comprese nei vari quadri mentali relativi al soggetto epistemico i nel mondo possibile w. Il soggetto i nel mondo w, a seconda del quadro mentale in cui si trova, ossia del contesto su cui focalizza la propria attenzione, crede talvolta che l'insieme di stati possibili sia T_1, talvolta che sia T_2, e così via. Agli operatori di credenza esplicita E_i viene assegnato in questa logica il seguente significato: una formula del tipo $E_i\alpha$ viene interpretata come "il soggetto i crede α in *qualche* quadro mentale". Questa nozione di credenza esplicita è quindi abbastanza diversa da quella discussa nei paragrafi precedenti. Fagin e Halpern chiamano *credenza locale* questo tipo di credenza. Per quanto riguarda la *credenza implicita*, essa viene definita nel modo seguente: un soggetto epistemico crede implicitamente α se α segue dal fatto di mettere insieme tutta la conoscenza presente nei vari quadri mentali. Cioè, una formula del tipo $B_i\alpha$ viene valutata in un mondo w valutando α rispetto all'insieme di mondi che si ottiene prendendo l'intersezione di tutti i quadri mentali T_j associati al soggetto i in w. In questo modo, ciò che è creduto da i nei vari quadri mentali, è creduto, a maggior ragione, implicitamente da i. Quindi, anche in questo caso vale che $E_i\alpha \rightarrow B_i\alpha$.

La relazione $\models$ per la logica del ragionamento locale viene definita nel modo seguente:

$M,w \models p$ (per p appartenente all'insieme delle formule primitive) sse $\varphi[w,p]=\mathbf{v}$
$M,w \models \alpha \wedge \beta$ sse $M,w \models \alpha$ e $M,w \models \beta$
$M,w \models \neg\alpha$ sse $M,w \not\models \alpha$
$M,w \models E_i\alpha$ sse esiste un $T \in C_i(w)$ tale che $M,w' \models \alpha$ per tutti i $w' \in T$.

---

[3]Fagin e Halpern riprendono la metafora della società delle menti da Minsky (1986), che la introduce tuttavia in un contesto diverso.

$$M,w \models B_i\alpha \text{ sse } M,w' \models \alpha \text{ per tutti i } w' \in \bigcap_{T\in C_i(w)} .$$

In base a questa definizione, la credenza esplicita non è chiusa rispetto all'implicazione; è infatti possibile, ad esempio, soddisfare formule del tipo: $E_i\alpha \wedge E_i(\alpha \rightarrow \beta) \wedge \neg E_i\beta$. Questo tuttavia non ha nulla a che fare con la consapevolezza: ciò accade perché α e $\alpha \rightarrow \beta$ sono creduti in quadri mentali diversi, e il soggetto epistemico i non è quindi in grado di "mettere assieme" queste informazioni per dedurre la formula β. Contrariamente a quanto accadeva nelle logiche della consapevolezza, sono invece credute esplicitamente tutte le formule valide. Inoltre, la credenza esplicita è chiusa rispetto all'implicazione valida: se $\alpha \rightarrow \beta$ è valida, lo è anche $E_i\alpha \rightarrow E_i\beta$. Questo perché tutti i mondi w appartenenti ai vari T_j sono corretti e completi. In questa logica, i soggetti epistemici sono quindi ragionatori perfetti all'interno di ciascun quadro mentale. Poiché consapevolezza e ragionamento locale catturano aspetti differenti della mancanza di onniscienza logica di soggetti razionali limitati, è possibile definire logiche ibride che combinino nella semantica il meccanismo dei quadri mentali con l'adozione di funzioni di consapevolezza.

I soggetti epistemici possono avere credenze incoerenti: una formula del tipo $E_i\alpha \wedge E_i\neg\alpha$ può essere soddisfatta. Ciò accade se α è creduta in un quadro mentale, e $\neg\alpha$ è creduta in un quadro mentale differente. Tuttavia, poiché i mondi sono coerenti, una contraddizione non può essere creduta all'interno di un singolo quadro mentale, cioè non possono essere soddisfatte formule del tipo $E_i(\alpha \wedge \neg\alpha)$. Si noti che, se in due quadri mentali diversi valgono due formule che si contraddicono a vicenda, allora l'intersezione dei vari T_j è vuota. In tal caso viene soddisfatta la formula $B_i(p \wedge \neg p)$, e di conseguenza il soggetto epistemico i crede implicitamente qualunque formula[4]. Cioè, nella logica per il ragionamento locale è valido il seguente schema: $E_i\alpha \wedge E_i\neg\alpha \rightarrow B_i(p \wedge \neg p)$. Questo significa che in questa logica non vale lo schema di assiomi P per la credenza implicita[5]. A differenza delle logiche dei paragrafi precedenti, quindi, gli operatori B_i non si comportano secondo gli assiomi di **S5** debole. E' facile vedere che non valgono neppure gli equivalenti degli schemi di assiomi A4 ed A5. Per quanto riguarda l'assioma di introspezione positiva (A4), esso non è valido in quanto in questo tipo di logica non vale l'equivalente della transitività delle relazioni di accessibilità R_i. Sia dato un mondo $w \in W$ ed un soggetto epistemico i. Supponiamo che w' sia un membro dell'intersezione dei T_j appartenenti a $C_i(w)$. Per come sono definite le interpretazioni, non è detto che tutti i w'' nell'intersezione degli elementi di $C_i(w')$ stiano a loro volta nell'intersezione degli elementi di $C_i(w)$. E' quindi possibile soddisfare $B_i\alpha \wedge \neg B_i B_i\alpha$. Analogamente, per quanto riguarda l'introspezione negativa (schema di assiomi A5), non vale l'equivalente del fatto che le relazioni di accessibilità sono euclidee. Sia dato un mondo $w \in W$ ed un soggetto epistemico i. Supponiamo che w' e w'' siano membri dell'intersezione degli elementi di $C_i(w)$. Anche in questo caso, non è detto che w'' stia nell'intersezione degli elementi di $C_i(w')$, e viceversa. Si può invece facilmente constatare che nella logica del ragionamento locale valgono gli assiomi A1 (assiomi del calcolo proposizionale) e A2 (assioma distributivo, formulato per gli operatori di cedenza implicita B_i) Sono soddisfatte inoltre le regole di inferenza R1 (*Modus ponens*) e R2 (regola di necessitazione per la credenza implicita). Per l'assiomatizzazione della logica del ragionamento locale prendiamo inoltre in considerazione i seguenti assiomi:

A8. $\neg E_i(\alpha \wedge \neg\alpha)$
A9. $E_i\alpha \rightarrow B_i\alpha$.

Di entrambi abbiamo constatato la validità in modo intuitivo. Abbiamo inoltre visto che i soggetti epistemici credono esplicitamente ogni formula valida, e che la credenza esplicita è chiusa rispetto all'implicazione valida. Questo giustifica le due seguenti regole di inferenza:

R3. da α si inferisce $E_i\alpha$ (necessitazione per la credenza implicita)
R4. da $\alpha \rightarrow \beta$ si inferisce $E_i\alpha \rightarrow E_i\beta$.

Si noti che A9 e R3 rendono la regola R2 per la conoscenza implicita ridondante.

Si può quindi dimostrare che:

Teorema: Il sistema formale che consiste degli assiomi A1, A2, A8 e A9, e delle regole di inferenza R1, R3 e R4 costituisce un'assiomatizzazione corretta e completa per la logica del ragionamento locale. (Fagin e Halpern 1988)

[4]Che valga $\neg E_i(p \wedge \neg p)$ è garantito invece dal fatto che i membri dei $C_i(w)$ non possono essere insiemi vuoti.

[5]Qui e nel seguito del paragrafo facciamo riferimento alla denominazione degli assiomi e delle regole data nel par. 8.2.

Ulteriori condizioni possono essere imposte sulle interpretazioni perché valgano gli equivalenti di A3, A4 e A5. Ad esempio, affinché sia valida $\neg B_i(p \wedge \neg p)$ bisogna fare in modo che, per ogni $w \in W$, l'intersezione dei $C_i(w)$ non sia mai vuota. Si noti tuttavia che questo impedisce che un soggetto possa avere credenza esplicite inconsistenti: $E_i p \wedge E_i \neg p$ non può più essere soddisfatta. Se si impone che, per ogni w, $w' \in W$, si abbia che, se $w' \in T \in C_i(w)$, allora $T \in C_i(w')$ (cioè, intuitivamente, se si richiede che in ogni quadro mentale un agente consideri possibile che egli si trovi in quel quadro mentale), allora valgono gli assiomi di introspezione positiva sia per la credenza implicita sia per quella esplicita: $E_i\alpha \to E_i E_i\alpha$ e $B_i\alpha \to B_i B_i\alpha$. Aggiungendo la condizione che, per ogni $T \in C_i(w)$ e per ogni $w' \in T$, si abbia $C_i(w') \subseteq C_i(w)$, valgono gli assiomi di introspezione negativa per la credenza esplicita ed implicita: $\neg E_i\alpha \to E_i \neg E_i\alpha$ e $\neg B_i\alpha \to B_i\neg B_i\alpha$.

Un caso particolare che Fagin e Halpern prendono in considerazione è quello in cui in ciascun quadro mentale un agente rifiuta di credere di potersi trovare in altri quadri mentali. Nella semantica, questo equivale ad imporre che, se $w' \in T \in C_i(w)$, allora $C_i(w')$ è uguale al singleton $\{T\}$. Fagin e Halpern chiamano un'interpretazione di questo genere una *Struttura di Kripke per agenti di vedute ristrette*. Un agente di vedute ristrette crede di essere consistente anche se, di fatto, non lo è. Infatti, in queste strutture semantiche $E_i(\neg(E_i\alpha \wedge E_i\neg\alpha))$ è valida, sebbene $E_i\alpha \wedge E_i\neg\alpha$ possa essere soddisfatta. Inoltre, poiché i soggetti epistemici sono ragionatori perfetti in ciascun quadro mentale, un ragionatore epistemico di vedute ristrette crede di essere un ragionatore perfetto *tout court*. Infatti il seguente schema risulta valido: $E_i(E_i\alpha \wedge E_i(\alpha \to \beta) \to E_i\beta)$.

Per costruire un modello del ragionamento locale per la conoscenza, bisogna imporre che il mondo reale sia compatibile con ciascun quadro mentale. Così, per ciascun $w \in W$, deve valere che w sia membro di ogni T_j appartenente a $C_i(w)$. Questo fa sì che, per ogni i e per ogni w, l'intersezione di tutti i T_j appartenenti a $C_i(w)$ non possa mai essere vuota. Quindi non è possibile che i soggetti epistemici abbiano credenze contraddittorie, neppure in quadri mentali differenti.

Fagin e Halpern evidenziano l'analogia fra un soggetto epistemico nelle logiche per il ragionamento locale e una comunità di soggetti epistemici (abbiamo visto che nelle logiche del ragionamento locale un soggetto epistemico può essere visto come una "società delle menti"). Sarebbe quindi possibile definire operatori modali che esprimano le relazioni fra i vari quadri mentali, come ad esempio un operatore che esprima che una data formula è *conoscenza comune* (cfr. par. 8.2) fra tutti i *frame of mind*. In (Fagin e Halpern 1985) era stato introdotto un operatore modale per esprimere che una data formula era creduta in tutti i quadri mentali.

E' interessante notare come la logica del ragionamento locale di Fagin e Halpern possa essere ottenuta come caso particolare della logica delle strutture di credenza descritta nel paragrafo 8.3, ponendo opportune restrizioni sui modelli minimali utilizzati. Si consideri il sistema formale **EMNP**, che si ottiene aggiungendo al sistema formale **E** gli schemi di assiomi M,N e P (cfr. par. 8.3). L'apparato deduttivo di **EMNP** è quindi il seguente[6]:

assiomi:

- tutti gli assiomi del calcolo proposizionale;

M. $E(\alpha \wedge \beta) \to (E\alpha \wedge E\beta)$

N. $E(\alpha \vee \neg\alpha)$

P. $\neg E(\alpha \wedge \neg\alpha)$

regole di inferenza :

MP. da α e da $\alpha \to \beta$ si inferisce β (*modus ponens*)

RE. da $\alpha \leftrightarrow \beta$ si inferisce $E\alpha \leftrightarrow E\beta$

Semanticamente, **EMNP** risulta corretto e completo rispetto alle strutture di credenza chiuse rispetto ai soprainsiemi, contenenti l'unità e tali che, per ogni $w \in W$, $\varnothing \notin N(w)$.

Limitandoci per semplicità al caso che prevede un solo soggetto epistemico, l'apparato deduttivo per la parte della logica per la conoscenza locale di Fagin e Halpern che concerne la credenza esplicita è il seguente:

assiomi:

- gli assiomi del calcolo proposizionale

A8. $\neg E(\alpha \wedge \neg\alpha)$

[6]Rispetto alla formulazione del par. 8.3, abbiamo sostituito, per uniformità con la notazione utilizzata in questo capitolo, l'operatore B con l'operatore di credenza esplicita E.

regole di inferenza:

R1. da α e da $\alpha \rightarrow \beta$ si inferisce β (*modus ponens*)
R3. da α si inferisce $E\alpha$
R4. da $\alpha \rightarrow \beta$ si inferisce $E\alpha \rightarrow E\beta$.

Si può provare che tutti i teoremi di **EMNP** sono anche teoremi della logica per la conoscenza locale, e viceversa. Infatti, tutti gli assiomi e le regole di **EMNP** sono ottenibili rispettivamente come teoremi o come regole derivate nella logica per la conoscenza locale. L'assioma P è lo stesso di A8, tutti gli esempi di N sono ottenibili applicando R3, e la regola RE è facilmente ottenibile come regola derivata utilizzando R4. Per quanto riguarda M, può essere derivato come segue:

(1) $\alpha \wedge \beta \rightarrow \alpha$
(2) $\alpha \wedge \beta \rightarrow \beta$
(3) $E(\alpha \wedge \beta) \rightarrow E\alpha$
(4) $E(\alpha \wedge \beta) \rightarrow E\beta$
(5) $E(\alpha \wedge \beta) \rightarrow (E\alpha \wedge E\beta)$

dove (1) e (2) sono tautologie proposizionali, (3) e (4) sono ottenute rispettivamente da (1) e (2) per mezzo di R4, e (5) è conseguenza proposizionale di (3) e (4).

Viceversa, tutti gli assiomi e le regole della logica per il ragionamento locale possono essere ottenuti rispettivamente come teoremi o come regole derivate di **EMNP**. Di nuovo, A8 è la stessa cosa di P, R3 si ottiene facilmente come regola derivata utilizzando RE e N, R4 si ottiene come regola derivata nel modo seguente:

(1) $\alpha \rightarrow \beta$
(2) $\alpha \leftrightarrow (\alpha \wedge \beta)$
(3) $E\alpha \leftrightarrow E(\alpha \wedge \beta)$
(4) $E(\alpha \wedge \beta) \rightarrow (E\alpha \wedge E\beta)$
(5) $E\alpha \rightarrow E\beta$

(dove (1) è l'ipotesi, (2) si ottiene da (1) per mezzo delle leggi del calcolo proposizionale, (3) si ottiene da (2) per mezzo di RE, (4) è l'assioma M, e (5) segue da (3) e da (4) per le leggi del calcolo proposizionale). Quindi la logica del ragionamento locale di Fagin e Halpern è corretta e completa (per la parte che concerne la credenza esplicita) rispetto alle strutture di credenza chiuse rispetto ai soprainsiemi, contenenti l'unità e tali che, per ogni $w \in W, \varnothing \notin N(w)$[7].

11.4 I diversi volti della credenza esplicita

Ciò che emerge a un primo esame del contenuto di questo capitolo e dei capitoli precedenti è probabilmente l'eterogeneità delle varie soluzioni proposte nell'ambito dell'IA logicista per evitare il problema dell'onniscienza logica[8]. Questa eterogeneità è analoga all'eterogeneità delle proposte avanzate in ambito filosofico[9]. Esistono anzi somiglianze precise fra le varie proposte avanzate in filosofia e nell'IA, e molte di queste ultime traggono ispirazione da ricerche di ambito filosofico. Nelle pagine precedenti abbiamo rapidamente accennato ad alcuni dei punti di contatto fra queste due tradizioni, ed altri punti di contatto potrebbero essere individuati. Sarebbe indubbiamente interessante indagare a fondo le relazioni fra questi due ambiti di ricerca, ma ciò costituirebbe un argomento di indagine ulteriore, che va oltre gli scopi di questo lavoro. Tuttavia, vale la pena sottolineare che la ricerca in IA ha messo in luce come l'eterogeneità delle varie proposte non possa essere considerata un fattore del tutto accidentale. Molte proposte che originariamente erano state interpretate come alternative per la soluzione di un unico problema, sono state invece in seguito inquadrate come possibili vie per affrontare aspetti diversi di un problema complesso. Ci sembra infatti essenziale la consapevolezza, emersa ad esempio con il lavoro di Fagin e Halpern (cfr. l'introduzione di questo capitolo), che la mancanza di onniscienza logica è un fenomeno che dipende dall'interazione di più fattori diversi, e che quindi può essere proficuo affrontare con strumenti differenti. Vardi (1986) sottolinea come alcuni problemi possano essere alleviati pur

[7]Anche una logica dell'inconsistenza basata su una intuizione simile a quella di Fagin e Halpern, proposta da Rescher e Brandom (1980), risulta equivalente al sistema **EMNP** - cfr. Vardi (1986).

[8]Per un modello che sintetizzi in una cornice unitaria varie di queste proposte si veda (Giunchiglia et al. 1993). Il quadro utilizzato è quello del ragionamento basato su contesti (Giunchiglia 1993). Per questo approccio al ragionamento modale ed epistemico si veda anche (Giunchiglia e Serafini 1994).

[9]Rimandiamo ancora alla già citata rassegna di Bäuerle e Cresswell (1988).

mantenendo l'assunzione che gli atteggiamenti proposizionali siano atteggiamenti verso *proposizioni* intese nel senso model teoretico classico (come accade ad esempio nelle logiche basate su modelli minimali). Per altri aspetti è necessario introdurre mondi non classici (come le situazioni incoerenti o incomplete di Levesque). Questo comporta un desiderabile indebolimento della logica, ma i soggetti epistemici così modellati sono pur tuttavia onniscienti rispetto a qualche tipo di logica non classica. Gli aspetti dell'onniscienza logica connessi in maniera più specifica alla finitezza dei processi deduttivi nei soggetti epistemici reali non sembrano risolvibili in altra via che tenendo conto, in qualche misura, delle proprietà sintattiche delle rappresentazioni utilizzate. Nella letteratura filosofica gli approcci al problema degli atteggiamenti proposizionali basati in qualche misura sulla forma sintattica delle rappresentazioni e quelli basati invece su concetti esclusivamente modellistici, come la nozione di proposizione in quanto insieme di mondi possibili, sono viste spesso come corrispondenti a visioni alternative del linguaggio e della mente. Ad esempio, Stalnaker (1984) a questo proposito distingue fra un *quadro linguistico* (*linguistic picture*) e un *quadro pragmatico* (*pragmatic picture*) della mente. Secondo il quadro linguistico, la razionalità è considerata soprattutto come una faccenda di uso del linguaggio. Gli stati mentali sono visti come rappresentazioni interne con una struttura simile a quella delle espressioni linguistiche. In generale, avere una certa credenza vuol dire stare in una relazione di un certo tipo con una rappresentazione linguistica (di tipo mentale o del linguaggio pubblico). Al contrario, il quadro pragmatico vede i soggetti razionali essenzialmente come agenti. Gli stati mentali sono visti soprattutto in relazione al loro ruolo nel determinare il comportamento. Essere in un certo stato mentale vuol dire porsi in un certo atteggiamento rispetto a un insieme di possibili stati del mondo. Ad esempio, in questa prospettiva, credere che α significa essere disposti ad agire (ad esempio per soddisfare i propri desideri o per ottenere i propri fini) *come se nel mondo fosse vero* α. In questo caso gli oggetti della credenza (e, in generale, degli atteggiamenti proposizionali) vengono visti essenzialmente come insiemi di mondi possibili. Secondo il quadro pragmatico, con cui Stalnaker si identifica, non viene negato che agli stati mentali corrisponda qualche forma di rappresentazione interna. Esso non si impegna tuttavia su quale tipo di rappresentazioni corrisponda agli stati mentali, e la forma delle rappresentazioni non viene considerata rilevante per affrontare, ad esempio, il problema degli atteggiamenti proposizionali. Il lavoro in IA ha posto in evidenza che le soluzioni tecniche che emergono dal quadro linguistico e dal quadro pragmatico rispondono ad esigenze in larga parte complementari. Ad esempio, Stalnaker (1984), in linea con la sua visione pragmatica, elabora una soluzione al problema dell'onniscienza logica che può essere assimilata alla logica del ragionamento locale di Fagin e Halpern (cfr. paragrafo precedente). La soluzione pragmatica ha l'indubbio pregio di non attribuire un ruolo prioritario all'uso del linguaggio (o di qualche rappresentazione di tipo simbolico) nella definizione e nell'attribuzione degli atteggiamenti proposizionali. In una prospettiva cognitiva, non vincola a ipotesi specifiche sulla natura degli stati mentali o sulle strutture cognitive da utilizzare, e consente di rendere conto in modo plausibile della possibilità di attribuire atteggiamenti proposizionali a soggetti cognitivi che non esibiscono un comportamento di tipo linguistico o simbolico. Tuttavia, vi sono aspetti del problema dell'onniscienza logica che non possono essere risolti in maniera naturale in una prospettiva del genere. Ne sono un esempio gli aspetti della mancanza di onniscienza logica dovuti ai limiti nella lunghezza dei processi inferenziali effettuabili da un soggetto epistemico finito, aspetti che è difficile tentare di spiegare se non si prendono in considerazione le elaborazioni effettuate sulla struttura delle rappresentazioni interne[10]. Ad esempio, secondo il quadro pragmatico così formulato, un giocatore di scacchi "finito", che non sia in grado di prendere in considerazione tutti i possibili esiti delle proprie mosse, risulterebbe probabilmente irrazionale. (Un giocatore di scacchi che conosca la configurazione iniziale dei pezzi e le regole del gioco dovrebbe essere in grado di dedurre tutte le possibili evoluzioni della partita, e di decidere quindi la mossa ottimale; egli dispone inoltre contemporaneamente di tutte le informazioni rilevanti, e non è quindi plausibile spiegare i suoi limiti in termini di *frame of mind*). Lo stesso dicasi di un matematico che non riesca a prendere in considerazione tutte le conseguenze logiche delle proprie premesse. In particolare, se si assume che le verità matematiche siano necessarie, chi sapesse un solo enunciato matematicamente vero dovrebbe sapere tutta la matematica. In questi casi sembra inevitabile prendere in considerazione qualche aspetto della sintassi delle rappresentazioni. La soluzione che propone Stalnaker a casi di questo genere è quella di considerare alcune forme di ragionamento come ad esempio il ragionamento matematico come forme di ragionamento di tipo metalinguistico:

> Ora, se le verità matematiche sono tutte necessarie, allora l'analisi in termini di mondi possibili non lascia spazio al dubbio circa la verità delle proposizioni stesse. Ci sono, sembra, solo due proposizioni matematiche, quella necessariamente vera e quella necessariamente falsa, e tutti noi sappiamo che la prima è vera e la seconda è falsa. Ma le funzioni che determinano quali delle due proposizioni sia espressa da un dato enunciato matematico sono di quel tipo che è ragionevolmente complesso da dare luogo a ragionevoli dubbi circa quale proposizione sia espressa da un enunciato. Quindi sembra ragionevole assumere che gli oggetti della credenza e del dubbio in matematica siano proposizioni relative alla relazione fra enunciati e ciò che essi esprimono. [...] *le proposizioni matematiche sono proposizioni che vertono su espressioni, o su strutture esibite dalle espressioni.* (Stalnaker 1984, pp. 73-4; corsivo nostro).

[10]Di questo sembra consapevole lo stesso Stalnaker quando, in un passo di un lavoro successivo che abbiamo già citato, afferma: "ogni tipo di elaborazione dell'informazione o di computazione è priva di senso se riferita a un agente onnisciente dal punto di vista deduttivo" (Stalnaker 1991, p. 429).

Una posizione di questo genere dà luogo a molte perplessità. In primo luogo, sembra che le soluzioni di tipo sintattico, cacciate dalla porta, siano rientrate dalla finestra. Inoltre, non si vede come su questa linea possa emergere una soluzione convincente per problemi come quello del giocatore di scacchi. Poco convincente risulta anche la visione della matematica (e della ricerca matematica) che ne emerge, e la possibilità di spiegare il ruolo che la conoscenza matematica può avere nel determinare il comportamento pratico. Si consideri l'aritmetica. Se gli enunciati aritmetici sono veri in tutti i mondi possibili, hanno tutti la stessa estensione, ed esiste quindi un sola proposizione aritmetica vera. Poiché, ragionevolmente, nessuno è in grado di comportarsi come se tutti gli enunciati aritmetici fossero veri, ne segue che nessuno conosce l'unica proposizione aritmetica vera. Quindi, nessuno *sa* che 2 + 2 fa 4 (e nessuno è in grado di comportarsi come se nel mondo 2 + 2 = 4 fosse vero, perché altrimenti dovrebbe essere in grado di comportarsi come se nel mondo fossero vere tutte le verità aritmetiche). La gente al massimo può sapere (e può comportarsi come se fosse vero che) elaborando in base a certe regole l'espressione "2 + 2" si ottiene l'espressione "4".

Il fatto che, almeno per alcuni aspetti del problema dell'onniscienza logica, sembri indispensabile tenere conto in qualche misura della forma sintattica delle rappresentazioni, non significa tuttavia che, per risolvere tali problemi, si debba necessariamente capitolare di fronte a una concezione di tipo citazionale, o puramente sintattica degli atteggiamenti proposizionali (come ad esempio, in IA, quella proposta da Konolige - cfr. par. 9.2). Proposte come quella di Fagin e Halpern di una logica della consapevolezza generalizzata (par. 11.2) consentono di affrontare il problema limitandosi a iniettare "piccole dosi di sintassi" (l'espressione è di Vardi 1986) in un quadro teorico di tipo genuinamente semantico e model teoretico. La situazione è ben sintetizzata dalla seguente citazione di Barbara Partee (1982):

> non intendo sostenere una teoria citazionale degli enunciati di atteggiamento proposizionale. Ma si assume che le persone non possano avere atteggiamenti verso insiemi di mondi possibili direttamente, senza dipendere da tutti i modi di rappresentarli in modo finito. (Partee 1982, p. 197 della trad. it.)

I motivi per rifiutare una soluzione puramente sintattica e citazionale sono molteplici. Ad alcuni abbiamo già fatto cenno alla fine del paragrafo 9.2. Alcune difficoltà ripetutamente esaminate nella letteratura filosofica sono legate al fatto di assumere enunciati "pubblici" di un linguaggio naturale come oggetti degli atteggiamenti proposizionali (che è la linea che è stata seguita dalla maggior parte delle situazioni di tipo "sintattico" e citazionale in ambito filosofico). Si supponga ad esempio che un enunciato del tipo:

S crede che α

sia analizzato come:

S è nella relazione di credenza con l'enunciato "α*",*

dove α è un certo enunciato dell'italiano. Questa strategia non funziona nel caso di un enunciato quale:

Mario crede che tutto quello che dico è vero,

che, evidentemente, non può essere parafrasato come:

Mario è nella relazione di credenza con l'enunciato "tutto quello che dico è vero".

In generale, la soluzione citazionale non è direttamente applicabile a tutti quei casi in cui il significato dell'enunciato che è oggetto di atteggiamento proposizionale dipende dal proprio contesto, come nel caso che vi compaiano espressioni indicali o anaforiche. Vi è poi l'ovvia osservazione che il soggetto di credenza S potrebbe non conoscere la lingua in cui è scritto α. Ad esempio, non avrebbe molto senso analizzare l'enunciato

(*) *Aristotele credeva che gli uomini sono animali razionali*

come:

Aristotele era nella relazione di credenza con l'enunciato "gli uomini sono animali razionali",

in quanto Aristotele certamente non poteva conoscere l'italiano. La difficoltà può essere aggirata facendo riferimento a una traduzione dell'oggetto di credenza, e parafrasando ad esempio (*) con

(**) *Aristotele era nella relazione di credenza con un enunciato la cui traduzione italiana è "gli uomini sono animali razionali".*

Anche questa soluzione pone tuttavia problemi, come ha posto in luce Church (1950). Church sviluppa il seguente argomento. In base a una teoria citazionale, (*) e (**) dovrebbero avere uguale significato. Tuttavia, si supponga di tradurre (*) e (**) in inglese, ottenendo rispettivamente:

(*') *Aristotle believed that men are rational animals*

e

(**') *Aristotle was in the belief relation with a sentence, whose Italian translation is "gli uomini sono animali razionali".*

Traducendo nella stessa lingua due enunciati con uguale significato si dovrebbero ottenere enunciati con uguale significato, mentre, afferma Church, (*') e (**') hanno chiaramente significato diverso. Infatti, ad esempio, essi comunicano contenuti completamente diversi a un lettore inglese che non conosca l'italiano. Di conseguenza, (*) e (**) non possono essere a loro volta sinonimi.

Difficoltà di questo tipo sembrano poter essere superate facilmente qualora si assuma che gli atteggiamenti proposizionali possano essere spiegati non nei termini di relazioni con enunciati "pubblici" di una lingua naturale, bensì con "enunciati interni" di un "linguaggio del pensiero" (una posizione di questo genere è sostenuta da Fodor 1978). In tal caso, tali difficoltà non costituirebbero un problema per le soluzioni puramente sintattiche di tipo cognitivo, o per i modelli sintattici proposti in IA, come ad esempio quello di Konolige. Secondo Cresswell (1980) tuttavia l'argomento di Church è di portata più generale, e si applica a tutte le teorie di tipo citazionale degli atteggiamenti proposizionali. In particolare, non può essere aggirato facendo riferimento a un sistema di rappresentazioni interne anziché al linguaggio pubblico. A questo proposito Cresswell utilizza alcune argomentazioni, che egli riconduce a varianti dell'argomento di Church, in cui non si fa uso del concetto di traduzione. Riportiamo uno di tali argomenti, dovuto a Bigelow (1978). Si consideri l'enunciato

(1) *Giorgio crede che la terra è piatta*

e la sua parafrasi citazionale:

(2) *Giorgio è nella relazione di credenza con l' enunciato "la terra è piatta".*

Si considerino poi due interpretazioni modellistiche del linguaggio in cui sono state formulate (1) e (2). Siano φ e ψ le rispettive funzioni interpretazione. In particolare, si abbia che:

φ[*Giorgio*] = ψ[*Giorgio*]
φ[*crede*] = ψ[*crede*]
φ[*che*] = ψ[*che*]
φ[*la terra*] = ψ[*la terra*]
φ[*è piatta*] = ψ[*è tonda*] ≠ ψ[*è piatta*].

Ora, è evidente che, in generale, si avrà φ[(1)] ≠ ψ[(1)]. D'altra parte, se le virgolette indicano la citazione in entrambe le interpretazioni, si ha che:

φ[*"la terra è piatta"*] = ψ[*"la terra è piatta"*].

Assumendo che φ e ψ differiscano esclusivamente per il significato che assegnano alle espressioni "*è tonda*" ed "*è piatta*", si avrà quindi che φ[(2)] = ψ[(2)]. Dunque, il significato di (2) resta lo stesso nelle due interpretazioni, a prescindere dal fatto che in esse il significato intuitivo di (1) è completamente diverso. Ne segue che (2) non può essere considerata un'analisi adeguata del significato di (1). Questo argomento non dipende dal fatto che si assumano come oggetti di credenza enunciati di un linguaggio pubblico, e può essere riformulato nel caso venga usato un sistema di rappresentazioni interne. In generale quindi, una spiegazione citazionale degli atteggiamenti proposizionali è inadeguata, a prescindere dal fatto che si faccia riferimento a un linguaggio pubblico o a un sistema di rappresentazione interne. Parafrasando Bäuerle e Creswell (1988), si può concludere che il problema centrale consiste nel fatto che la relazione di credenza coinvolge il significato. "Credere" non può essere una relazione fra un soggetto di credenza e un enunciato, in quanto lo stesso enunciato può significare cose completamente diverse in base a diverse interpretazioni. Se quindi, almeno per certi aspetti dell'onniscienza logica, sembra necessario far riferimento in qualche misura alla forma sintattica delle rappresentazioni, la soluzione delle "piccole dosi di sintassi" (come ad esempio nella logica della consapevolezza generalizzata di Fagin e Halpern) sembra inevitabile. In ogni caso, gli argomenti di Church-Cresswell rendono inaccettabile affermazioni come quella di Konolige (1986b), secondo cui l'impostazione alla base della logica della consapevolezza generalizzata sarebbe "una cattiva idea: finisce per essere equivalente a una versione più complicata del *sentential approach* [cioè del modello sintattico proposto da Konolige stesso] in un quadro a mondi possibili" (Konolige 1986b, p. 242).

Un ulteriore punto che accomuna le ricerche sull'onniscienza logica in IA, distinguendole dal lavoro svolto a tale proposito in ambito model teoretico e filosofico, può essere individuato nell'abbandono del tentativo di individuare gli "oggetti" degli atteggiamenti proposizionali, nei termini di entità semantiche (di "significati"), che si collochino a un livello intermedio fra enunciati e proposizioni, e che siano dotati di una consistenza ontologica di tipo extramentale e intersoggettivo, analoga a quella dei sensi fregeani. Questa è la prospettiva che viene adottata dalla maggior parte delle ricerche nell'ambito della semantica filosofica, e fra esse da alcune che seguono la via di introdurre elementi sintattici nella semantica, o che comunque tengono conto della struttura sintattica delle rappresentazioni. L'esempio recente più esplicito a questo proposito è la teoria dei *significati strutturati* (*structured meaning*) di Cresswell (1985), sviluppata a partire dall'isomorfismo intensionale di Carnap attraverso le proposte di Lewis (1972). Secondo Cresswell un significato strutturato è "una 'entità pubblica' nel senso che la sua struttura dipende dalla struttura dell'enunciato usato nell'attribuzione della credenza" (Bäuerle e Cresswell 1988, p. 499). Queste affermazioni sono motivate in base a considerazioni di diretta ascendenza fregeana: "I significati non possono essere entità private. Due persone possono avere la stessa credenza, e la credenza che essi hanno in comune deve essere qualcosa cui devono poter fare pubblico riferimento" (ivi, p. 498). In ultima analisi, questo tipo di scelte è dettato da preoccupazioni di origine comportamentista e verificazionista (per questo tipo di critica applicata alla teoria dell'isomorfismo intensionale di Carnap si veda Fodor 1978). Lo scopo dell'analisi degli enunciati di atteggiamento proposizionale viene visto in ultima istanza nell'individuazione di criteri per l'attribuzione degli atteggiamenti stessi, criteri che devono essere intersoggettivamente accessibili, e quindi non possono far riferimento a entità di tipo privato (ad esempio di tipo mentale). Viceversa, nelle logiche dell'IA il tentativo è comunque sempre quello di tener conto delle strutture e delle rappresentazioni, e dei processi computazionali, associati ai singoli soggetti epistemici, pure se in un quadro di tipo, in senso lato, logico. Torneremo su questi problemi nel prossimo capitolo. Per il momento, ci limitiamo ad osservare che l'attribuzione di una credenza a un dato soggetto è sempre un'operazione a carattere ipotetico, sottodeterminata dal comportamento (linguistico o di altro tipo) del soggetto stesso. Non c'è nulla quindi di problematico nel fatto di riferirsi a costrutti teorici e a entità di tipo non "osservabile" (come appunto le rappresentazioni o gli stati mentali). D'altro canto, circa la possibilità che più soggetti possano avere la stessa credenza, ciò non è incompatibile con l'assunzione della natura "mentale" degli oggetti di credenza se si accettano alcune assunzioni sulla natura degli stati mentali che sono ampiamente condivise nella psicologia contemporanea (si veda oltre, nel par. 12.4, la discussione sulla distinzione fra competenza ed esecuzione nella psicologia cognitiva in rapporto all'antipsicologismo di Frege).