

## 9. Credere enunciati

### 9.1 Modelli sintattici per gli atteggiamenti proposizionali

Abbiamo visto che, rispetto al problema dell'onniscienza logica, il limite delle semantiche a mondi possibili per le logiche della credenza può essere considerato una conseguenza del fatto di assumere come oggetto della credenza proposizioni, ossia intensioni di enunciati. Nella semantica di tipo modellistico una proposizione è una funzione da mondi possibili a valori di verità. Poiché due enunciati logicamente equivalenti sono veri esattamente negli stessi mondi possibili, allora due enunciati equivalenti hanno la stessa intensione, e sono quindi indiscernibili dal punto di vista semantico: se è creduto uno, allora, inevitabilmente, deve essere creduto anche l'altro. Tuttavia, per come sono definite le intensioni, è del tutto legittimo che enunciati con la stessa intensione si comportino in maniera differente rispetto ai contesti di atteggiamento proposizionale. Una possibile soluzione può essere cercata tentando di introdurre una classificazione "a grana più fine" per gli oggetti degli atteggiamenti proposizionali. In questa direzione si muoveva già, ad esempio, la proposta dell'isomorfismo intensionale di Carnap, cui abbiamo fatto cenno nel paragrafo 2.1. Una via estrema, esplorata in ambiti diversi (filosofico, cognitivo, e, appunto, in intelligenza artificiale) consiste nell'assumere che gli oggetti degli atteggiamenti proposizionali e, in particolare, del credere, siano enunciati, vale a dire oggetti sintattici, espressioni linguistiche di qualche tipo di linguaggio<sup>1</sup>. Nell'ambito dell'intelligenza artificiale, un "manifesto" di questo modo di concepire la credenza può essere considerato il lavoro di Moore e Hendrix (1979 e 1982). La posizione di Moore e Hendrix è formulata in termini di tipo psicologico, ed è legata all'ipotesi della mente come sistema simbolico, e dell'esistenza di un linguaggio del pensiero. In questa prospettiva, quindi, gli oggetti degli atteggiamenti proposizionali vengono interpretati come espressioni di un linguaggio mentale interno. "Nel nostro modello - affermano Moore e Hendrix - il credere è spiegato dal fatto che un sistema è in una certa relazione computazionale con espressioni di un linguaggio interno" (Moore e Hendrix 1982, p. 112). Secondo Moore e Hendrix, perché un'espressione del linguaggio mentale esprima una credenza del soggetto cognitivo devono essere soddisfatte alcune condizioni. Innanzi tutto, tale espressione deve essere presente esplicitamente nella memoria del sistema, o comunque deve stare in una relazione opportunamente specificata (ad esempio nei termini di certi tipi di inferenza) con le espressioni immagazzinate esplicitamente in memoria. Inoltre, perché un'espressione esprima una credenza di un soggetto non è sufficiente che sia immagazzinata nella sua memoria, in quanto altri atteggiamenti proposizionali oltre al credere (ad esempio temere, desiderare, e così via) possono essere spiegati nei termini di relazioni con espressioni memorizzate del linguaggio mentale. Perché possano essere differenziate dagli oggetti degli altri atteggiamenti proposizionali, le credenze devono essere specificate funzionalmente, sulla base del tipo di processi (che qui non ci interessa indagare) che operano su di esse e che le connettono agli *input* e agli *output* del sistema.

L'insieme delle espressioni che costituiscono le credenze del sistema viene detto da Moore e Hendrix *insieme di credenza* (*belief set*) del sistema stesso. Essi assumono che le espressioni di un *belief set* possano essere considerate, almeno in prima approssimazione, analoghe a formule del calcolo dei predicati del primo ordine, esteso con operatori intensionali per gli atteggiamenti proposizionali.

E' chiaro che in questa prospettiva il problema dell'onniscienza logica, e in particolare il problema di discriminare fra credenze logicamente equivalenti, non si pone. Nel caso di due credenze logicamente equivalenti, ad esempio, se esse corrispondono a enunciati diversi, vale a dire a diverse espressioni simboliche del linguaggio interno del sistema, esse devono comunque essere considerate distinte, ed è del tutto concepibile che una di esse appartenga al *belief set* del sistema senza che vi appartenga anche l'altra.

Si supponga - dicono Moore e Hendrix - che le credenze siano individuate più o meno come formule del linguaggio interno. Si supponga inoltre che il sistema abbia una particolare formula  $P$  nel suo *belief set* che sia logicamente equivalente a un'altra formula  $Q$ , nel senso che esiste un modo di applicare le procedure di inferenza di base del sistema per inferire  $P$  da  $Q$  e viceversa. Tuttavia, il sistema potrebbe non inserire  $Q$  nel suo *belief set*, perché non ha mai provato a derivare  $Q$ , o perché le sue euristiche per applicare le sue procedure di inferenza non sono sufficienti per trovare la derivazione di  $Q$ , o perché la derivazione di  $Q$  è talmente lunga che esaurirebbe le risorse di tempo e di memoria del sistema. Solleviamo l'attenzione su questo punto perché la possibilità che 'A crede  $P$ ' sia vera e 'A crede  $Q$ ' sia falsa, sebbene  $P$  e  $Q$  siano logicamente equivalenti, è correntemente considerata uno dei problemi principali nella semantica degli enunciati di credenza, specialmente per teorie basate sulla semantica dei mondi possibili. (Moore e Hendrix 1982, p. 114).

Moore e Hendrix non sviluppano in senso tecnico la loro proposta. La loro è una sorta di dichiarazione di intenti sul metodo che ritengono adeguato per affrontare il problema degli atteggiamenti proposizionali. D'altro canto, in

---

<sup>1</sup>Nella letteratura filosofica sono state formulate numerose proposte di assumere come oggetti degli atteggiamenti proposizionali degli enunciati. Queste posizioni, dette anche di tipo *citazionale*, si distinguono da quelle analoghe sviluppate in IA e nelle scienze cognitive per il fatto di utilizzare enunciati del linguaggio pubblico anziché rappresentazioni linguistiche "interne". Le due prospettive sono messe a confronto ad esempio in (Fodor 1978).

letteratura si possono trovare esempi di modelli sintattici delle credenze formulati nei termini di insiemi di enunciati anche in lavori più antichi di quelli di Moore e Hendrix (come ad esempio Eberle 1974). Noi prenderemo in considerazione nel prossimo paragrafo il modello di Konolige (1984, 1985a, 1986a), il quale sviluppa la proposta sintattica in termini che, dal punto di vista logico, risultano più ricchi ed articolati. Konolige utilizza un linguaggio epistemico modale al quale associa una semantica in cui le credenze dei soggetti di credenza sono modellate come insiemi di formule. In generale, tali insiemi di formule sono deduttivamente chiusi rispetto a insiemi di regole di inferenza logicamente incomplete.

Prima di procedere ad esporre il lavoro di Konolige è tuttavia opportuna una precisazione. Le trattazioni della credenza che prendiamo in considerazione in questo paragrafo sono caratterizzate dal fatto di considerare enunciati, cioè entità di tipo sintattico, quali oggetti del credere. Ciò ha fatto sì che a questo proposito si parlasse talvolta di trattamenti *sintattici* della credenza (ad esempio Levesque (1984b) parla di *syntactic approach* a proposito di Konolige e di Moore e Hendrix). Questo ha generato alcune ambiguità terminologiche rispetto a un'altra linea di ricerca sul trattamento logico degli atteggiamenti proposizionali e della credenza in particolare, che è stata anch'essa spesso qualificata come *approccio sintattico* al problema della credenza. Tale linea di ricerca è caratterizzata dal fatto di utilizzare un linguaggio predicativo del primo ordine anziché un linguaggio modale per formulare la logica degli atteggiamenti proposizionali. Ad esempio, anziché introdurre nel linguaggio un operatore modale di credenza  $B$  che assuma come argomenti formule del linguaggio (per cui "è creduto che  $\alpha$ " si esprime mediante la formula  $B\alpha$ ), la credenza viene espressa mediante un predicato ad un posto (ad esempio  $bel(x)$ ), che assume come argomenti *nomi* degli oggetti di credenza. Così, a prescindere dai dettagli tecnici, se " $\alpha$ " è il nome dell'oggetto di credenza espresso dalla formula  $\alpha$ , "è creduto che  $\alpha$ " si esprime mediante la formula  $bel(" \alpha ")$ <sup>2</sup>. In intelligenza artificiale questo modo di procedere è stato proposto ad esempio da McCarthy(1979). Le formalizzazioni predicative al primo ordine degli atteggiamenti proposizionali sono, di norma, più espressive dei linguaggi proposizionali o del primo ordine modali. Il punto centrale è che, poiché gli oggetti degli atteggiamenti proposizionali sono rappresentati mediante termini del linguaggio, diventa possibile quantificare su di essi. Così, ad esempio, è possibile scrivere formule del tipo:

$$\exists x(\alpha(x) \wedge bel(x))$$

oppure:

$$\forall x(\alpha(x) \rightarrow bel(x))$$

(dove  $\alpha(x)$  è una formula qualunque del linguaggio in cui compare libera la variabile  $x$ ), le quali non hanno corrispettivo in un linguaggio modale del primo ordine.

Tale potenza espressiva comporta però l'insorgere di inconsistenze. Nel caso si utilizzi la tecnica della gödelizzazione come strumento per far riferimento alle formule di una teoria dall'interno della teoria stessa, Montague (1963) ha dimostrato che teorie del primo ordine che includano assiomi corrispondenti a quelli del sistema modale  $\mathbf{T}$  assieme agli assiomi per l'aritmetica elementare sono inconsistenti<sup>3</sup>. Consideriamo una teoria del primo ordine  $\mathbf{P}_T$  il cui linguaggio comprenda i simboli per l'aritmetica elementare e un predicato a un posto  $know$  (trattandosi di un "equivalente" di  $\mathbf{T}$ , in termini epistemici avremo a che fare con una teoria della conoscenza anziché della credenza). Assumiamo inoltre che, per ogni formula  $\alpha$  del linguaggio di  $\mathbf{P}_T$ , " $\alpha$ " sia il numerale che corrisponde al numero di Gödel di  $\alpha$ . Montague dimostra che, se  $\mathbf{P}_T$  comprende l'assioma per l'aritmetica di Robinson, e se in  $\mathbf{P}_T$  si ha che, per qualsiasi formula  $\alpha$  e  $\beta$ , sono dimostrabili:

- (i)  $know(" \alpha ") \rightarrow \alpha$ ;
- (ii)  $know(" \alpha \rightarrow \beta ") \rightarrow (know(" \alpha ") \rightarrow know(" \beta "))$ ;

e se inoltre vale la regola:

$$\text{PNec: da } \alpha \text{ segue } know(" \alpha "),$$

allora  $\mathbf{P}_T$  è inconsistente. Si noti che (i) è l'equivalente, in un linguaggio del primo ordine, dell'assioma T, (ii) corrisponde all'assioma di distribuzione, e PNec è l'equivalente della regola di necessitazione. La fonte dell'inconsistenza sta nel fatto che la rappresentabilità dell'aritmetica in  $\mathbf{P}_T$  fa sì che il linguaggio sia autoreferenziale, per cui è possibile derivare in  $\mathbf{P}_T$  una formula "paradossale" analoga a quella utilizzata da Tarski per dimostrare

<sup>2</sup>Nel caso di una logica a più soggetti epistemici  $bel$  può essere definito come un predicato a due argomenti, di cui il primo è un termine che denota un soggetto epistemico, e il secondo è il nome di un oggetto di credenza. Ad esempio,  $bel(Nicola, " \alpha ")$  significherebbe che Nicola crede che  $\alpha$ .

<sup>3</sup>Montague (1963) faceva riferimento al trattamento delle modalità aletiche, ma il suo risultato è automaticamente estendibile ai contesti epistemici.

l'impossibilità di introdurre in maniera consistente un predicato di verità in un linguaggio dotato di autoriferimento. La dimostrazione del teorema di Tarski si basa sul fatto che in un sistema formale del primo ordine che comprenda il proprio predicato di verità e che sia in grado di far riferimento alle sue stesse espressioni, può essere dimostrata una formula equivalente al paradosso del mentitore, ossia un enunciato che asserisce la propria falsità (vale a dire, qualcosa del tipo: "questo enunciato è falso"). Analogamente, in una formalizzazione al primo ordine di una logica della conoscenza che comprenda l'aritmetica di Robinson, esiste una formula  $\alpha$  per la quale è possibile dimostrare:

$$(iii) \quad \alpha \leftrightarrow know(" \neg \alpha ").$$

Intuitivamente,  $\alpha$  afferma che è conosciuta la sua stessa negazione<sup>4</sup>. La formula (iii), assieme a (i), (ii) e PNec, consente di derivare una contraddizione<sup>5</sup>. Si tratta di quello che è usualmente noto come *the knower paradox*.

Il risultato di Montague non dipende dal fatto di utilizzare la gödelizzazione. E' possibile derivare il paradosso ogni qual volta si disponga in una teoria del primo ordine di termini che denotano gli oggetti della conoscenza, e si disponga inoltre di un apparato di operatori sintattici che garantiscano la possibilità di effettuare le necessarie manipolazioni formali sulla struttura di tali termini (come ad esempio un operatore di sostituzione che operi sui nomi delle formule)<sup>6</sup>.

Thomason (1980) ha mostrato che un risultato analogo a quello di Montague può essere esteso a taluni casi di teorie per la credenza espresse mediante un linguaggio del primo ordine. In particolare, sorgono problemi di consistenza nel trattamento mediante un linguaggio predicativo del primo ordine di **S5** debole. Sia  $\mathbf{P}_{WS5}$  una teoria del primo ordine, il cui linguaggio comprenda i simboli per l'aritmetica e un predicato a un posto *bel*. Per ogni formula  $\alpha$  e  $\beta$  del linguaggio, in  $\mathbf{P}_{WS5}$  sia dimostrabile quanto segue:

- (i)  $bel("Q")$ , dove  $Q$  è l'assioma per l'aritmetica di Robinson;
- (ii)  $bel(" \alpha \rightarrow \beta ") \rightarrow (bel(" \alpha ") \rightarrow bel(" \beta "));$
- (iii)  $bel(" \alpha ") \rightarrow bel("bel(" \alpha ")");$
- (iv)  $\neg bel(" \alpha ") \rightarrow bel(" \neg bel(" \alpha ")");$

e inoltre valga la regola:

$$PNec': \text{ da } \alpha \text{ segue } bel(" \alpha ").$$

---

<sup>4</sup>La possibilità di dimostrare (iii) in  $\mathbf{P}_T$  è una conseguenza del cosiddetto *lemma diagonale* (*diagonal lemma*) (si veda ad esempio il cap. 15 di Boolos e Jeffrey 1974). Secondo il lemma diagonale, in una teoria del primo ordine che comprenda l'aritmetica di Robinson (e nella quale quindi sono rappresentabili tutte le funzioni ricorsive) vale quanto segue: per ogni formula  $\beta(x)$  della teoria, con  $x$  come unica variabile libera, esiste una formula chiusa  $\alpha$  tale che è dimostrabile:

$$\alpha \leftrightarrow \beta(" \alpha ").$$

<sup>5</sup>Il risultato di Montague è più forte. Per derivare l'inconsistenza da (iii) è sufficiente disporre di (i) e di PNec. Infatti, mediante la logica proposizionale, da (i) e da (iii) si ottiene:

$$\alpha \rightarrow \neg \alpha$$

e quindi:

$$(*) \quad \neg \alpha,$$

da cui, applicando PNec, si ottiene:

$$K(" \neg \alpha ").$$

$K(" \neg \alpha ")$ , in base a (iii), è equivalente ad  $\alpha$ , ed è quindi in contraddizione con (\*). Alternativamente, Montague dimostra che si ottiene l'inconsistenza qualora in una teoria del primo ordine si disponga, oltre che dell'aritmetica di Robinson, di (i), (ii) e di:

- (iv)  $know(" \alpha ")$ , dove  $\alpha$  è un assioma della logica del primo ordine;
- (v)  $know("know(" \alpha ") \rightarrow \alpha ")$

(dove (iv) e (v), assieme, costituiscono un indebolimento di PNec).

<sup>6</sup>Un tale apparato è necessario ad esempio per mettere in relazione un termine come " $\alpha$ " (ossia, il nome della formula  $\alpha$ ) con il termine " $\alpha \rightarrow \beta$ " (il nome della formula  $\alpha \rightarrow \beta$ ). Nel caso che si adotti una tecnica di gödelizzazione, tali strumenti sono garantiti dalla presenza degli assiomi dell'aritmetica (e dalla conseguente possibilità di rappresentare le funzioni ricorsive nella teoria).

$P_{WS5}$ , a differenza di  $P_T$ , non è inconsistente, ma, per ogni formula  $\beta$ , si può derivare  $bel("β")$ : essa è, per così dire, inconsistente "dentro" i contesti di credenza. E, ovviamente, diventa inconsistente non appena si aggiunga una premessa del tipo  $\neg bel("β")$ . Come si può constatare, (iii') e (iv') sono gli equivalenti nel linguaggio di  $P_{WS5}$  rispettivamente degli assiomi di introspezione positiva e negativa (si noti tuttavia che, per ottenere il risultato sopra enunciato, l'introspezione negativa è un requisito eccessivo; è sufficiente che sia derivabile lo schema seguente, più debole: (v')  $bel("bel("α") \rightarrow α")$ ).

I linguaggi modali, proposizionali e del primo ordine, non presentano questo tipo di problemi, e ciò è dovuto al fatto che il loro potere espressivo è più limitato. E' possibile dimostrare infatti che neppure le formalizzazioni predicative al primo ordine degli atteggiamenti proposizionali danno luogo a inconsistenze qualora il loro potere espressivo venga ridotto a quello dei linguaggi modali (des Rivieres e Levesque 1986). Varie ricerche sono state effettuate per individuare teorie per gli atteggiamenti proposizionali formulate al primo ordine estendendone l'espressività al di là di quella dei linguaggi modali senza tuttavia presentare problemi di inconsistenza. Poiché questi temi hanno vaste implicazioni teoriche per l'intelligenza artificiale e per la semantica cognitiva, torneremo su questo punto in sede di conclusioni generali. Il punto che qui ci interessa tuttavia è che l'impostazione di Moore e Hendrix, ripresa da Konolige, ha a che fare con il problema di *come interpretare* una teoria degli atteggiamenti proposizionali, e non del tipo di linguaggio adeguato per esprimerla (che è il problema posto dalla contrapposizione fra linguaggi modali e linguaggi del primo ordine non modali). Le due istanze, quella che verte sul tipo di linguaggio da adottare e quella che verte su come interpretare gli enunciati di atteggiamento proposizionale, sono fra loro ortogonali (Konolige 1984, McArthur 1988). Sebbene vi sia una sorta di "affinità storica" fra linguaggi modali e semantica a mondi possibili da un lato, e fra semantica basata su enunciati e linguaggi del primo ordine non modali dall'altro, tuttavia si può elaborare una semantica basata su insiemi di enunciati per un linguaggio modale, così come una semantica a mondi possibili può essere associata a un linguaggio predicativo non modale. Konolige (1984) cita esempi per ciascuna di queste possibilità. In un lavoro precedente Konolige aveva proposto un sistema basato su un linguaggio del primo ordine con semantica a mondi possibili<sup>7</sup>, mentre un esempio di linguaggio modale con semantica enunciativa è quello che verrà descritto nel resto di questo capitolo.

## 9.2 Il sistema di Konolige

Passiamo ora ad esporre il sistema di Konolige per la logica della credenza. Prima di iniziare è necessaria qualche premessa sulla formulazione utilizzata. Konolige formula la sua logica in un linguaggio modale del primo ordine (escludendo in un primo tempo la quantificazione all'interno dei contesti epistemic, ed estendendo in seguito la logica in maniera da affrontare il problema del *quantifying in* - Konolige 1984, 1986a). Per la teoria della dimostrazione, Konolige non utilizza una formulazione in termini hilbertiani, bensì adotta una variante del calcolo dei sequenti di Gentzen, ossia il sistema dei *block tableau* sviluppato da Beth e da Hintikka (si veda a questo proposito Smullyan 1968). Qui, per uniformità con la trattazione degli altri sistemi presentati, ci limiteremo al caso proposizionale, e presenteremo il sistema di Konolige in maniera più omogenea possibile ad una formulazione di tipo hilbertiano.

Konolige parte dall'assunto che le credenze di un agente possano essere descritte ipotizzando un *sottosistema delle credenze* (*belief subsystem*), vale a dire, un insieme di enunciati formulati in un linguaggio interno, che rappresentano le credenze di base dell'agente, su cui sono definiti dei meccanismi che consentano di derivare nuove credenze a partire dalle credenze di base. Si assume inoltre che come linguaggio interno venga utilizzato un linguaggio di tipo logico, e che i meccanismi inferenziali siano formulabili in termini di regole logiche. Tale modello, che Konolige chiama *deduction model*, viene proposto come un modello astratto delle capacità effettive di agenti epistemic reali (siano essi naturali o artificiali). In particolare, non vengono presi in considerazione tutta una serie di aspetti quali la distinzione fra memoria a lungo e breve termine, il ragionamento su informazioni incerte, il dimenticare, la revisione delle credenze in presenza di informazioni inconsistenti. Vengono presi in considerazione esclusivamente quelli che, secondo Konolige, sono i due aspetti più importanti del credere e del ragionamento di senso comune: il fatto che la gente può trarre conseguenze da ciò che crede, e il fatto che non ne può trarre tutte le conseguenze logicamente possibili.

Formalmente, nella logica di Konolige i *belief subsystem* vengono modellati per mezzo di *strutture di deduzione* (*deduction structure*). La *deduction structure* associata a un soggetto di credenza  $i$  è una coppia  $d(i) = \langle b(i), \rho(i) \rangle$ , dove  $b(i)$  è un insieme di formule di qualche linguaggio logico  $L$ , le quali rappresentano le credenze di base del soggetto epistemic, e  $\rho(i)$  è l'apparato deduttivo associato al soggetto epistemic. In linea di principio, *deduction structure* diverse potrebbero essere dotate di linguaggi diversi. Nel seguito, tuttavia, assumeremo che il linguaggio sia lo stesso per tutte le *deduction structure*. In una formulazione nello stile della deduzione naturale l'insieme  $\rho(i)$  sarebbe costituito esclusivamente da regole in senso stretto; se invece, come nel nostro caso, si adotta una formulazione di tipo hilbertiano,

<sup>7</sup>A questo proposito egli cita un suo lavoro inedito dal titolo "Modal logics for belief".

$\rho(i)$  deve comprendere, oltre alle regole di inferenza vere e proprie, anche un certo numero di assiomi logici. Tuttavia, poiché gli assiomi logici possono essere considerati regole a zero premesse, assumeremo, per semplicità, che anche nel nostro caso tutti i membri di  $\rho(i)$  siano regole.

Konolige impone una serie di condizioni che devono essere soddisfatte dalle *deduction structure*. Una è la cosiddetta *proprietà di chiusura (closure property)*, in base alla quale l'insieme delle credenze di ogni soggetto epistemico  $i$  deve essere chiuso rispetto alle deduzioni consentite da  $\rho(i)$ . Data una *deduction structure*  $\langle b(i), \rho(i) \rangle$ , sia  $Bel(\langle b(i), \rho(i) \rangle)$  l'insieme delle credenze associate a  $\langle b(i), \rho(i) \rangle$ . Indichiamo con  $\Gamma \vdash_{\rho(i)} \alpha$  il fatto che la formula  $\alpha$  è derivabile dall'insieme di formule  $\Gamma$  per mezzo delle regole  $\rho(i)$ . Allora, in base alla proprietà di chiusura, l'insieme delle credenze di  $i$  può essere definito come segue:

$$Bel(\langle b(i), \rho(i) \rangle) = \{ \alpha \mid b(i) \vdash_{\rho(i)} \alpha \},$$

vale a dire, l'insieme delle credenze associate a  $\langle b(i), \rho(i) \rangle$  è l'insieme di formule  $\alpha$  che sono derivabili a partire dalle credenze di base  $b(i)$  per mezzo delle regole di  $\rho(i)$ .

La richiesta di chiusura potrebbe sembrare di primo acchito in conflitto con l'esigenza di modellare le prestazioni di soggetti epistemici che non siano logicamente onniscienti. Tuttavia il problema non si pone in quanto una delle caratteristiche centrali delle *deduction structure* consiste nell'ammettere la possibilità che l'insieme  $\rho(i)$  di regole sia logicamente incompleto. In questo modo la chiusura deduttiva richiesta da Konolige non coincide con la chiusura rispetto alla conseguenza logica. La prima è una proprietà sintattica del processo di derivazione, mentre la seconda è una proprietà di tipo semantico, che impone che ogni conseguenza logica delle formule dell'insieme sia a sua volta membro dell'insieme. Se le regole in  $\rho(i)$  non sono logicamente complete, allora un insieme di formule può essere deduttivamente chiuso senza per questo essere chiuso rispetto alla conseguenza logica. Dal punto di vista della definizione del modello, risulta più conveniente utilizzare un apparato deduttivo incompleto e richiedere la chiusura deduttiva, piuttosto che rinunciare a quest'ultima proprietà mantenendo completo l'insieme delle regole di inferenza. Un modello in cui si dovesse tenere conto di complesse strategie di controllo della derivazione risulterebbe estremamente complicato ed inelegante. D'altra parte, l'impiego di regole incomplete consente di prendere in considerazione gran parte dei fenomeni rilevanti di incompletezza logica. Due tipi di incompletezza logica delle credenze sono individuati da Konolige come imputabili rispettivamente a ignoranza di alcune regole di inferenza oppure a limiti nelle risorse computazionali dei soggetti epistemici. Come esempio del primo tipo Konolige cita uno studente che non è in grado di risolvere un'equazione del tipo  $x+a = b$  semplicemente perché non conosce la regola che consente di sottrarre la stessa grandezza da entrambi i termini di un'eguaglianza. Un altro esempio di questo genere potrebbe essere individuato nella difficoltà, studiata sperimentalmente dagli psicologi (Wason e Johnson-Laird 1972; Johnson-Laird 1983), che molti soggetti presentano nell'applicare la regola di contrapposizione. Come esempio del secondo tipo si può citare il gioco degli scacchi: ogni giocatore è perfettamente in grado di effettuare ogni singola inferenza che gli consentirebbe, in linea di principio, di esaminare tutti i possibili esiti di una partita, tuttavia il numero di tali inferenze è talmente grande da eccedere le possibilità computazionali di ogni soggetto razionale sottoposto a vincoli di finitezza realistici.

E' chiaro che è possibile modellare l'ignoranza di regole logiche mediante insiemi di credenze deduttivamente chiusi: è sufficiente che le regole che il soggetto non conosce non siano disponibili, né in forma esplicita, né come regole derivate, nella *deduction structure* che ne modella le credenze. Anche nel secondo caso, quello di incompletezza logica dovuta a limiti di risorse, ci si può spesso ricondurre a insiemi di credenze chiusi rispetto a un insieme di regole di inferenza parziali. Si consideri il caso in cui si voglia imporre un vincolo locale sulle risorse disponibili per ogni singola derivazione. In generale, è sempre possibile ricondurre casi di questo genere a processi di derivazione che godano della proprietà di chiusura. Konolige fa l'esempio di un sistema che disponga del *modus ponens* come regola di inferenza, e che sia sottoposto a un vincolo in base al quale in ogni derivazione il *modus ponens* non possa essere applicato più di un certo numero  $k$  di volte. Per modellare questa situazione con una *deduction structure* che goda della proprietà di chiusura, è possibile introdurre una regola analoga al *modus ponens* tradizionale, cui tuttavia sia associato una sorta di "contatore" che viene incrementato ad ogni applicazione della regola stessa, in maniera da controllare che la regola non venga applicata più di  $k$  volte.

Benché le regole di  $\rho(i)$  possano essere incomplete, Konolige assume tuttavia che esse siano logicamente corrette rispetto alla semantica del linguaggio. Egli assume infatti che al linguaggio di ogni *deduction structure*, che, come abbiamo visto, deve essere un linguaggio di tipo logico, sia possibile associare una semantica di tipo modellistico ben definita. Questo affinché sia possibile, ad esempio, stabilire quando le credenze di un soggetto sono vere rispetto ad un certo dominio, o per poter imporre che gli agenti siano razionali, nel senso, appunto, che le regole di inferenza che utilizzano siano corrette. Tuttavia, sottolinea Konolige, deve essere ben chiaro che l'interpretazione semantica non è un processo che avviene "nella testa" del soggetto epistemico. Essa fa parte della metateoria utilizzata da un osservatore esterno per analizzare la natura delle credenze di un agente. Dal punto di vista dell'agente, un *belief subsystem* non è nulla di più di un insieme di formule e di processi computazionali definiti su di esse. Inoltre viene richiesto che, in ogni *deduction structure*, le regole in  $\rho(i)$  siano *effettive*, nel senso che deve sempre esistere una procedura algoritmica per ottenere la conclusione a partire dalle premesse. Esse devono inoltre godere della proprietà di "provincialità"

(*provinciality*): il numero delle premesse di ogni regola deve sempre essere fissato e finito. Questo garantisce che ogni inferenza effettuata a partire da un certo insieme di credenze valga a prescindere dall'insieme globale delle credenze nel *belief set*.

Un'ulteriore proprietà viene imposta da Konolige sulle *deduction structure*, al fine di rendere conto delle credenze iterate. Si tratta della cosiddetta *proprietà di ricorsione (recursion property)*, in base alla quale ogni soggetto di credenza ragiona sulle credenze di altri agenti (comprese le proprie) associandogli a loro volta altre *deduction structure*, che costituiscono un modello di come il soggetto stesso si rappresenta i sistemi di credenze degli altri agenti (o le sue proprie). Così, ad esempio, se Tizio crede che Caio creda che  $\alpha$ , Tizio assocerà a Caio una *deduction structure* in cui sia derivabile  $\alpha$ . Ovviamente, tale *deduction structure* non è quella che rappresenta nel modello le credenze di Caio, bensì quella che rappresenta il *punto di vista* di Tizio sulle credenze di Caio. Così, se  $d(\text{Caio})$  è la *deduction structure* che rappresenta le credenze di Caio, indicheremo con  $d(\text{Tizio}, \text{Caio})$  la *deduction structure* che rappresenta ciò che Tizio crede che creda Caio.  $d(\text{Tizio}, \text{Caio})$ , come ogni *deduction structure*, comprenderà un insieme  $b(\text{Tizio}, \text{Caio})$  di credenze di base, e un insieme  $\rho(\text{Tizio}, \text{Caio})$  di regole di inferenza (di norma, logicamente incomplete). Tali insiemi costituiscono rispettivamente l'insieme delle credenze di base e l'insieme delle regole di inferenza che Tizio attribuisce a Caio. In generale, si avrà che sia  $b(\text{Tizio}, \text{Caio})$  che  $\rho(\text{Tizio}, \text{Caio})$  saranno diversi da  $b(\text{Caio})$  e  $\rho(\text{Caio})$ . Questo procedimento può essere iterato ricorsivamente per rendere conto di qualsiasi livello arbitrario di annidamento di operatori epistemiche. Quindi, ad esempio,  $d(\text{Tizio}, \text{Caio}, \text{Sempronio})$  sarà la *deduction structure* che rappresenta ciò che Tizio crede che Caio creda a proposito di ciò che crede Sempronio. Chiameremo *punto di vista (view)* una sequenza di nomi di soggetti epistemiche che corrisponde alla descrizione di un sistema di credenze ad un qualsiasi livello di iterazione. Così,  $v' = \text{Tizio}, \text{Caio}, \text{Sempronio}$ ,  $v'' = \text{Sempronio}, \text{Caio}$  sono esempi di punti di vista. Per ogni punto di vista  $v$ ,  $d(v)$  sarà la *deduction structure* associata a  $v$ . Considereremo i singoli soggetti epistemiche come casi limite di punti di vista. Così, ad esempio,  $v''' = i$  (dove  $i$  è un soggetto epistemico) è il punto di vista "oggettivo" sulle credenze di  $i$ . Per ogni punto di vista  $v$ , nella *deduction structure* che gli è associata si potrà derivare che il soggetto  $i$  crede che  $\alpha$  se  $\alpha$  è derivabile nella *deduction structure*  $d(v, i)$ . Per ogni punto di vista  $v$ ,  $d(v)$  deve ottemperare alle condizioni sopra formulate per le *deduction structure* (proprietà di chiusura, decidibilità delle regole di inferenza in  $\rho(v)$ , e così via). Casi particolari di punti di vista sono quelli di tipo  $v = i, i$  (dove  $i$  è un soggetto epistemico), che rappresentano i punti di vista dei vari soggetti sui loro stessi sistemi di credenze. Ad esempio,  $d(\text{Tizio}, \text{Tizio})$  è la *deduction structure* che corrisponde a ciò che Tizio crede a proposito del suo stesso sistema di credenze.

Partendo dal concetto di *deduction structure*, Konolige definisce una classe di logiche della credenza in cui una *deduction structure*  $d(i)$  viene associata a ciascun soggetto epistemico  $i$ , e, ricorsivamente, una *deduction structure*  $d(v)$  viene associata a ciascun punto di vista  $v$ . Poiché Konolige prende in considerazione il caso con un numero  $n$  di soggetti epistemiche, il linguaggio utilizzato sarà un linguaggio modale  $\mathbf{L}_n$  dotato di  $n$  operatori modali di credenza  $B_i$  (con  $1 \leq i \leq n$ ), uno per ciascun agente epistemico  $i$ . Per semplicità, assumeremo che il linguaggio utilizzato in tutte le *deduction structure*  $d(v)$  sia sempre  $\mathbf{L}_n$  (in linea di principio tale assunzione non è necessaria, e ogni agente potrebbe utilizzare un linguaggio "interno" differente, ma questo complicherrebbe notevolmente la notazione). Come nel capitolo precedente utilizzeremo esclusivamente i connettivi  $\neg$  e  $\wedge$ , assumendo che gli altri operatori verofunzionali vengano introdotti mediante definizioni. Si noti che assumere che valga l'interdefinibilità dei connettivi nei linguaggi "interni" delle *deduction structure* è arbitrario. Anche questa assunzione tuttavia è dovuta esclusivamente a comodità espositiva, e può facilmente essere evitata.

Semanticamente, la parte non modale del linguaggio viene interpretata mediante le usuali tecniche di teoria dei modelli, mentre, per quanto riguarda le formule modali, per ogni soggetto  $i$ , una formula del tipo  $B_i \alpha$  viene considerata vera nel caso che  $\alpha$  sia deducibile in  $d(i)$ . Quindi, una interpretazione  $M$  per una logica delle *deduction structure* è una coppia  $M = (\varphi, D)$ , dove  $\varphi$  è una funzione interpretazione che assegna valori di verità alle formule del linguaggio, e  $D$  è un insieme di *deduction structure*  $d(v)$ , una per ciascun punto di vista  $v$ . La relazione  $\models$  è così definita:

$$\begin{aligned} M \models p & \text{ (per } p \text{ atomica) se e solo se } \varphi[p] = v; \\ M \models \neg \alpha & \text{ se e soltanto se } M \not\models \alpha; \\ M \models \alpha \wedge \beta & \text{ se e soltanto se } M \models \alpha \text{ e } M \models \beta; \\ M \models B_i \alpha & \text{ se e soltanto se } \alpha \in \text{Bel}(d(i)) \text{ (dove } d(i) \in D). \end{aligned}$$

Come usuale, una formula  $\alpha$  si dice *valida* se e soltanto se, per ogni interpretazione  $M$ , si ha che  $M \models \alpha$ . Una formula  $\alpha$  è *soddisfacibile* se esiste un'interpretazione  $M$  per cui  $M \models \alpha$ .

Si noti che le prime tre clausole della definizione di  $\models$  sono del tutto analoghe alle corrispondenti clausole per la logica proposizionale classica. Per quanto riguarda la verità delle formule modali, come si è detto essa è definita sulla base degli insiemi di credenza associati alle *deduction structure* dei vari soggetti epistemiche. Tali insiemi di credenze sono appunto oggetti sintattici, che Konolige inserisce nei modelli della sua logica a fianco agli usuali costrutti di tipo insiemistico. Un soggetto epistemico  $i$  crede che  $\alpha$  (cioè, è vera la formula  $B_i \alpha$ ) se e soltanto se  $\alpha$  è derivabile dall'insieme  $b(i)$  delle formule che rappresentano le credenze di base di  $i$  per mezzo delle regole di inferenza  $\rho(i)$  associate alla sua *deduction structure*. Quindi,  $i$  è logicamente onnisciente se e solo se le regole di  $\rho(i)$  sono logicamente

complete. E' evidente inoltre che, in generale, se un soggetto  $i$  crede che  $\alpha$  e  $\alpha$  è logicamente equivalente a  $\beta$ , non è detto che  $i$  creda anche  $\beta$ : nel modello  $\alpha$  e  $\beta$ , in quanto oggetti di credenza, sono formule, e quindi entità sintattiche distinte.

Per quanto riguarda la teoria della dimostrazione, le logiche delle *deduction structure* possono essere assiomatizzate mediante un sistema formale **BK** caratterizzato dal seguente apparato deduttivo:

assiomi:

- assiomi del calcolo proposizionale

regole di inferenza:

- da  $\alpha$  e da  $\alpha \rightarrow \beta$  segue  $\beta$  (*modus ponens*)

- da  $B_i\alpha_1 \dots B_i\alpha_k$  e da  $\alpha_1 \dots \alpha_k \vdash_{\rho(i)} \beta$  segue  $B_i\beta$  (*Regola di collegamento RC*)

- un insieme di regole  $\rho(v)$  per ciascun punto di vista  $v$  (fra cui un insieme di regole  $\rho(i)$  per ciascun soggetto di credenza  $i$ )

Le inferenze che coinvolgono formule nell'ambito degli operatori modali di credenza vengono effettuate per mezzo della *regola di collegamento procedurale RC* (*procedural attachment rule*), che consente di effettuare inferenze in **BK** "collegandosi" alle strutture sintattiche comprese nel modello. RC fa infatti riferimento esplicito alle relazioni di derivazione  $\vdash_{\rho(i)}$  comprese nelle varie *deduction structure* associate ai vari soggetti epistemicici. In base a RC, date le formule  $B_i\alpha_1 \dots B_i\alpha_k$ , se da  $\alpha_1 \dots \alpha_k$  segue  $\beta$  sulla base delle regole  $\rho(i)$  della *deduction structure*  $d(i)$ , allora si può inferire che il soggetto  $i$ -esimo crede che  $\beta$  (è derivabile cioè  $B_i\beta$ ).

Il meccanismo delle regole di collegamento procedurale è anche ciò che permette di dedurre formule con modalità iterate in **BK**, consentendo di mettere fra loro in relazione le *deduction structure* corrispondenti ai diversi punti di vista. Si è detto che il linguaggio "interno" associato alle *deduction structure* relative ai vari soggetti epistemicici è lo stesso linguaggio modale  $\mathbf{L}_n$  della teoria. Per come è definito il modello, la possibilità di derivare formule con modalità iterate è legata alla possibilità di dedurre formule modali nelle *deduction structure* associate ai soggetti di credenza. Ciò è garantito da regole di collegamento interne alle *deduction structure* relative ai vari punti di vista. Dato un punto di vista  $v' = v, i$  (dove  $v$  è a sua volta un punto di vista e  $i$  è un soggetto epistemicico), le formule derivabili in  $d(v, i)$  devono poter influire sulle formule modali derivabili in  $d(v)$ . Così, per ogni punto di vista  $v$ ,  $\rho(v)$  comprenderà la regola di collegamento relativa ad ogni soggetto "visibile" da  $v$ , vale a dire la regola:

$$\text{da } \vdash_{\rho(v)} B_i\alpha_1 \dots \vdash_{\rho(v)} B_i\alpha_k \text{ e da } \alpha_1 \dots \alpha_k \vdash_{\rho(v, i)} \beta \text{ segue } \vdash_{\rho(v)} B_i\beta$$

per ogni soggetto epistemicico  $i$  "collegato" al punto di vista  $v$ .

Si può dimostrare che **BK** è corretto e completo rispetto alla semantica delle *deduction structure* sopra definita.

Si è visto che, nella logica di Konolige, ogni soggetto epistemicico  $i$  è logicamente onnisciente se e soltanto se le regole comprese in  $\rho(i)$  sono classicamente complete. Generalizzando questa condizione ad ogni punto di vista  $v$ , si può dimostrare che, come caso limite, la logica **BK** equivale alla logica epistemicica classica, basata sulla semantica a mondi possibili. Diremo che una logica è *satura* se, per ogni punto di vista  $v$ , le regole associate a  $v$  sono proposizionalmente complete e se comprendono una regola di collegamento definita per ogni soggetto  $i$ . Chiameremo **BK<sub>S</sub>** la versione satura del sistema **BK**. E' possibile dimostrare l'equivalenza fra **BK<sub>S</sub>** e **K<sub>n</sub>** (il sistema **K** ad  $n$  soggetti epistemicici). Vale cioè il seguente risultato:

*Teorema* (Konolige 1984, 1986a): per ogni formula  $\alpha$ ,

$$\vdash_{\mathbf{K}_n} \alpha \text{ se e soltanto se } \vdash_{\mathbf{BK}_S} \alpha.$$

Varie proprietà degli operatori di credenza possono essere ottenute imponendo ulteriori vincoli sulle *deduction structure*. Ad esempio, affinché valga il principio di introspezione positiva:

$$B_i\alpha \rightarrow B_i B_i\alpha,$$

deve valere che, per ogni soggetto  $i$  e per ogni formula  $\alpha$ :

$$\alpha \in Bel(d(i)) \Rightarrow B_i\alpha \in Bel(d(i)).$$

In questo caso, secondo la terminologia di Konolige, la *deduction structure*  $d(i)$  gode della proprietà di *completamento positivo* (*positive fulfillment*). Affinché valga il principio di introspezione negativa:

$$\neg B_i \alpha \rightarrow B_i \neg B_i \alpha,$$

deve valere che, per ogni soggetto  $i$  e per ogni formula  $\alpha$ :

$$\alpha \notin Bel(d(i)) \Rightarrow \neg B_i \alpha \in Bel(d(i)).$$

In tal caso  $d(i)$  gode della proprietà di *completamento negativo* (*negative fulfillment*). Per un trattamento approfondito dei principi di introspezione nelle logiche delle *deduction structure* e dei problemi che essi pongono si vedano (Konolige 1984, 1985b, 1986a).

Il modello di Konolige, come d'altronde tutti i modelli sintattici della credenza, ha suscitato molte perplessità in ambito logico. Levesque (1984a, 1984b) ad esempio, sostiene che i modelli sintattici presentano il difetto opposto rispetto alla semantica dei mondi possibili: se quella è di grana troppo grossolana per modellare gli atteggiamenti proposizionali, questi, considerando ogni enunciato un'entità semantica distinta, ammettono distinzioni di grana troppo fine. Levesque sceglie come esempio la disgiunzione di due formule  $\alpha$  e  $\beta$ . Dal punto di vista intuitivo, ai fini del ragionamento epistemico l'ordinamento dei due disgiunti sembra essere del tutto irrilevante: non c'è ragione di pensare che si possa credere " $\alpha$  oppure  $\beta$ " senza credere anche " $\beta$  oppure  $\alpha$ ". Tuttavia,  $\alpha \vee \beta$  e  $\beta \vee \alpha$  sono formule sintatticamente distinte, e perciò, in linea di principio, una di esse può comparire in un *belief set* senza che vi compaia anche l'altra. Che in una disgiunzione i due disgiunti debbano apparire in un certo ordine è un fatto puramente accidentale, dovuto esclusivamente alla natura della notazione. Tuttavia, "l'approccio sintattico - dice Levesque - rende l'ordinamento da sinistra a destra dei disgiunti *semanticamente* significativo, in quanto possiamo credere un ordinamento senza credere l'altro" (Levesque 1984b, p. 5). Ovviamente si può risolvere il problema ponendo restrizioni sulla struttura dei *belief set*. Ad esempio, nel modello di Konolige, si può imporre che ogni  $\rho(v)$  comprenda la regola:

$$\text{da } \vdash_{\rho(v)} \alpha \vee \beta \text{ segue } \vdash_{\rho(v)} \beta \vee \alpha.$$

Tuttavia si tratta di una soluzione *ad hoc*, che, secondo Levesque, risulta del tutto immotivata dal punto di vista semantico. Quali altre condizioni di questo tipo dovrebbero essere imposte sui *belief set*? Si potrebbe richiedere, continua Levesque, che, se in un *belief set* è compresa la formula  $\neg\neg\alpha$ , allora vi debba essere compresa anche la formula  $\alpha$ ; che se vi sono due formule  $\alpha$  e  $\beta$ , allora vi debba essere anche la loro congiunzione  $\alpha \wedge \beta$ , che tutte le tautologie "ovvie", come ad esempio  $\alpha \rightarrow \alpha$ , siano comprese in ogni *belief set* (con tutti i problemi che derivano dallo stabilire quali tautologie siano "ovvie"). Il modello non dice assolutamente nulla su quali di queste scelte effettuare.

Il problema principale è che la "semantica" di Konolige si può definire tale solo in un senso estremamente debole: il significato di un sottoinsieme del linguaggio (quello più problematico, cioè le formule che compaiono nell'ambito degli operatori modali di credenza) viene definito semplicemente mappando le strutture del linguaggio su loro stesse. In questo modo tutto si appiattisce su di un livello sintattico indistinto. Il vincolo che pone Konolige, in base al quale il linguaggio delle *deduction structure* deve essere a sua volta un linguaggio logico cui sia possibile associare una semantica di tipo modellistico, resta lettera morta, rimane cioè del tutto estrinseco rispetto a come la semantica della logica viene definita. Così, anche la richiesta che le regole di inferenza di ogni *deduction structure* siano logicamente corrette risulta del tutto "immotivata" ed ininfluenza dal punto di vista di come i modelli sono definiti. Anche lo stesso meccanismo dei punti di vista e delle regole di collegamento procedurale che connettono i vari punti di vista fra loro, introdotto per giustificare le modalità iterate, resta completamente "opaco" per la semantica del sistema, rimane cioè un puro meccanismo formale, in quanto nel modello tutto quanto si trovi nell'ambito di un operatore modale si riduce ad un piano puramente sintattico. La semantica del sistema resterebbe esattamente la stessa qualunque cosa ci fosse all'interno degli insiemi di credenza  $Bel(d(i))$  associati ai soggetti epistemici.

Nei paragrafi seguenti esamineremo alcuni tentativi per superare, almeno in parte, questo genere di problemi. Nel prossimo paragrafo, in particolare, prenderemo in considerazione la proposta di Levesque di fornire una logica per il ragionamento epistemico limitato dotata di "una semantica che (come quella dei mondi possibili) sia basata su qualche concetto di *verità* piuttosto che su di una collezione di restrizioni *ad hoc* su insiemi di enunciati" (Levesque 1984b, p.5).