

MODELLI MENTALI: UN'ANALISI EPISTEMOLOGICA

di Alberto Binazzi

“The limits of our models are the limits of our world”.

P.N. Johnson-Laird

Abstract

The aim of this paper is to investigate the concept of ‘mental model’ from an epistemological point of view. Since mental models theory accounts for many knowledge’s domains (from semantic competence to probabilistic reasoning) we will focus on deductive reasoning and particularly on syllogistic reasoning and propositional reasoning. We will argue that a correct comprehension of Johnson-Laird’s theory for deductive reasoning must include a philosophical analysis of the concept of mental model itself: a logical foundation of mental models theory. In order to understand the ontological nature of mental models we will stress the importance of the concept of logical consequence as a property that belongs to the mental models and that relies on the operations and on the structure of models. We will underline that mental models are cognitive structures, which, like theoretical models in logic, are conceived as interpretation I of a set of well-formed formulas such that each member of the set is true *in I*. *Mental* models represent, only what is true in the (mental) models. We will stress that this property, apparently abstract, can impede valid deductions. Besides, we will show that the operations of recursive manipulations of models incorporate (psychologically) the principle of logical consequence, as is described in formal logic. In order to appreciate the theoretical consequences of mental models theory for philosophy of mind as well, we will discuss the relation between mental models and the architecture of mind and we will underline the importance to consider models as embodied structures.

Keywords: mental models, logical consequence, truth, local mental models, models theory, isomorphism, embodied structures.

La teoria dei modelli mentali di Philip Johnson-Laird si applica a tre domini della conoscenza: ragionamento, percezione, linguaggio. È, però, nella sua applicazione al ragionamento deduttivo che la teoria ha dimostrato maggior efficacia esplicativa.

Per comprendere pienamente le implicazioni della teoria dei modelli mentali nella sua applicazione al ragionamento deduttivo, occorre, crediamo, indagare i fondamenti epistemologici del concetto di modello mentale. Concetti come quello di conseguenza logica, verità, validità di una deduzione, non sono soltanto i ben noti principi formalizzati in logica matematica, ma sono, altresì, gli ingredienti principali con cui il concetto stesso di modello mentale prende forma e si struttura.

Crediamo che un’analisi di questo genere possa essere utile sia per chi si occupa di psicologia cognitiva, sia per chi si occupa di filosofia della mente, di logica o di IA.

Gli studiosi della teoria dei modelli mentali hanno dato, finora, molto spazio alle applicazioni pratiche della teoria e ai suoi risultati sperimentali. Indagare, invece, i fondamenti epistemologici del concetto di modello mentale significa analizzare i rapporti che intercorrono tra i modelli come strutture cognitive e i principi logici di verità e validità come sono formalizzati in logica (classica).

Dimostreremo, anche attraverso alcuni esempi, come la teoria per la competenza deduttiva proposta dallo psicologo britannico, si fonda su due principi fondamentali: il principio di verità e il principio semantico di conseguenza logica.

Tali principi, come vedremo, sono costitutivi dei modelli. Il modo migliore di apprezzare la proposta di Johnson-Laird per il ragionamento deduttivo, è mettere in evidenza come la proprietà di conseguenza logica, almeno per i domini finiti (che si ipotizza realizzabile nella mente) diventa una proprietà *derivata* di una data proposizione, nel senso che è dipendente dalla struttura e dalle *procedure* ricorsive che costruiscono e modificano i modelli mentali *di* quella proposizione o *di* un insieme (finito) di enunciati.

I modelli mentali sono un tipo di struttura dati dinamica in grado di rappresentare, attraverso un formato semi-analogico, vari domini di conoscenza (reali, ipotetici, immaginari). Quali sono le particolarità di un modello mentale rispetto ad altre strutture cognitive? Secondo Johnson-Laird (1989) un modello mentale è caratterizzato dalle seguenti proprietà:

- 1) La struttura di un modello corrisponde alla struttura della *situazione* rappresentata nel modello.
- 2) Un modello contiene elementi corrispondenti solo ad entità percepite (perceptible entities) ma, all'occorrenza, può contenere anche elementi corrispondenti a concetti astratti (abstract notions) che dipendono, in modo decisivo, dalle procedure utilizzate per la *manipolazione* dei modelli.
- 3) È un tipo di rappresentazione mentale che non contiene variabili e che può includere soltanto un numero *finito* di occorrenze individuali.
- 4) Un modello è da considerarsi il risultato finale (output) di cinque procedure generali (START, ADD, COMBINE, VERIFY, USE) e due procedure ricorsive (REVISE, REVISE) operanti nei confronti delle occorrenze contenute nel modello. Tali procedure, secondo Johnson-Laird, sono in grado di riprodurre correttamente entità, proprietà e relazioni tra le entità *in* un modello, nonché di incorporare un metodo esaustivo per trarre inferenze deduttive.

Luccio (1998) descrive i modelli mentali come strutture rappresentazionali costituite da occorrenze e relazioni congruenti con i processi e le operazioni incorporabili nei modelli (pp. 25).

Una definizione più formale di modello mentale data da Bara, Bucciarelli e Lombardo (2001) descrive tale struttura come una tripla $\langle T, R, S \rangle$ dove T è un *insieme* non vuoto di tokens (occorrenze) R è un insieme di relazioni che valgono in T e S è un insieme di annotazioni su T (per esempio, il segno \neg che individua una negazione di una data occorrenza).

Recentemente, Johnson-Laird (2007) ha proposto di considerare i modelli come un tipo di struttura cognitiva 3D (un medium rappresentazionale tridimensionale) suscettibile di manipolazione interna.

Prendiamo, per esempio, l'ormai classico caso delle relazioni spaziali. Si consideri questa coppia di premesse (Johnson-Laird, 2006):

Il triangolo è a destra del cerchio.

Il cerchio è a destra del quadrato.

Nel caso della prima premessa, il dispositivo cognitivo costruisce il seguente modello mentale:

(START, ADD):

● ▲

La seconda premessa 'elicita' la costruzione di un secondo modello:

(START, ADD):

■ ●

I due modelli possono, a questo punto, essere integrati in un *unico* modello che conterrà i tre 'tokens', cioè, le tre occorrenze corrispondenti a 'cerchio', 'triangolo', 'quadrato', in una precisa disposizione spaziale.

(COMBINE):

■ ● ▲

Per il teorico dei modelli mentali, tali formati devono essere analogici (iconic as far as possible) e occupare, come gli oggetti dello spazio fisico, uno spazio mentale finito dove possono essere suscettibili di manipolazione geometrica. Inoltre, tali modelli rappresentano lo stretto necessario (le occorrenze precedenti possono essere bidimensionali o tridimensionali, ma questo è superfluo per l'economia della rappresentazione). Di certo, per rappresentare fedelmente la transitività della relazione

precedente, è necessario rappresentare gli elementi allineati in un piano euclideo e non, per esempio, posizionati a formare un triangolo.

Il terzo modello contiene adesso tre elementi e con essi, anche la relazione spaziale tra il quadrato e il triangolo, non esplicitamente *menzionata* nelle premesse. Tale relazione (il triangolo è alla destra del quadrato) può essere letta, estrapolata dal modello, senza premesse aggiuntive che definiscono la transitività della precedente relazione.

In logica, però, una deduzione come quella precedentemente inferita, non sarebbe valida perché la transitività della relazione deve essere esplicitamente aggiunta come premessa della derivazione:

Per ogni x,y,z , se x è alla destra di y e y è alla destra di z , allora x è alla destra di z .

Si noti che da tale 'premessa' aggiuntiva non è consentito accedere al significato della relazione: dal punto di vista della stringa considerata, infatti, non vi è né una destra, né una sinistra. Nella stringa, localmente, non c'è nulla che individui il significato di 'alla destra di'. Tale significato è colto, invece, dal modello mentale precedente, dove la relazione di transitività 'emerge' dalla struttura del modello.

Le due procedure ricorsive possono essere considerate le funzioni più caratteristiche dei modelli: infatti, tali procedure, se applicate al modello precedente (■ ● ▲) cercano di modificare (REVISE) il modello, così da rendere, compatibilmente con le premesse, *falsa* la conclusione provvisoriamente accettata. Se tale modello esiste, allora esso rappresenta un controesempio significativo. A tale modello, si applica, allora, la stessa procedura (REVISE) così da invalidare nuovamente la conclusione e così il processo continua, fino a quando non è più possibile trovare modelli alternativi delle premesse, cioè, nuove disposizioni spaziali delle occorrenze compatibili con la verità delle premesse.

Come sarà chiarito successivamente, secondo i teorici dei modelli mentali, il ragionamento può essere considerato il risultato dei particolari processi mentali implicati nella costruzione, manipolazione, integrazione ricorsiva di tali strutture cognitive.

Un altro elemento di divergenza rispetto alle strutture ricordate è che i modelli mentali sono, per definizione, isomorfici a ciò che rappresentano, nel senso che nel modello viene mantenuta 'la stessa struttura di rapporti' che esiste tra gli elementi del dominio rappresentato (si osservi l'esempio precedente). Strutture rappresentazionali come frames, scripts e schemi, inoltre, non descrivono in che modo le parole si riferiscono al mondo (Johnson-Laird, 1988). La teoria dei modelli mentali assume, invece, che la

comprensione di un enunciato corrisponda ad afferrare le sue condizioni di verità (grasping truth conditions).

Compatibilmente con una concezione internista della verità, Johnson-Laird sostiene che il riferimento di un dato enunciato viene individuato dallo stato di cose immaginato che *rende vera* la situazione descritta dall'enunciato in oggetto: si immagina uno stato di cose (si costruisce un modello mentale dell'enunciato) che corrisponde a "come sarebbe il mondo se l'asserzione venisse considerata vera" (ibidem pp.110).

Lo stato di cose immaginato corrisponde alla sua rappresentazione mentale, mentre il valore di verità dell'asserzione considerata può essere accertato comparando la rappresentazione mentale con lo stato di cose del mondo (percepibile) che essa descrive (ibidem pp.110).

Cosa significa riconoscere che un dato enunciato è vero? Si tratta, crediamo, di effettuare un'*immersione* (in un senso matematico, darne un'interpretazione relativa) del modello mentale dell'enunciato considerato *nel* modello percettivo originario.

L'ipotesi che formuliamo si basa sull'accezione matematica del termine 'immersione'. In matematica, infatti, tale termine indica una relazione tra due strutture tali che una delle due contiene all'interno una 'copia' dell'altra: in questo senso la struttura di un modello percettivo 'contiene' al suo interno una copia del modello mentale, dato che questo è considerato come la versione schematica del modello percettivo originario.

Se, però, è possibile immergere un modello in un altro, data una certa compatibilità strutturale, in che senso possiamo immergere un enunciato in un modello? Facciamo notare che per Johnson-Laird i modelli percettivi costituiscono l'interfaccia privilegiata con il nostro ambiente sensoriale, sono, cioè, rappresentazioni 'compatte' del mondo (Schwamb, 1990). Senza tale 'incorporazione', un dato enunciato risulta privo di significato, inaccessibile alla mente.

Si tratterebbe di costruire e recuperare tali rappresentazioni 'compatte' del mondo, internalizzarle (ecco che diventano mentali) e utilizzarle per ragionare e per comprendere un qualsiasi insieme di enunciati. Tale processo consentirebbe alla mente un accesso naturale alla verità, intesa come meta-proprietà di un enunciato che si 'afferra' solo se è possibile immaginare la struttura percettivo-spaziale di quell'enunciato.

Secondo tale concezione, la verità è una proprietà costitutiva dei modelli, interna, sostanziale (condivisa con la struttura dei nostri corpi, così come dei nostri ambienti di vita). In tal senso, comprendere la verità di un enunciato equivale, cognitivamente, ad ancorare percettivamente l'enunciato. Supponiamo, con Johnson-Laird, che sia così: si

potrebbe dire che a *parti* di modello corrispondono *parti* di mondo proiettati nel modello. Le proprietà geometriche della struttura del vivente (quindi anche dei nostri corpi) saranno mantenute nei modelli mentali: le caratteristiche topologiche di un oggetto dovranno essere mantenute nel modello di tale oggetto. Un modello mentale di un pallone che ruota nel campo visivo, non avrà la struttura di un toro, per intenderci; se, infatti, diamo a qualcuno il compito di disegnare un albero, con molta probabilità, disegnerà una figura che manterrà il principio di continuità nella rappresentazione (mentale): la persona, pur deformando a piacere la figura, utilizzerà rappresentazioni continue della figura rappresentata. È interessante notare come la continuità sia mantenuta anche nella rappresentazione 'mimetica' dell'oggetto che ruota: la mia mano simula una cicloide!

Quando diciamo che un modello mentale (che è una versione più schematica di un'immagine mentale e che può contenere contrassegni simbolici) è *isomorfo* al dominio rappresentato, facciamo alcune ipotesi sul tipo e sul formato di rappresentazioni mentali manipolate.

Geometricamente, tale isomorfismo sarebbe la realizzazione mentale di un'iniezione, nel preciso senso matematico, tra i 'punti' del modello percettivo e i 'punti' del modello *mentale*: in una fotografia, per esempio, dobbiamo poter riconoscere un'immagine rappresentata ed è necessario che siano conservate determinati invarianti strutturali. Punti allineati dell'oggetto devono andare in punti allineati della fotografia, le lunghezze e le dimensioni non saranno alterate eccessivamente, non saranno, cioè, tutte rappresentate in un punto di singolarità e dovrà essere conservata la relazione di vicinanza tra punti: nel modello, come nella fotografia, verrà rispettata la struttura topologica del dominio rappresentato.

Tali assunzioni, precisiamolo, riguardano la natura dei modelli concepiti come strutture cognitive astratte (model theoretic): non sappiamo se ad un livello corticale o sub corticale vi sia un rapporto in scala tra mappe neurali o se esista una corrispondenza punto a punto tra ipotetiche mappe corticali in grado di implementare i processi mentali coinvolti nella genesi di un modello mentale.

Johnson-Laird recupera e sviluppa l'idea introdotta da Wittgenstein e da Peirce di una 'struttura pittografia della forma logica di un enunciato' soggiacente all'enunciato (il quale, aggiungiamo, ne emergerebbe come una figura dallo sfondo). Tale struttura, abbiamo detto, si manifesta nella possibilità logica e quindi psicologica, di poter

immaginare (e quindi interagire con) un modello mentale della situazione descritta dall'enunciato.

Recupera, inoltre, l'idea tarskiana di verità come proprietà relazionale sempre riferita a interpretazioni (modelli) (Mangione, Bozzi 2001).

Se per Tarski la verità è sempre riferita ad interpretazioni insiemistiche, per Johnson-Laird è sempre riferita a specifiche strutture cognitive, i modelli mentali, appunto, che non sono oggetti matematici astratti, ma rappresentazioni mentali originate da input di tipo motorio, percettivo, spaziale, sensoriale e linguistico.

Consideriamo questo esempio. Se diciamo: 'La mela è sul tavolo' (ψ) comprendiamo l'enunciato se siamo in grado di immaginare una situazione che *rende vera* tale asserzione.

-----O-----

Immaginare una situazione che rende vera questa asserzione equivale a 'costruire un modello mentale' dell'enunciato. In questa accezione, la verità non è una proprietà interna all'enunciato, dato che questo (la sua struttura proposizionale) non viene messo in corrispondenza *diretta* con il mondo. Semmai viene messo in corrispondenza con il modello corrispondente, ma soltanto dopo questa fase si può dire che (ψ) è vero.

Facciamo notare che le procedure ricorsive precedentemente illustrate (REVISE) non possono applicarsi alla struttura proposizionale dell'enunciato, ma solo al suo modello mentale.

Se alterassimo, infatti, la struttura superficiale dell'enunciato, come in:

la è tavolo mela sul

non staremmo modificando la struttura di un modello e né saremmo in grado di inferire niente di significativo; sarebbe possibile percepire unicamente l'alterazione sintattica degli elementi grammaticali che compongono l'enunciato o individuare le relazioni spaziali tra gli elementi grammaticali come in 'la è *a sinistra di* mela': ma a questa relazione non corrisponderebbe niente che si trova al di fuori della struttura dell'enunciato. Una caratteristica vitale del linguaggio, invece, è proprio quella di riferirsi ad elementi estrinseci alla propria architettura interna (si esamini il confronto con le rappresentazioni proposizionali).

Se operiamo a livello del modello dell'enunciato come in:

----- ○

Siamo in grado, facilmente, di affermare che 'la mela *non* è più sul tavolo'.

La rappresentazione schematica dei modelli descritti, ripetiamolo, non corrisponde ai modelli così come sono prodotti nei cervelli, né corrisponde all'esperienza cognitiva soggettiva. Ciò che è degno di nota, invece, è che la struttura dei modelli percettivi viene tradotta fedelmente nei corrispondenti modelli mentali, ancorando, così, gli enunciati al mondo (to the models of the *real* world).

Gli esseri umani, infatti, sono in grado di comprendere un dato enunciato per procura (by proxy), perché capaci, potremmo aggiungere, di guardare *nel* linguaggio e individuarvi – recuperare il mondo (e viceversa). Sono, cioè, in grado di percepire attraverso il linguaggio: si potrebbe dire che il linguaggio possieda un paio di occhi (chiasma logico-linguistico). Tale visione, portata all'estremo, consentirebbe di inferire che linguaggio, ragionamento, percezione condividono tutti un livello rappresentazionale comune.

I modelli mentali, lo abbiamo già accennato, si candidano come interfacce cognitive in grado di interpretare domini logici. Crediamo che impostare una riflessione epistemologica nei confronti dei modelli mentali, risulta proficuo anche per chi è interessato alla comprensione del rapporto tra logica e psicologia.

Come mai la logica risulta così ostica a molte persone e perché determinate inferenze risultano più semplici di altre? Cos'è che nei nostri processi mentali fa sì che una forma argomentativa invalida finisca per essere tranquillamente accettata e utilizzata nella pratica quotidiana?

Ci sono delle limitazioni costitutive nella architettura cognitiva della mente che impediscono, del tutto naturalmente, la correttezza di determinati processi mentali? Come è possibile attribuire condizioni di correttezza al nostro stesso 'fenotipo' mentale? Come facciamo a isolare le forme argomentative tipiche della logica dai processi e dalle strutture psicologiche che le implementano fisicamente? È corretto ipotizzare strutture mentali isomorfe a calcoli logici per spiegare i fenomeni mentali implicati nei calcoli logici stessi? Se la scala che conduce verso la logica parte dalla mente, come possiamo, una volta raggiunta la logica, gettare via la scala? In che modo i nostri cervelli sono in grado di implementare processi mentali in accordo o in disaccordo a principi logici? Ipotizzare relazioni logiche tra enti mentali significa concepire, inconsapevolmente, gli enti mentali stessi, siano essi modelli o immagini o altro, come proposizioni. Una relazione come quella di conseguenza logica non è qualcosa che esiste tra i nostri lobi parietali e la percezione

sensoriale centrata sull'ego; ma se la conseguenza logica è una proprietà che si applica (solo) a enunciati, come Frege ci ha indicato, com'è possibile che si possa ragionare per mezzo di strutture cognitive che incorporano tale principio? In che senso possiamo guardare alla struttura logica senza guardare alla forma di quella struttura?

Come vedremo, l'architettura dei modelli è cruciale per ragionare correttamente. È possibile, inoltre, intervenire ricorsivamente su tale architettura interna. La scansione di un modello percettivo o cinematografico, per esempio, può essere naturalmente effettuata da sinistra a destra o da destra a sinistra e le operazioni eseguibili nel modello, come spostare un'occorrenza o eliminarla e sostituirla con un'altra, hanno un 'costo computazionale' misurabile in bit, dipendente dal numero e dal formato dei modelli (Ragni, Knauff, Nebel, 2005). Inoltre, il modo in cui i modelli vengono integrati (eseguendo trasformazioni isometriche) risulta essere fondamentale per leggere dai modelli le informazioni pertinenti (relative alle relazioni logiche tra gli elementi del modello).

La teoria dei modelli mentali nasce per contrastare l'idea che le nostre menti siano dotate di un insieme finito di schemi formali di regole di derivazione che permetterebbero di eseguire inferenze deduttivamente valide, grazie a proprietà logiche insensibili al contesto e innate (Rips, 1994, Braine, o Brien, 1998).

Supponiamo che esista un apparato logico minimale 'naturale' (un modulo fodoriano per la logica): da dove originerebbe la precisa struttura di una forma come $((\psi \rightarrow \phi) \& \psi) \rightarrow \phi$? Dall'innata configurazione delle connessioni neurali o dalla successione temporale di precisi meccanismi biochimici? Dalla struttura logica di derivazioni sintattiche su determinate rappresentazioni mentali? Se qualsiasi schema logico innato esistesse, dovrebbe anche essere, citando Fodor, "incapsulato informazionalmente" e questo, stando alle prestazioni deduttive umane, non si verifica quasi mai.

Un dato schema inferenziale potrebbe applicarsi anche a rappresentazioni mentali che possono, così, venir composte secondo le usuali funzioni di verità. Se assumiamo che i pensieri abbiano un formato proposizionale, simile ad enunciati, e diciamo che la conseguenza logica è un principio che si applica anche ai pensieri e non solo agli enunciati, dovremmo osservare un'ottima competenza deduttiva nei soggetti sperimentali. Se i nostri dispositivi cognitivi fossero dotati di schemi formali operanti su rappresentazioni mentali, allora, nel caso di una deduzione classica come il modus ponens $((\psi \rightarrow \phi) \& \psi) \rightarrow \phi$, il nostro modulo per le deduzioni dovrebbe comprendere una serie di istruzioni come la seguente: se la prima rappresentazione ψ è identica a (sintatticamente unificabile) alla seconda rappresentazione ϕ , e tra ψ e ϕ c'è una 'freccia', allora, scrivi

(deduci formalmente) ϕ . Tali 'istruzioni', completamente automatiche, data la loro natura puramente sintattica, dovrebbero applicarsi a *qualsiasi* insieme di rappresentazioni mentali.

Rips (1994) ha sostenuto che la mente è equipaggiata di un insieme (innato) di regole formali di inferenza. Ma, sostiene Johnson-Laird, il teorico degli schemi formali mentali deve rispondere ad una serie di domande imbarazzanti: che tipo di logica o schemi logici possiede il ragionatore medio? Quali teoremi considerare come schemi? In più, dovremmo legittimamente supporre che esista una qualche logica minimale, un apparato logico completo e decidibile e insensibile al contesto. Ebbene, non sembra che il ragionamento quotidiano e, più in generale, una buona competenza deduttiva, dipendano dall'utilizzo di schemi formali di ragionamento, pronti per l'uso (come la serie di scariche neurali necessarie per la contrazione muscolare) e indipendenti dal contesto.

Consideriamo questo esempio:

Solo una delle seguenti frasi è vera:

Se fra le mie carte c'è un re, allora c'è un asso.

Se fra le mie carte non c'è un re, allora c'è un asso.

Fra le mie carte c'è un re, cosa segue?

Tutti i soggetti sperimentali ai quali è stato posto questo problema, hanno risposto che "c'è un asso tra le mie carte" (Johnson-Laird e Savary, 1999). La risposta sembra corretta, ma si tratta di un'inferenza illusoria: infatti, uno dei due condizionali è falso e un condizionale è falso se il suo antecedente è vero e il suo conseguente è falso. Siccome sappiamo dalle premesse iniziali che uno dei due condizionali è necessariamente falso, deve essere necessariamente vero che 'non c'è un asso tra le mie carte'. Se il comportamento del condizionale materiale fosse incapsulato informazionalmente, sostiene Johnson-Laird, i nostri soggetti sperimentali non dovrebbero compiere inferenze invalide: ma è proprio questo che fanno!

Per la verità, la circolarità dell'argomentazione è evidente già a livello della logica pura: siamo in grado di dedurre validamente perché adoperiamo schemi di assiomi simili a quelli formalizzati (appunto!) nel calcolo proposizionale. È piuttosto sorprendente (e imbarazzante) constatare che sono stati necessari numerosi esperimenti di psicologia cognitiva per sbarazzarsi di questa idea.

Ciò che sembra sfuggire sia ai teorici dei modelli mentali, sia ai sostenitori della logica naturale, come vedremo, è il fatto che uno schema formale come quello precedentemente espresso, non prende forma dal vuoto: tale struttura è formale rispetto ad una costruzione già data, cioè all'interno di un universo semantico *costruito* (Longo, 2002).

L'ipotesi fondamentale della teoria dei modelli mentali postula che il ragionamento deduttivo si basa sul processo di costruzione, integrazione, modifica di modelli mentali semi-analogici di stati di cose (Bara, 1990).

Tale definizione, sosterrò, si può ritenere 'l'implementazione architettonica', del principio di conseguenza logica (per domini finiti) che trova la sua formalizzazione matematica all'interno della teoria dei modelli. Come un punto appartenente ad un cerchio in movimento disegna una cicloide, così i modelli mentali vanno concepiti come oggetti dinamici che disegnano la forma strutturale (cinematica) di un ragionamento.

Il principio di conseguenza logica, strumento dimostrativo di logici e matematici, stabilisce che un dato enunciato ψ deriva da una teoria T se e solo se non c'è un modello di T che non sia *anche* modello di ψ , così per provare che ψ non deriva da T , basta *trovare* un modello di T che non sia modello di ψ , cioè un modello che soddisfi T e $\neg\psi$ (Bozzi, Mangione, 2001).

La procedura corretta per il ragionamento deduttivo offerta da Johnson-Laird, prevede tre tipi di 'operazioni di alto livello':

- 1) costruire un modello mentale provvisorio compatibile con le premesse e formulare una conclusione che sia vera nel modello.
- 2) ricercare un ulteriore modello che sia compatibile con un'interpretazione alternativa delle premesse e formulare un'altra conclusione che sia vera adesso nei due modelli fin qui considerati.
- 3) decidere se accettare questa conclusione come definitiva o cercare di interpretare nuovamente le premesse per verificare la verità delle conclusioni formulate o dedurre la loro falsità rispetto ai modelli fin qui presi in considerazione.

Facciamo notare che trovare un modello che soddisfi T e $\neg\psi$ equivale 'cognitivamente' a cercare un modello mentale alternativo delle premesse e formulare, così, una nuova conclusione vera nel modello. Il processo diventa ricorsivo, perché la ricerca di modelli ulteriori continua (si tenga presente, però, che bastano tre modelli per sovraccaricare la memoria di lavoro).

La teoria dei modelli è considerata come la parte più strettamente matematica della logica matematica:

Nel famoso testo di Chang-Keiser (*Model Theory, North-Holland, Amsterdam-London 1973*) tale disciplina veniva sinteticamente descritta come algebra universale + logica, ovvero un'estensione dello studio generale delle strutture matematiche in cui il linguaggio predicativo sostituisce quello equazionale [...]. Un paragone forse più illuminante, espresso da Hodges in *A Shorter Model Theory*, afferma che la teoria dei modelli è una generalizzazione della geometria algebrica, dove si descrivono proprietà e relazioni geometriche attraverso equazioni e sistemi di equazioni algebriche (su corpi, negli sviluppi più astratti). Analogamente, in teoria dei modelli la struttura di riferimento è una qualsiasi struttura matematica e il linguaggio non è equazionale ma predicativo.

[...] Quello che la logica predicativa ha permesso di fare con la teoria dei modelli, è un passo ulteriore nella linea cartesiana. La generalizzazione della teoria dei modelli è duplice: strutture qualsivoglia *hanno sostituito lo spazio* (corsivo nostro) e il linguaggio predicativo ha sostituito quello algebrico-equazionale.¹

Ora, la costruzione di opportuni modelli per una teoria consente di studiarne la struttura logica: si può dimostrare, ad esempio, che una teoria non è contraddittoria (è soddisfacibile) se possiede almeno un modello o che una data teoria è categorica se tutti i suoi modelli sono isomorfi tra loro (Odifreddi, 2003). Lo studio di operazioni e relazioni strutturali porta, così, al concetto di traduzione tra sistemi simbolici diversi e alla *rappresentazione logica* di relazioni entro modelli e teorie (Manca, 2001).

Consideriamo il quinto postulato di Euclide: come dimostrare che non è conseguenza logica degli altri? Interpretando gli assiomi in modelli alternativi a quello classico. Le geometrie iperboliche di Lobacëvskij e di Bólyai, la geometria ellittica di Riemann, il modello di Klein o di Poincaré costituiscono dei modelli nei quali non vale il postulato delle parallele. Se consideriamo, per esempio, 'piano' una superficie sferica, i suoi punti euclidei come punti del piano, e le circonferenze massime della superficie sferica come 'rette', allora, *in questo modello*, l'assioma delle parallele non è più valido: tutte le circonferenze massime di una superficie sferica si incontrano in due punti diametralmente opposti. Interpretando diversamente i postulati, collocandoci, cioè, al di fuori del modello euclideo, il quinto postulato non segue più logicamente dagli altri.

Si consideri il seguente esempio (Manca, 2001): siano dati una segnatura con 4 predicati, P, R entrambi unari e ε , \perp , binario il primo e ternario il secondo: "Essi esprimono le

¹ Manca (2001) pp.152.

proprietà e relazioni geometriche di punto, retta, appartenenza (di un punto ad una retta) e di stare tra due punti (\perp a, b, c se b sta tra a e c)” (pp. 65).

Manca elenca (senza dimostrarle) sei proposizioni che esprimono determinate proprietà geometriche tra gli enti presi in considerazione:

- 1) Per ogni due punti distinti esiste una e una sola retta che passa per entrambi.
- 2) Ogni retta contiene almeno due punti distinti.
- 3) Esistono una retta e un punto fuori di essa.
- 4) La relazione di ‘stare tra’ è simmetrica rispetto al primo e al terzo argomento.
- 5) Dati tre punti distinti allineati, uno di questi è sempre tra gli altri due.
- 6) Data una retta e un punto fuori di essa, vi è sempre una e una sola retta che passa per il punto e che non ha punti in comune con la prima retta.

Come facciamo a dimostrare che l’ultima proposizione *non* è conseguenza logica delle altre?

“Basta prendere come modello i tre vertici di un triangolo, che verificano il predicato P e i tre lati del triangolo che verificano il predicato R, imponendo che ciascun vertice appartenga ai lati di cui è estremo (il predicato di ‘stare tra’ non vale per nessuna terna di punti)” (pp.65):

I lati del triangolo “corrispondono” al predicato R :

P

P Δ P

Tale modello verifica tutte le proposizioni dell’esempio precedente, tranne l’ultima.

Sulla nozione di modello si basano le dimostrazioni di indipendenza: pensiamo all’indipendenza dell’assioma di scelta dagli altri assiomi insiemistici dimostrata da Cohen negli anni quaranta o alle suddette geometrie non euclidee.

Bene, ora pensiamo alle procedure offerte da Johnson-Laird: non è interessante notare che la teoria proposta sembri incorporare la stessa dinamica logica?

Potremmo cogliere il senso della proposta di Johnson-Laird utilizzando la nozione di validità in un modello, così come è formalizzata, appunto, in teoria dei modelli:

$$(M \models \psi \rightarrow \phi) \leftrightarrow (M \models \psi \rightarrow M \models \phi)$$

I modelli mentali, diversamente dai modelli logico-algebrici, incarnano la possibilità di intervenire ricorsivamente nei confronti dei propri processi mentali in atto: si va alla

ricerca di modelli (finiti) compatibili con interpretazioni alternative delle premesse, si valuta la validità delle conclusioni fin qui accettate, si decide se andare alla ricerca di modelli alternativi per mettere alla prova la correttezza delle conclusioni raggiunte. Il processo si svolge ricorsivamente, dato che i modelli incorporano anche le procedure in grado di modificarli e sostituirli.

Consideriamo il seguente sillogismo:

Nessun A è B

Tutti i B sono C

Cosa segue?

Si costruisce un primo modello mentale che realizza le premesse e si formula una conclusione che sia vera nel modello (o meglio, che in base alla disposizione delle occorrenze nel modello, risulta vera, cioè, leggibile dalla situazione immaginata).

Ci fermiamo qui? No, se cerchiamo un altro modello in grado di falsificare la conclusione da noi considerata (che diventa provvisoria). Esiste il modello cercato? Bene, se siamo in grado di disporre le occorrenze in un modo alternativo, allora trarremo una seconda conclusione, che dovrà valere anche nel primo modello costruito per essere considerata valida e così il processo continua:

Ecco i *tre* modelli compatibili con le premesse del sillogismo precedente (Johnson-Laird, 2005).

(START-ADD-COMBINE) → (VERIFY-READ)

[A] ¬ B

[A] ¬ B

[A] [B] C

[B] C

...

(REVISE) → (VERIFY-READ)

[A] ¬ B C

[A] ¬ B

[A] [B] C

[B] C

...

(REVISE) → (VERIFY-READ)

[A] ¬ B C

[A] ¬ B C

[B] C

[B] C

...

Il simbolo “¬” denota il simbolo per la negazione, ciascuna linea denota una determinata relazione logica tra individui, le parentesi quadre indicano che l’insieme degli elementi A e B è stato rappresentato esaustivamente e i tre puntini (...) indicano la possibilità di inserire, a discrezione del solutore, altri tokens che si ipotizza non siano rappresentati esplicitamente nel modello.²

Perché tali strutture si chiamano *modelli* mentali e non immagini o schemi mentali? Semplicemente perché, lo abbiamo già accennato, come i corrispettivi modelli logico-algebrici, i modelli mentali rappresentano soltanto ciò che è *vero*: per definizione, un modello *M* di una teoria *T* contiene soltanto gli elementi di *T* che sono *veri in M*.

Tale principio, osserviamo, impone delle restrizioni operative decisive ai modelli stessi: nel caso dell’applicazione alla logica proposizionale, per esempio, i modelli di una disgiunzione inclusiva, non rappresentano lo stato (o gli stati) di cose in cui la disgiunzione è falsa. Inoltre, “il principio stabilisce che in ogni modello i letterali presenti nelle premesse siano rappresentati solo quando sono veri in un caso possibile” (Johnson-Laird, 2001).

Per esempio, nel caso di una disgiunzione inclusiva:

$(\psi \vee \phi)$

nello stato di cose in cui ψ è vero e ϕ è falso, il modello non conterrà contenuto semi analogico per ϕ . A causa di questa limitazione, non viene conseguentemente rappresentata e percepita internamente l’asserzione “è falso lo stato di cose rappresentato da ϕ ”.

Data, per esempio, la disgiunzione esclusiva precedente, i ragionatori elencano *due* possibilità che corrispondono a *due* modelli mentali:

² A dire il vero, i tre modelli rappresentati sono soltanto l’output dei processi mentali implicati nella deduzione. I modelli vengono generati uno alla volta secondo l’ordine rappresentato. In base a numerosi esperimenti con i sillogismi, è emerso che i soggetti, di norma, costruiscono un primo modello e non prendono in considerazione la possibilità di disposizioni *alternative* delle occorrenze. Chi fosse interessato, invece, al programma per calcolatore sviluppato da Johnson-Laird per la risoluzione dei sillogismi può visitare il sito web: <http://webscript.princeton.edu/~mentmod/models.php/> dove sono presenti, tra l’altro, modelli computazionali (in Lisp) di modelli mentali per la logica proposizionale, il ragionamento modale e il ragionamento temporale.

- 1) ψ
- 2) ϕ

Il primo modello mentale non rappresenta ϕ , che è falso in questo primo modello ($\neg\phi$) e il secondo modello mentale non rappresenta ψ , che è falso in questo secondo modello ($\neg\psi$) (quindi non viene rappresentata internamente l'asserzione che ' ψ è falso'). "Le persone tendono ad ignorare questi casi" (Johnson-Laird 2005).

Gli individui possono però produrre delle note mentali (mental footnotes) dei casi falsi e utilizzare queste note mentali per trasformare (to flesh out) i modelli *mentali* in modelli *espliciti* (fully explicit models).

Per esempio, i modelli mentali del condizionale ($\psi \rightarrow \phi$) sono inizialmente due:

- 1) ψ ϕ
- 2)

dove i puntini denotano un modello implicito (implicito, perché non rappresentato in un primo momento) dei possibili stati di cose in cui l'antecedente del condizionale è *falso*.

Il modello esplicito del condizionale, una volta 'sviluppato' (fleshing out) risulta essere:

- 1) ψ ϕ
- 2) $\neg\psi$ ϕ
- 3) $\neg\psi$ $\neg\phi$

Si consideri questo esempio (Johnson-Laird, 2005):

Ann is in Alaska or Beth is in Barbados, or both.
Beth is in Barbados or Cath is in Canada or both.
What follows?

Nell'esempio considerato, la prima premessa prevede la costruzione di tre modelli, così come la seconda. Il termine 'medio' Beth, permette di integrare i modelli, tramite sovrapposizione *isometrica*, riducendone, così, il numero a tre.

Modelli mentali:

- a) Ann Cath
- b) Beth
- c) Ann Beth Cath

Conclusioni: (((Ann & Cath) v (Beth) v (Ann & Beth & Cath))).

Secondo la teoria dei modelli mentali, un’inferenza deduttiva valida dipende dalla “proporzione di modelli” di volta in volta presi in considerazione dal ragionatore: infatti, ciò che conta è il *numero* di modelli che il ragionatore costruisce e manipola. Questo è un punto controverso: alcuni teorici hanno focalizzato l’attenzione sul numero e sulla complessità di *operazioni* effettuabili nei modelli, o sul tipo di conoscenze (in formato proposizionale o analogico) che possono interferire nello svolgimento corretto della procedura, piuttosto che sul numero di modelli (Ragni, Knauff, Nebel, 2005). Tali studiosi, che hanno proposto un loro modello computazionale basato sui modelli mentali per il ragionamento spaziale, (SRM, Spatial Reasoning by Models) considerano i modelli mentali come ‘mappe spaziali bidimensionali della memoria di lavoro’. Il loro modello computazionale, infatti, dispone le occorrenze in un assetto bidimensionale (2D array) e manipola le occorrenze in modo da ‘inferire’ relazioni di tipo spaziale non espressamente dichiarate nelle premesse.

Altri studiosi si sono soffermati, invece, sulle difficoltà incontrate dai soggetti durante il processo di ‘model editing’, ovvero, sul tipo di operazioni eseguibili all’interno di un modello, in relazione alle informazioni del solutore nei confronti di un dato dominio di conoscenza.

Johnson-Laird ha implementato diversi programmi per calcolatore in cui le conoscenze dichiarative del solutore, incorporate nella struttura dati della memoria, influenzano il corso della deduzione: il fenomeno chiamato ‘modulazione’ (semantic modulation of models).

Si consideri, infatti, tale condizionale:

Se è un gioco, allora non è il calcio (Johnson-Laird, 2005).

La cui rappresentazione ‘modellistica’ (fully explicit) è la seguente:

- gioco – calcio
- gioco – calcio
- gioco calcio

Il significato del termine 'calcio' implica che il calcio sia anche un gioco e sostiene Johnson-Laird, a causa di questo naturale fenomeno, il terzo modello non verrà costruito. Il terzo modello non viene così né immaginato né utilizzato per ragionare: semplicemente 'non esiste' nella mente del solutore.

Secondo Bazzanella, Bouquet e Warglien (2005) i soggetti sperimentali costruiscono modelli piuttosto statici, fissi, non facilmente sostituibili 'in toto' durante il processo di ragionamento (una contraddizione tra modelli dovrebbe generare l'eliminazione dei modelli, ma questo non sembra avvenire 'concretamente'). I soggetti preferiscono, invece, basandosi spesso su un solo modello, modificare *parti* del modello, piuttosto che andare alla ricerca di modelli ulteriori. Un altro elemento divergente con la formulazione di Johnson-Laird è relativo alla confusione, presente nel sistema di credenze dei soggetti, tra soddisfacibilità (verità in almeno un modello) e validità (verità in tutti i modelli considerati).

Altri teorici, mettendo in luce maggiormente gli effetti di natura pragmatica nel ragionamento, evidenziano come i soggetti sono abituati, nella vita quotidiana, a 'interagire' con elementi discorsivi (utterances) piuttosto che con enunciati (sentences). Per questo, non sarebbero in grado di immaginare *sempre* stati di cose alternativi, se tale operazione non viene loro espressamente motivata (Sacco, Bucciarelli, Adenzato, 2001).

Recenti risultati sperimentali (Hagen, C., Knauff, M., Rauch, R., Schlieder, C., Strube, G., 2005) evidenziano come, in realtà, soggetti alle prese con deduzioni di tipo spaziale prendono in considerazione un solo modello e realizzano delle trasformazioni locali del modello considerato. Tale procedura contrasta con la visione classica offerta da Johnson-Laird relativa alla costruzione di modelli alternativi delle premesse 'letti' uno per volta (si osservi i tre modelli corrispondenti alle due premesse del sillogismo precedente: il secondo modello non è una trasformazione del primo, ma è un altro modello, così come il terzo non è una modificazione del secondo).

Johnson-Laird fa spesso esempi in cui sembra considerare il soggetto stesso come *parte* del processo di costruzione – integrazione - modifica dei modelli mentali. Il soggetto dovrebbe decidere di cercare un modello alternativo alle premesse, dovrebbe essere in grado di ripercorrere a ritroso i passi mentali svolti, dovrebbe essere capace di 'guardare' il modello da destra a sinistra, dovrebbe spostare i tokens in modo da invalidare le sue conclusioni provvisorie, per esempio. Benissimo, ma come fa a sapere di poter essere in

grado di realizzare tutta questa ricchezza di operazioni, senza qualcuno che lo istruisca a tal fine?

Insomma, il ragionatore proposto da Johnson-Laird dovrebbe essere capace di effettuare, spontaneamente, una consistente quantità di riflessioni meta-logiche. Come diventa naturale questo processo?

Secondo Yang e Bringsjord (2001) la teoria dei modelli mentali e la teoria della logica naturale (natural logic) non sono altro che le versioni psicologiche, rispettivamente, dell'approccio semantico e sintattico alla logica simbolica. Entrambi i partiti dimenticano di essere, in realtà, due facce della stessa medaglia. È per questo che i due autori sopra citati credono che sia necessaria una terza teoria: la 'meta-logica mentale' (Mental Meta Logic). Tali autori considerano tale progetto come un vero e proprio nuovo settore di ricerca in psicologia cognitiva.

A dire la verità, non c'è bisogno di scomodare teorie cognitive 'apparentemente' risolutive. Già a livello della logica pura la distinzione tra sintassi e semantica è considerata molto labile: le tavole di verità, infatti, vengono *costruite* per mezzo di regole sintattiche codificate dalle tavole dei connettivi e le regole sintattiche per i sequenti sono introdotte in base a considerazioni semantiche legate ai valori di verità (Odifreddi, 2003).

Il problema, sosteniamo, non è nella distinzione tra metodi semantici e sintattici in logica, quanto in una loro 'versione speculare' in psicologia. In logica, infatti, la differenza tra il livello sintattico e il livello semantico è molto meno netta di quanto i due partiti dimostrano di conoscere: quali regole siano sintattiche e quali siano semantiche è una nozione relativa, non sono, cioè, nozioni assolute (Lolli, 2000). In logica, così come in IA, "si lavora con due domini di conoscenza, uno dei quali computa l'altro; il secondo dà la semantica al primo, ma potrebbe essere sintattico rispetto ad un terzo" (pp. 86).

Supponiamo, però, che i teorici di una metalogica mentale abbiano ragione: enunciati metalogici potrebbero essere incorporati in un modello mentale? E nel caso sia possibile, come potrebbe diventare meno ostico per i soggetti incorporare all'interno dei modelli uno specifico contenuto proposizionale in grado di 'orientare' la costruzione di modelli più adeguati? La teoria, cioè, dovrebbe essere dotata di euristiche o criteri di interpretazione in grado di selezionare i modelli pertinenti (Evans, Over 1996). Inoltre, aggiungiamo, una data credenza o è un modello mentale o è una rappresentazione proposizionale. Nel caso sia una rappresentazione proposizionale, rimane da capire come possa essere ancora salvaguardata la natura *analogica* dei processi inferenziali. Per i sostenitori della teoria dei modelli mentali, ricordiamolo, il ragionamento è un'attività cognitiva non soggetta al

controllo sintattico di un ipotetico sottoinsieme 'naturale' di regole come quelle usate nei cosiddetti sistemi di derivazione naturale (Rips, 1994).

Le nostre perplessità sono relative al tipo di ontologia attribuibile ai modelli: si dice che derivano da dati di tipo percettivo, e allo stesso tempo si ammette che sono 'più astratti delle immagini mentali' (Johnson-Laird, 2007). Benissimo, ma come fa un modello mentale, che è un tipo di rappresentazione analogica 3D, a incorporare contrassegni come '-' o '[]'? Il nostro dispositivo computazionale può fare a meno di sapere a cosa tali simboli si riferiscono, ma il soggetto sperimentale *non può*.

È interessante mettere in evidenza come l'attuale dibattito in psicologia cognitiva nei confronti del ragionamento deduttivo rifletta la contrapposizione tra l'impostazione della logica data da Frege, Russel e Whitehead, Post, Hilbert, e la visione della logica del primo Wittgenstein. Secondo Wittgenstein, infatti, in logica, regole di inferenza come il modus ponens, sono da considerarsi inessenziali: "Leggi d'inferenza che come in Frege e Russel giustificano le conclusioni, sono prive di senso e superflue" (5.132). Con le dovute distanze, tale posizione è ancora viva nei sostenitori della teoria dei modelli mentali: infatti, i modelli, come le proposizioni del *Tractatus*, non 'rappresentano' internamente schemi formali d'inferenza. Le relazioni di tipo logico tra date occorrenze 'emergono' dalla struttura dei modelli, così come, per Wittgenstein, la 'derivabilità' di un enunciato non poteva essere letta o inferita dall'enunciato stesso, cioè, dalla sua forma (superficiale) enunciativa. Per Wittgenstein, infatti, le proposizioni di una teoria non potevano essere interpretate per parlare di qualcos'altro, tanto meno per le proprie formule (Lolli, 2004). Tale eredità, poi superata dagli sviluppi della logica matematica moderna con gli importanti risultati limitativi di Tarski, Gödel, Turing, è stata raccolta, con le dovute distanze, ripetiamolo, da Johnson-Laird con i suoi modelli mentali. Nei modelli, infatti, non c'è spazio per regole formali di inferenza: uno schema come il modus ponens non è una regola che guida il succedersi di una proposizione da un insieme di proposizioni. Lo schema modus ponens è da considerarsi, semplicemente, come il risultato dell'immaginabilità e della combinazione di determinati modelli mentali. Immaginiamo ψ , immaginiamo ϕ e 'leggiamo' dalla struttura delle rappresentazioni mentali di (occorrenze mentali di) ψ e ϕ una specifica relazione logica tra le due proposizioni.

In che senso dal modello (e soltanto da questo) possono essere *estrapolate* relazioni logiche?

Ripensiamo ai contrassegni utilizzati per rappresentare i modelli dei sillogismi: operazioni logiche come la quantificazione o la disgiunzione richiedono una qualche

rappresentazione di tipo simbolico, e come ha sostenuto Wilks (1988), anche i modelli mentali non sfuggono alla fallacia simbolica.

Secondo Marco Colombetti (2000) ricercatore presso il Dipartimento di Elettronica e Informazione del Politecnico di Milano, i modelli mentali non possono essere concepiti come strutture completamente analogiche:

Per inferire che due enunciati sono contraddittori, dobbiamo utilizzare assiomi logico-formali o un meccanismo equivalente, dato che l'ancoraggio causale realizzato dall'interfaccia percettiva *non compensa* la natura formale dei simboli (pp. 167).

In logica, i simboli extralogici non hanno un'interpretazione fissa e ammettono infinite interpretazioni (modelli). È necessario, allora, vincolare le possibili interpretazioni di tali simboli e assumere, a seconda dei casi, un bel po' di assiomi. Anche i modelli mentali, sostiene Colombetti, non sono completamente analogici, dato che incorporano contrassegni per la negazione, si uniformano al comportamento delle tavole di verità dei connettivi e a quello dei quantificatori per la sillogistica: per rappresentare correttamente domini logici, è necessario includere assiomi logico-formali nei modelli mentali.

Cosa significa, infatti, rappresentare determinate relazioni logiche in un modello mentale? Immaginiamo una scena (costruiamo un modello) di una particolare disposizione di oggetti in un determinato spazio metrico come in:



Supponiamo di possedere un insieme finito di assiomi incapsulati in un modulo per il ragionamento deduttivo. Che formato avrebbero tali schemi? E, soprattutto, perché dovrebbero avere un formato simbolico? Supponiamo che abbiano un formato proposizionale (D'Mello, S. Grasser, A., Ventura, M., 2005):

Left (▲,■) if Left (▲,●) and Left (●,■).

Left (▲,●) = Right (●,▲).

Right (■,▲) if Right (■,●) and Right (●,▲).

.....

I sostenitori della teoria dei modelli mentali spesso affermano che alcune inferenze, se vi fosse una logica mentale 'naturale', non dovrebbero essere nemmeno *psicologicamente* possibili. Noi tutti, però, nella vita quotidiana, come i soggetti sperimentali, eseguiamo inferenze invalide. Come spiegare questi fenomeni, se nei nostri 'data base' mentali vivessero dei sistemi formali pronti per l'uso? Come garantire la possibilità di ragionare in

modo deduttivamente valido se disponiamo solo di modelli mentali, che, a causa di valori di default, sono il più delle volte, incompleti? I modelli, infatti, non rappresentano tutto lo spazio logico, ma, soltanto un *sottoinsieme* possibile di stati di cose compatibile con un insieme di premesse (Bouquet, 2005).

In uno scenario spaziale come quello precedentemente raffigurato (che si suppone sia neuropsicologicamente localizzabile) siamo in grado di inferire, del tutto naturalmente, relazioni spaziali tra le occorrenze, ma nessuno sano di mente si sognerebbe di dire che, nel ricavare queste conclusioni, noi siamo guidati da assiomi: il fatto è che esistono condizioni naturali (embodied) delle 'restrizioni operative', come ha sostenuto il filosofo Putnam (2003), per le quali alcune inferenze risultano spontanee e altre no (come quelle sillogistiche). Ciò significa, secondo Putnam, che "certi predicati devono applicarsi a certi oggetti ogniqualvolta gli input del computer neurale (vale a dire, gli output dei moduli percettivi) sono di un certo tipo" (2003, pp. 31).

Allo stesso tempo, i modelli mentali, a causa delle limitazioni computazionali del dispositivo inferenziale umano, non sono sempre in grado di implementare conclusioni deduttivamente valide.

Come ha affermato Lolli:

Perché l'autore non conclude che non vi è nessuna cosa come la competenza logica nei suoi soggetti sperimentali? Le persone fanno errori sistematici prima di imparare la nozione di conseguenza, di riconoscere la sua utilità e la necessità di impadronirsene attraverso metodi opportuni. Johnson-Laird ha cercato un modello del comportamento effettivo che potesse estendersi ad un calcolo logico corretto e in ciò ha avuto largo successo; ma l'esperienza fatta con la matematica ci avverte che i più efficienti algoritmi che impariamo e usiamo molto spesso non hanno niente a che fare con le manipolazioni approssimative e sperimentali con le quali incominciamo a trattare i numeri e lo stesso potrebbe essere vero per gli algoritmi logici³.

In lavori recenti, Johnson-Laird ha introdotto all'interno dei modelli mentali varie nozioni logico-matematiche: secondo Vosgerau (2006) risulta molto arduo armonizzare tali nozioni logiche con il formato percettivo dei modelli. Cosa ne sarebbe allora del principio di 'naturalzza' dei modelli? Essi si strutturano a partire da dati percettivi e allo stesso tempo, contengono astratti contrassegni simbolici, come, per esempio, il contrassegno logico della negazione '¬'.

³ Lolli (2000) pp. 97.

Autori come Rips, Wilks, Lolti, appunto, hanno sostenuto che i modelli mentali non sono meno ‘formali’ di altri tipi di strutture dati: infatti, la manipolazione di tali strutture dipende in maniera sostanziale dalle procedure algoritmiche, via via applicate a seconda del dominio di entità che viene modellizzato. Per esempio, nel caso della logica proposizionale, esistono una serie di vincoli alle procedure di integrazione dei modelli alla base dei processi cognitivi implicati.

In *Mental Models and Thought* Johnson-Laird illustra particolari procedure per il trattamento della congiunzione di una coppia di modelli. Le procedure in oggetto non riguardano solo modelli espliciti, ma anche quelli impliciti (non rappresentati).

Elenchiamo le regole espresse da Johnson-Laird (valide per la logica proposizionale):

La congiunzione di due modelli impliciti dà come risultato un modello implicito.

La congiunzione di un modello implicito con un modello che rappresenta proposizioni dà come risultato un modello nullo (null model) come nel caso dell’insieme vuoto.

La congiunzione di due modelli che rappresentano rispettivamente una proposizione e la sua negazione, danno come risultato un modello nullo ($A \neg B$ and A yields nil).

La congiunzione di due modelli in cui la proposizione B, rappresentata in un modello ma non nell’altro, dipende dall’insieme di modelli (set of models) di cui quel modello è membro. Se B occorre in almeno uno di questi modelli, allora la sua assenza nel modello corrente viene trattata come una negazione ($A B$ and A yields nil).

La congiunzione di due modelli espliciti (fully explicit models) aggiorna il secondo modello con tutte le nuove proposizioni del primo modello ($\neg A B$ and $\neg A C$ yields $\neg A B C$)⁴.

Lolti (2000) ha osservato che è opinabile modellizzare “il ragionamento logicamente corretto” assumendo, come fa Johnson-Laird, che i soggetti si uniformino ai principi di verità e validità, e affermando, allo stesso tempo, che “è possibile ragionare validamente senza ricorrere alla logica” (Johnson-Laird, 1983).

Abbiamo sostenuto, infatti, che Johnson-Laird trasporta la nozione di conseguenza logica all’interno di una teoria della mente: ora, per ironia della sorte, questa viene proposta per contrastare la teoria ‘apparentemente’ opposta della logica mentale (mental logic), la quale postula l’attivazione di schemi formali insensibili al contesto o al contenuto di determinare argomentazioni.

Benissimo, ma se questo avviene proprio per mezzo di un principio (semantico) metalogico, come facciamo a riconoscere che è possibile ragionare validamente senza ricorrere alla logica?

⁴ Johnson-Laird (2005) pp. 191.

Inoltre, secondo Lolli, dato che la procedura di Johnson-Laird si basa sulla restrizione ai modelli finiti, non è in grado di applicarsi a modelli infiniti. Ora, siccome l'indecidibilità della logica del prim'ordine è dovuta proprio all'esistenza di modelli infiniti, in che senso la procedura offerta da Johnson-Laird pretende di essere più efficiente di qualsiasi metodo sintattico (pensiamo ai tableaux analitici)?

Lolli chiarisce questo punto delicatissimo:

In realtà la logica dei modelli mentali è la logica dei modelli finiti 'piccoli'; tuttavia, 'piccoli' è un confine disperatamente vago e non computabile. E per ogni confine che si fissa si potrebbe facilmente inventare un esempio che richiede un elemento in più. [...] Così, se la competenza del dominio non è nota, o non è computabile, e per cominciare scriviamo un programma, non siamo in grado di sapere se quel programma è un programma per una funzione ricorsiva che modella la competenza o piuttosto sappiamo che non lo è. Allora, di che cosa è un programma il nostro programma? Possiamo dare come risposta solo l'insieme dei casi praticamente coperti dall'uso del programma e un numero finito di casi non determina un'unica funzione computabile. In conclusione a venir prima è l'algoritmo, cioè l'esecuzione, non la competenza.⁵

Se concepiamo il ragionamento deduttivo come un insieme di funzioni che mappano un insieme di premesse in un insieme di conclusioni (e gli algoritmi di Johnson-Laird realizzerebbero le funzioni suddette) che garanzie abbiamo, nella vita reale e non in una situazione sperimentale, che gli algoritmi selezionati implementino *solo* le conclusioni valide? Nessuna garanzia. Si prenda, per esempio, questo esempio:

If a pilot falls from a plane without a parachute, the pilot dies.
This pilot did not die, however. Why not? (Johnson-Laird, 2005).

Molti soggetti rispondono che l'aeroplano volava vicino a terra, o che l'aereo era finito nella neve o nel mare ecc, ma soltanto una piccola parte risponde in modo logicamente valido:

The pilot did not fall from the plane without a paracadute.

Questo esempio è veramente significativo e sorprendente: la conclusione valida (il pilota non si è buttato senza paracadute) non viene automaticamente immaginata. Si direbbe che sta lì sotto il naso, ma che non la vediamo, perchè preferiamo spiegazioni di tipo

⁵ Lolli (2000) pp. 96.

causale. La maggior parte di noi, assume, infatti, che il pilota si sia effettivamente buttato, come descritto dall'antecedente del condizionale. Costruiamo, in modo del tutto inconscio, modelli mentali cinematografici di un uomo (non pensiamo mai ad un pilota donna) che si butta da un aereo senza paracadute. Non abbiamo immediatamente accesso al modello mentale che descrive, per esempio, il pilota seduto nell'aeroplano, in tutta tranquillità. È incredibile, ma questo tipo di fenomeni, così ben descritti da Johnson-Laird, si presentano continuamente con soggetti alle prese con problemi di logica. In verità, le deduzioni effettivamente valide nella vita quotidiana sono una piccola parte.

È sicuramente merito di Johnson-Laird aver messo in luce come anche il ragionamento deduttivo sia, in parte, guidato da processi mentali automatici che, in prima battuta, non sono del tutto 'penetrabili'.

Abbiamo detto i modelli mentali, a causa dei ben noti limiti della memoria operativa, costituiscono la versione schematica, in un formato simbolico, dei corrispondenti modelli percettivi, nel senso che sono rappresentati da minor grado di specificità e dettaglio. Abbiamo visto, inoltre, che il modello percettivo di una situazione e il corrispondente modello mentale devono essere isomorfi e che tale isomorfismo si conserva anche tra modelli mentali.

Che cosa significa postulare tale 'somiglianza strutturale' nel caso in cui i modelli mentali si applicano alla logica? Analizziamo, per esempio, il comportamento del connettivo \rightarrow (implicazione materiale) in L0 (logica proposizionale). Si tenga in mente la tavola di verità per il connettivo suddetto: la tavola individua tre stati di cose veri su quattro (si considerino due formule ψ e ϕ): (v,v), (f,v), (f,f). Questi sono esattamente gli stati di cose rappresentati nel modello.

Modelli mentali:

ψ	ϕ
$\neg\psi$	ϕ
$\neg\psi$	$\neg\phi$

Ora, di tale tavola è possibile dare la cosiddetta forma normale disgiuntiva (considerando i casi in cui la tavola di verità produce il valore vero): $(\psi \& \phi) \vee (\neg\psi \& \phi) \vee (\neg\psi \& \neg\phi)$.

Notiamo una qualche somiglianza con la descrizione offerta da Johnson-Laird? Questi tre casi non sono forse gli stati di cose rappresentabili in un modello mentale totalmente esplicito? Nel caso di una forma argomentativa con due premesse disgiuntive (due

disgiunzioni inclusive) per esempio, integrare i modelli mentali *corrisponde* cognitivamente ad una 'composizione isometrica' di tavole di verità.

Consideriamo, infatti, tale formula (consequentia mirabilis) : $(\neg\psi \rightarrow \psi) \rightarrow \psi$

Quali sono i modelli mentali corrispondenti?

- 1) $\neg\psi \quad \psi \quad \psi$
- 2) ...

E i modelli espliciti?

- 1) $\neg\psi \quad \psi \quad \psi$
- 2) $\psi \quad \psi \quad \psi$
- 3) $\psi \quad \neg\psi \quad \psi$
- 4) $\neg\psi \quad \neg\psi \quad \psi$
- 5) $\neg\psi \quad \neg\psi \quad \neg\psi$

In logica si distingue tra connettivi e tavole di verità: i primi sono considerati come operatori sintattici su formule, le seconde come matrici che incorporano operazioni semantiche su valori di verità (Odifreddi, pp. 78). Ciò che conta, in logica, è il comportamento vero-funzionale delle formule. Osserviamo, adesso, la struttura del modello dell'implicazione materiale: esso è isomorfo alla tavola di verità dell'implicazione (tranne per la riga corrispondente al caso (v,f) non rappresentato nel modello). Johnson-Laird sostituisce ai valori di verità, formule e negazioni di formule, corrispondenti a formule vere o false. Ricordiamoci, però, che a enunciati di un calcolo logico, corrispondono modelli mentali di quelle formule. Per Johnson-Laird, tale struttura:

$\psi \quad \phi$

corrisponde ad un modello mentale in cui ψ e ϕ sono veri (la prima riga della tavola di verità).

Tale interpretazione, ripetiamolo, deriva direttamente dalla concezione della logica esposta da Wittgenstein nel *Tractatus* (4.441):

È chiaro che al complesso dei segni 'f' e 'v' non corrisponde alcun oggetto o complesso di oggetti, così come nessun oggetto o complesso di oggetti corrisponde alle righe orizzontali o verticali (della tavola di verità). *Oggetti logici non vi sono* (corsivo nostro).

Anche nei modelli mentali, 'oggetti logici' non sono rappresentati: i modelli nascono 'particolarizzati' e corrispondono sempre a *un* caso particolare tra gli infiniti stati di cose rappresentabili (Johnson-Laird, 1983).

In logica, se a formule sintatticamente differenti corrisponde la stessa tavola di verità, allora, tali formule risultano indistinguibili da un punto di vista semantico. Consideriamo due formule equivalenti da un punto di vista semantico come: $\psi \leftrightarrow \phi \approx (\psi \rightarrow \phi) \& (\phi \rightarrow \psi)$ e $(\psi \vee \phi) \approx \neg(\psi \& \neg\phi)$.

Avranno gli stessi modelli mentali, cioè, risulteranno indistinguibili anche da un punto di vista cognitivo? Tali formule avrebbero modelli mentali isomorfi tra loro?

La risposta a tale interrogativo da parte del teorico dei modelli mentali, sarebbe, però, negativa: a causa delle assunzioni di default, dei limiti fisici della memoria di lavoro e della modulazione semantica, i modelli mentali di tali formule nascono incompleti e dunque non possono, cognitivamente, essere logicamente isomorfi tra loro (rileviamo che, se le assunzioni della teoria sono corrette, allora, questa non è una conclusione da poco per chi è interessato ad una proficua didattica della logica).

Ogni tavola di verità può essere rappresentata da un albero in cui le alternative sono rappresentate dalle lettere proposizionali, le biforcazioni dai loro possibili valori di verità e i rami dalle righe della tavola. Ogni ramo è dunque una congiunzione di lettere o loro negazioni e l'intera tavola è la disgiunzione di rami corrispondenti alle righe 'vere'; se pensiamo di rappresentare i modelli come tavole di verità 'incomplete' e postulando tale struttura ad albero come 'forma logica' comune anche ai modelli, possiamo visualizzare il processo di ragionamento (valido) come l'attraversamento di alcuni rami (quelli che corrispondono alle righe vere) dell'albero in questione.

La struttura ad albero di un modello mentale sarebbe, però, a causa delle limitazioni a carico della memoria di lavoro, incompleta (alcuni rami sarebbero *assenti*) e non sarebbe, così, sempre possibile individuare tutti i modelli mentali di una data formula.

I modelli mentali per la logica proposizionale possono essere considerati come il risultato della *proiezione* della struttura geometrica della forma logica (rappresentata, wittgensteinianamente, dalla tavola di verità): tale proiezione conserverebbe la forma logica in piani al massimo 3D (realizzabili nel cervello) di uno spazio logico che può avere infinite dimensioni (spazi logico-vettoriali).

Un essere ipersferico in grado di vedere la profondità dei nostri corpi e dei nostri cervelli, costruirebbe, probabilmente, modelli mentali molto diversi dai nostri. Lo farebbe anche nei confronti di un sillogismo? Svilupperebbe anch'esso una o molteplici logiche? Ma se lo spazio logico è infinito, come è possibile rappresentarci mentalmente tale ricchezza per mezzo di modelli mentali che possono essere al massimo 3D? Al variare dello spazio logico variano anche le relazioni tra gli enti di tale spazio? Pensiamo all'impostazione della

geometria data dal matematico tedesco Felix Klein: secondo Klein, le geometrie studiano le proprietà delle figure geometriche che risultano *invarianti* rispetto ad un certo gruppo di trasformazioni. La stessa impostazione può essere applicata alla logica (o alle logiche). Immaginiamo di immergere lo spazio logico in una superficie elastica, le proprietà logiche tra gli enti di tale spazio subiranno anch'esse delle trasformazioni. Al variare dei modelli, cosa rimane *invariante* nei modelli, rispetto a un certo gruppo di trasformazioni? Tale ricchezza logica potrebbe essere, cognitivamente, rappresentabile in una classe di modelli mentali? In che senso il formato (simbolico) dei modelli può essere considerato causalmente indipendente dal tipo di operazioni effettuabili nel modello e tra i modelli? Consideriamo, adesso, un'applicazione della teoria dei modelli mentali nei confronti della competenza semantica.

Immaginiamo di avere una lista T di infinite rappresentazioni proposizionali (di enunciati). Siano T_1, T_2, T_3 , la lista di tutti i sottoinsiemi (finiti) di T . Supponiamo di riuscire a trovare un modello per ogni sottoinsieme considerato e siano M_1, M_2, M_3 ... tali modelli mentali (locali, che soddisfano, cioè, parti di T , finite). Ora, per il teorema di compattezza che qui estendiamo anche ai modelli mentali, l'esistenza di tali modelli mentali, garantisce l'esistenza di un *modello unico* che soddisfa tutte le formule di T (che può essere diverso da ciascuno dei modelli locali).

La teoria dei modelli mentali afferma che tale modello, qualunque esso sia, debba essere isomorfo a un modello percettivo. Allora, dato un qualsiasi insieme di rappresentazioni proposizionali, è sempre logicamente possibile individuare un modello percettivo di tale insieme. Ora, se tale modello esiste ed è unico, a meno di isomorfismi, che relazione esiste, da un punto di vista cognitivo, tra tale modello e gli altri modelli locali come i modelli motori, sensoriali, percettivi, spaziali e concettuali?

Supponiamo che sia *logicamente* garantita, così, l'esistenza di un modello mentale unico. Adesso, utilizzando lo stesso procedimento adoperato dal matematico John von Neumann nel generare i numeri naturali a partire dall'insieme vuoto, possiamo, partendo da tale modello, generare, ricorsivamente, un modello di questo modello, e così via. Se però, tale modello è isomorfo ad un modello percettivo, esso deriva, necessariamente, da dati di tipo interno (ai nostri sistemi nervosi e ai nostri corpi) e da dati di tipo esterno (relativi alla struttura della realtà). L'incapsulamento di modelli in modelli e la loro successiva memorizzazione potrebbe aver generato un medium rappresentazionale dove ancorare i simboli linguistici?

Supponiamo che un qualsiasi insieme di enunciati attivi il reclutamento automatico di modelli locali (motori, sensoriali, spaziali, percettivi, ecc.). Tali modelli sarebbero suscettibili di connessione solo perché dispongono di un 'modello' che guida il collegamento tra i modelli locali. Tale unità (prototipica?) sarebbe necessaria, ipotizziamo, non solo per comprendere un qualsiasi insieme di enunciati, ma anche per possedere una propria psicologia. Mente e linguaggio, secondo tale visione, sarebbero due occhi in uno stesso volto.

Nei confronti della mente, siamo come un astronomo bidimensionale 'appiattito' su un palloncino (l'universo) che vede intorno a sé le galassie allontanarsi. Mentre il pallone si gonfia, l'astronomo vede le galassie fuggire ad una data velocità, ma non è in grado di vedere il centro di questa espansione, che si trova all'interno del palloncino. Per lui, concetti come 'all'interno di' non avrebbero senso, non sarebbero concepibili, sarebbero, semplicemente, assenti; ma anche per un neurone o per un modulo o per una rete neurale, concetti come 'conseguenza logica', 'verità', 'modello' sono assenti: non troveremo mai un modello mentale guardando dentro una scatola cranica!

Per completezza espositiva, alcune considerazioni più generali sul rapporto tra modelli mentali e architettura della mente.

Jerry Fodor (2001) ha proposto di considerare due tipologie di spiegazioni utilizzate in scienza cognitiva: "le spiegazioni computazionali in senso stretto" e "le spiegazioni architettoniche". Le prime sono teorie che generano derivazioni, ovvero sequenze causali di rappresentazioni mentali, la cui ultima riga è "un'istruzione che comunica ad opportuni meccanismi effettori di produrre il comportamento" (pp. 56). Le seconde invece, "rispondono a questioni su come (in virtù di quale processo causale) la mente passa da una riga all'altra della derivazione in questione" (pp. 57).

"Data una derivazione che include formule della forma A e $A \rightarrow B$, un processo architettonico separa B automaticamente se A e B sono espressioni primitive" (pp. 57). Secondo Fodor, o tali "decisioni di ordine superiore" sono realizzate in modo computazionale (localmente) oppure sono realizzate in modo non computazionale ma automatico: ossia sono conseguenze *causali* del modo in cui le occorrenze delle rappresentazioni mentali interagiscono con l'architettura cognitiva.

Cos'è che in un modello mentale prende 'decisioni di ordine superiore'? Il programma su calcolatore che implementa i modelli mentali 'realizza' tali decisioni localmente, citando Fodor. Ma cosa dire degli esseri umani? Se le forme logiche sono proprietà locali, sostiene Fodor, allora, dovrebbero essere insensibili alle proprietà globali dei sistemi di credenze.

Ma questo, come ha dimostrato Johnson-Laird per la logica, non avviene. Ora, però, se concordiamo con Fodor sul fatto che anche determinate computazioni su rappresentazioni sono anch'esse proprietà locali, allora, anche tali computazioni (le procedure ricorsive nel caso dei modelli) dovrebbero essere insensibili alle proprietà globali dei sistemi di credenze: ma questo, ipotizzando la natura computazionale dei processi mentali, di fatto, non avviene. Cosa dedurre allora? Dobbiamo ammettere l'inadeguatezza di qualsiasi teoria computazionale del ragionamento? Se ammettiamo che i processi cognitivi implicati in un'inferenza deduttiva sono influenzati dalle proprietà globali del sistema di credenze, com'è possibile spiegare le relazioni causali tra tali processi cognitivi per mezzo di computazioni *nei* modelli (locali)? Se la costruzione di un dato modello mentale dipende da computazioni, in che senso, allora, il soggetto è in grado di agire intenzionalmente nei confronti di tali computazioni, modificandole? Com'è possibile mantenendo *fisso* il modello percettivo, manipolare un maggior numero di modelli mentali di questo? Una volta costruito un modello mentale, siamo in grado di ritornare, a partire da questo, al modello percettivo originario o il processo è unidirezionale? Perché questo 'sganciamento' dal dominio basico, rappresentato dalla costruzione dei modelli mentali, dovrebbe mantenere una somiglianza strutturale (schematica) con i modelli percettivi? Esistono prove di natura neuropsicologica in grado di spiegare tale fenomeno?

I nostri sistemi nervosi possiedono un'intricatissima architettura tridimensionale originatasi da un tessuto bidimensionale, il neuroepitelio, la cui struttura si modifica durante lo sviluppo. Le neuroscienze contemporanee sembrano concordare sull'esistenza, nei nostri cervelli, di moduli funzionali spazialmente separati e operanti in parallelo. Il tipo di informazione elaborata da ciascun modulo (visiva, auditiva, olfattiva, tattile, ecc.) può essere analizzata in parallelo, in modo simultaneo, grazie a circuiti neurali specializzati, concepiti come unità locali distribuite in parti diverse del sistema nervoso centrale. L'attivazione congiunta di tali unità attraverso le fibre nervose forma differenti tipologie di sistemi funzionali (Schlosser, Wagner, 2004).

I modelli mentali, come Johnson-Laird a più riprese ha sostenuto, originano da input di tipo percettivo. Abbiamo proposto di considerare una maggiore varietà sensoriale come originaria dei modelli (non solo di tipo visivo). Se l'attivazione congiunta di determinati modelli mentali locali risulta essere determinante per l'ancoraggio cognitivo di un'insieme di enunciati, risulta fondamentale comprendere, neuropsicologicamente, come tale 'combinazione coesiva' avvenga.

Facciamo notare, però, che armonizzare tale richiesta con l'assunto del formato *simbolico* dei modelli mentali diventa problematico. Da una parte, Johnson-Laird sostiene che i modelli contengono soltanto entità di tipo percettivo, da un'altra, sostiene che all'occorrenza possono contenere anche contrassegni simbolici, da un'altra parte ancora, sottolinea, come elemento distintivo della sua teoria, l'assunto della naturalezza (naturalness) dei modelli.

Il punto è che se vogliamo comprendere come i modelli mentali sono 'computati' dal cervello, non possiamo assumere il formato simbolico dei modelli: il cervello non è composto da simboli o da occorrenze di enunciati! Noi assumiamo che durante la comprensione di un dato enunciato, accade realmente qualcosa dentro la scatola cranica: ma se esiste qualcosa come la combinazione coesiva dei modelli mentali locali, come abbiamo congetturato, questa non può essere caratterizzata da un'insieme di processi di tipo simbolico sconnessi da qualsiasi esperienza neurologica. Se un illustre scienziato cognitivo come Johnson-Laird, è disposto a rigettare qualsiasi ipotesi di tipo dualistico nei confronti della mente, allora, è in debito di una spiegazione che non sia esclusivamente computazionale, relativa al formato simbolico dei suoi modelli mentali.

Ricapitolando, abbiamo cercato di dimostrare, anche con alcuni esempi, che la proprietà di 'essere conseguenza logica' di una proposizione, diventa, nella teoria dei modelli mentali, una proprietà *dei* modelli mentali (delle procedure incorporare nei modelli mentali).

Tale proprietà non è qualcosa che si può scorgere da fuori, per così dire, guardando la forma proposizionale di una data formula. Soltanto dopo aver costruito il modello corrispondente, e soltanto a quel livello, possiamo dire che una data proposizione è conseguenza logica di un insieme di premesse. A differenza dei computer, gli esseri umani, però, possono generare modelli mentali direttamente da dati di tipo percettivo-spaziale facendo a meno di dati di tipo proposizionale come input. Sono, inoltre, in grado di costruire rappresentazioni mentali come i modelli 3D da dati linguistici in ingresso (e viceversa). Se Johnson-Laird ha ragione, i processi di ragionamento sono fortemente condizionati dal tipo di modelli costruiti (che incorporano sempre assunzioni di default) e dal tipo di operazioni che è possibile effettuare nei modelli e tra i modelli (scanning, updating, flashing out, integrating, superimposition, ecc).

Diventa necessario, allora, studiare le tipologie di operazioni che è possibile effettuare nei modelli e tra i modelli, oltre che impegnarsi in una tassonomia dei modelli.

Come abbiamo precedentemente affermato, crediamo che occorra ridimensionare l'assunto molto limitante del formato *esclusivamente* simbolico dei modelli mentali. Proponiamo di considerare i modelli (percettivi, sensoriali, spaziali, motori, ecc.) come rappresentazioni interne derivate dai nostri sistemi sensoriali e motori. Tali modelli interni sarebbero, altresì, attivati a seconda dei predicati, così interpretati: predicati motori susciteranno la costruzione di modelli cinematici, predicati spaziali e percettivi favoriranno la costruzione di modelli spaziali e percettivi, e così via.

È opportuno evidenziare che i modelli mentali, considerati solamente come strutture 3D suscettibili di manipolazione interna, non sono in grado di incorporare processi mentali di tipo deduttivo. Per fare ciò, ricordiamolo, occorre incorporare nei modelli determinate procedure ricorsive operanti in conformità al principio di conseguenza logica; ma che questo sia realizzabile in un programma per calcolatore, non significa *affatto* che la mente umana o il cervello umano eseguano proprio quel tipo di computazioni (si ricordi l'argomento di Lolli).

I modelli *mentali* potrebbero essere considerati come le 'tracce' (termine mutuato dal vocabolario della topologia) dei modelli (embodied, cioè motori, spaziali, percettivi, sensoriali, ecc) originari. Resta problematico, a nostro avviso, conciliare l'ipotesi del contenuto simbolico dei modelli, con le caratteristiche bio-fisiche dei nostri cervelli e dei nostri corpi. Detto questo, come si 'interconnettono' a livello sub-simbolico i vari modelli? Da un punto di vista computazionale, ci sono delle differenze tra modelli mentali motori e spaziali, per esempio? Tali differenze tra modelli si manifesterebbero anche a un livello linguistico? In che senso le procedure ricorsive che utilizzano i modelli stessi come input, possono applicarsi a modelli motori oltre che a modelli spaziali?

Immaginiamo di ascoltare questo enunciato: "muovi il braccio ritmicamente una volta verso il petto e una volta verso il lato opposto". In tale enunciato, è già contenuto il modello mentale (motorio) dell'enunciato: il modello sarebbe, centrato sull'ego, cinematico, e 'cognitivamente penetrabile', infatti, possiamo modificarlo volontariamente in un'istantanea, in un'immagine statica. Si può parlare, qui, di un livello di ricorsività sub-simbolica, intesa come la possibilità di utilizzare le *stesse* computazioni che hanno generato il modello motorio anche per modificarlo? O sono computazioni qualitativamente diverse? La mente stessa sarebbe il risultato del modo in cui i modelli (locali) si 'collegano' tra loro in un procedimento continuo e tale 'combinazione coesiva' sarebbe mantenuta anche *in* un insieme di enunciati.

Veniamo, adesso, alle conclusioni finali. I modelli mentali incorporano il principio di verità in questa accezione: i modelli mimano il comportamento dei modelli logici come insiemi o strutture che rendono veri tutti gli assiomi di un dato sistema formale.

Possiamo, inoltre, considerare i modelli mentali come ambienti virtuali dove è possibile eseguire determinate operazioni su determinati oggetti in conformità a principi logici (verità e validità). Johnson-Laird recupera, inoltre, l'idea che la verità di un enunciato dipende dal tipo di ambiente matematico (il modello) che realizza l'enunciato. La versione cognitiva di tale principio considera la verità di un enunciato come una proprietà (accessibile dalla mente) dipendente dal tipo di modelli (l'interfaccia tra percezione e linguaggio) che costruiamo, compatibili con l'*immaginabilità* dell'enunciato. Non si tratta di strutture insiemistiche o algebriche, ma di strutture percettivo-spaziali che i nostri cervelli sono in grado di computare. Recupera, come abbiamo visto, alcune idee sulla natura della logica espresse da Wittgenstein nel *Tractatus*.

In secondo luogo, la teoria postula che proprio a causa del principio di verità (il modello rappresenta, per definizione, soltanto i casi veri) i ragionatori non andranno sempre alla ricerca di controesempi (di modelli aggiuntivi) che, appunto, falsificano le premesse e rendono invalide le conclusioni e fa delle previsioni sul numero di modelli che i soggetti prenderanno in considerazione.

Il modello di soggetto epistemico che ne deriva è, come ha sostenuto Legrenzi, (2001, pp.40) quello di un soggetto che si accontenta di rappresentazioni incomplete, di un 'satisfacer', molto simile al soggetto tipico descritto dalla teoria della razionalità limitata di H. Simon.

A differenza della teoria della razionalità limitata, però, la teoria dei modelli mentali cerca di penetrare nei mondi rappresentazionali dei soggetti e ne descrive il funzionamento. Che questi mondi siano, per la maggior parte, fondati su rappresentazioni di natura percettivo-spaziale, anche quando il contenuto delle rappresentazioni assume un formato simbolico, significa che il prossimo passo della ricerca sarà lo studio comparato dei modelli mentali (motori, spaziali, percettivi, sensoriali, concettuali, ecc) compatibili con le differenti architetture del cervello umano, animale e artificiale.

Bibliografia

- Adenzato, M., Bucciarelli, M., Sacco, K. (2001) Mental Models and the meaning of connectives: A study on children, adolescents and adults, Proceedings of the XXIII Conference of the Cognitive Science Society. In J.D. Moore and K. Stenning (eds.). Mahwah, NJ, Lawrence Erlbaum Associates, pp.875-880.
- Bara, B. (1990) *Scienza cognitiva: un approccio evolutivo alla simulazione della mente*, Bollati Boringhieri, Torino.
- Bara, B.G., Bucciarelli, M., Lombardo, V. (2001) Model theory of deduction: a unified computational approach. *Cognitive Science*, 25, pp.839-901.
- Bazzanella, B., Bouquet, P., Warglien, M. (2005) What do external representations tell about Mental Models? An exploratory study in deductive reasoning, Proceedings of the XXVII Annual Conference of the Cognitive Science Society (Cogsci, 2005) 21-23, July 2005, Stresa, Italy.
- Bouquet, P., Warglien, M., (1998) Mental models and Local models semantics: the problem of information integration, Proceedings of the European Conference of Cognitive Science (AICS'98) University College Dublin (Ireland) 19-21th August, 1998.
- Bozzi, S., Mangione, C. (2001) *Storia della logica: da Boole ai nostri giorni*, Garzanti, Milano.
- Braine, M.D.S. (1978) On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85, pp.1-21.
- Cherubini, P., Giarretta, P., Mazzocco, A. a cura di (2000) *Ragionamento: psicologia e logica*, Giunti, Firenze.
- Chiu, C.Y, Lau, Y.Y.M., Lee, S. L., Kiesler, S. (2004) Human Mental Models of humanoid robots, IEEE, Computer Society, International Conference on Robotics and Automation Proceedings, 2005, pp.2767-2772.
- Colombetti, M. (2000) Ragionare senza significato, in *Ragionamento: psicologia e logica*, a cura di Cherubini, P., Giarretta, P., Mazzocco, A., Giunti, Firenze, 2000, pp. 157-169.
- Cosmides, L., Toby, J. (1996) A logical design for the mind? (Review of the *Psychology of Proof*, Lance Rips, MIT Press, 1994) *Contemporary Psychology*, 41, pp.448-450.

- Davidson, D. (2005) *Truth and Predication*, The Estate of Donald Davidson (trad.it. *Sulla Verità*, Laterza, Roma-Bari, 2006).
- Dennet, D.C. (1996) *Kinds of Minds*, Basic Books, New York (trad.it. *La mente e le menti*, Sansoni, Milano, 1997).
- D’Mello, S., Grasser, A., Ventura, M. (2005) A computational approach to Mental Models, Proceedings of the XXVII Annual Conference of the Cognitive Science Society (pp.2289-2294) Stresa, Italy.
- Eco, U., Santambrogio, M., Violi, P. a cura di (1988) *Meaning and mental representations*, Indiana University Press, Bloomington.
- Evans, J.St.B.T., Over, D.E. (1996) *Rationality and reasoning*, Psychology Press, Hove.
- Fagin, R., Helpert, J.Y. (1988) Belief, awareness and limited reasoning, *Artificial Intelligence*, 34, pp.39-76.
- Ferretti, F., Gambarara, D., a cura di (2005) *Comunicazione e scienza cognitiva*, Laterza, Roma-Bari.
- Fodor, J. (2001) *La mente non funziona così: la portata e i limiti della psicologia computazionale*, Laterza, Roma-Bari.
- Frege (1923) *Ricerche Logiche*, a cura di Michele Di Francesco, Angelo Guerini e Associati, Milano, 1988.
- Frixione, M., (1994) Intelligenza artificiale e semantica: il problema degli atteggiamenti proposizionali. Apparso sugli Atti del Congresso Annuale dell’Associazione Italiana Calcolo Automatico (AICA) Palermo, Settembre 1994, Vol. I, pp.425-437.
- Frosch, C.A, Johnson-Laird, P.N. (2006) The revision of beliefs about causes and enabling conditions, Proceedings of the XXVIII Annual Meeting of the Cognitive Science Society, pp. 1329-1333.
- Hagen,C., Knauf, M., Rauh, R., Schlieder, C., Strube, G. (2005) Preferred and alternative Mental Models in spatial reasoning, *Spatial Cognition and Computation*, Vol.5, N° 2,3, pp.239-269.
- Johnson-Laird, P.N. (1983) *Mental Models*, Cambridge UP, Cambridge (trad.it. *Modelli Mentali*, il Mulino, Bologna, 1988).
- Johnson-Laird, P.N. (1988) How is meaning mentally represented? In *Meaning and mental representations* (a cura di) Eco, U., Santambrogio, M., Violi, P., Indiana University Press, Bloomington, pp.99-119.

- Johnson-Laird, P.N. (1989) *Mental Models*, in Posner, M.I. (Ed) *Foundations of Cognitive Science*, Cambridge, MA: Bradford, MIT Press.
- Johnson-Laird, P.N. (1989) *Pensiero e modelli mentali*, in *Mente umana, mente artificiale* (a cura di) Riccardo Viale, pp. 260-278, Feltrinelli, Milano, 1989.
- Johnson-Laird, P.N. (1993) *Human and Machine Thinking*, Erlbaum, Hillsdale (NJ) (trad.it. *Deduzione, Induzione, Creatività. Pensiero umano e pensiero meccanico*, il Mulino, Bologna, 1994).
- Johnson-Laird, P.N. (1993) *The computer and the mind. An introduction to cognitive science*, London, William Collins Sons e Co.Ltd. (trad.it. *La mente e il computer. Un'introduzione alla scienza cognitive*, Il Mulino, Bologna, 1997).
- Johnson-Laird, P.N. (1996) *Mental Models, deductive reasoning and the brain*, in *The Cognitive Neuroscience* (Eds) Michael S.Gazzaniga, Bradford Book, MIT Press, Cambridge Massachusetts, London, England, pp.999-1008.
- Johnson-Laird, P.N., Savary, F. (1999) *Illusory inferences: a novel class of erroneous deductions*, *Cognition*, 71, pp.191-229.
- Johnson-Laird, P.N. (2000) *Perché la logica non è un buon modello del ragionamento deduttivo*, in *Ragionamento: psicologia e logica* (a cura di) Cherubini P., Giarretta, P., Mazzocco, A., Giunti, Firenze, 2000, pp.17-68.
- Johnson-Laird, P.N. (2001) *Mental Models and deduction*, *Trends in cognitive science*, 5, pp.434-442.
- Johnson-Laird, P.N. (2004) *The history of Mental Models*, in Manktelow, K. and Chung, M.C. (Eds) *Psychology of reasoning: theoretical and historical perspectives*, NY: Psychology Press, pp.179-212.
- Johnson-Laird, P.N., Lee Louis, N.Y. (2004) *Creative strategies in problem solving*, in Forbus, K. Gentnend and Regier, T. (Eds) *Proceedings of the XXVI Annual Conference of the Cognitive Science Society*, Mahwah, N.J.:LEA, pp.813-818.
- Johnson-Laird, P.N. (2005) *Mental Models and Thought*, in Holyak,K., Sternberg, R.J. (Eds) *The Cambridge Handbook of Thinking and Reasoning*, Cambridge University Press, pp.179-212.
- Johnson-Laird, P.N., Yang,Y. (2007) *Mental logic, mental models and computer simulations of reasoning*, in Sun, R. (Ed) *Cambridge Handbook of Computational Psychology*, Cambridge University Press, in press.

- Kandel, E.R., Schwarz, J.H., Jessel, T.M. (1995) *Essentials of Neural Science and Behaviour*, Appleton and Lange (trad.it. *Fondamenti delle neuroscienze e del comportamento*, Ambrosiana, Milano, 1999).
- Knauff, M., Nebel, B., Ragni, M. (2005) A computational model for spatial reasoning with mental models, in Bara, B., Barsalau, B., Bucciarelli, B. (Eds), *Proceedings of the XXVII Annual Cognitive Science Conference (Cogsci 2005)* pp.1064-70, Erlbaum.
- Knauff, M., Johnson-Laird, P.N. (2007) Preferred Mental Models in spatial reasoning, *Memory and Cognition*, 35, pp.2075-2086.
- Lakoff, G., Nunez. (2000) R. *Where mathematics comes from: how the embodied mind brings mathematics into being*, BasicBooks, New York.
- Legrenzi, P. (2001) *Come funziona la mente*, Laterza, Roma-Bari.
- Lolli, G. (2000) Un logico esamina i modelli mentali, in *Ragionamento: psicologia e logica*, a cura di Cherubini, P., Giarretta, P., Mazzocco, A., Giunti, Firenze, 2000, pp.81-93.
- Lolli, G. (1989) *Mente e matematica*, in *Mente umana, mente artificiale* (a cura di Riccardo Viale, Feltrinelli, Milano, pp.84-93.
- Lolli, G. (2004) *Da Euclide a Gödel*, Il Mulino, Bologna.
- Lombardo Radice, L. (1970) Un nuovo livello di astrazione: la teoria delle categorie, in *Insiemi, Gruppi, Strutture, Le scienze Quaderni*, n.21.
- Longo, G. (1999) Lo spazio, i fondamenti della matematica e la resistibile ascesa della metafora: il cervello è un calcolatore digitale. Apparso in *L'uomo e le macchine*, a cura di M.Bresciani Califano, Leo S. Olschki, Firenze, 2002.
- Luccio, R. (1998) *Psicologia generale: le frontiere della ricerca*, Laterza, Roma-Bari.
- Manca, V. (2001) *Logica matematica: strutture, rappresentazioni, deduzioni*, Bollati Boringhieri, Torino.
- Odifreddi, P. (2003) *Il diavolo in cattedra: la logica da Aristotele a Gödel*, Einaudi, Torino.
- Palladino, D., Palladino, C. (2007) *Logiche non classiche: un'introduzione*, Carocci, Roma.
- Peruzzi, A.(2004) *Il significato inesistente: lezioni sulla semantica*, Firenze University Press.
- Peruzzi, A. (2005) Il lifting categoriale dalla topologia alla logica, *Annali del Dipartimento di Filosofia* 2006, Firenze University Press.

- Putnam, H. (2003) *Mente, corpo, mondo*, Il Mulino, Bologna.
- Rips L.J. (1994) *The psychology of proofs*, MIT Press, Cambridge, MA.
- Schlosser, G., Wagner, G.P., ed. by (2004) *Modularity in development and evolution*, The University of Chicago Press, Chicago and London.
- Schwamb, K.B. (1990) Mental Models: a survey, url: citeseer.nj.nec.com/schwamb90Mental.html
- Viale, R., a cura di (1989) *Mente umana, mente artificiale*, Feltrinelli, Milano.
- Vosgerau, G. (2006) The perceptual nature of Mental Models, in *Mental Models in cognitive psychology, neuroscience and philosophy of mind*, (eds) Held, C., Knauff, M., Vosgerau, G. Elsevier B.V., pp. 225-275.
- Wilks, Y. (1988) Reference and its role in computational models of mental representations, in *Meaning and Mental Representations* (ed.by) Eco, U., Santambrogio, M., Violi, P., 1988, pp.221-236.
- Wittgenstein, L. (1922) *Tractatus Logico-philosophicus e Quaderni 1914-1916*, Einaudi, Torino, 1994.
- Yang, Y., Bringsjord, S. (2001) Mental MetaLogic: a new paradigm in psychology of reasoning, in L.Chen, Y.Zhou (Eds) *Proceedings of the Third International Conference of Cognitive Science (ICCS 2001)* Press of the University of Science and Technology of China, Hefei, China, pp.199-204.