

# INTERVENTI



# L'APPROCCIO COMPUTAZIONALE IN PSICOLOGIA COGNITIVA

MARCO ZORZI

*Università di Padova*

## INTRODUZIONE

La ricerca psicologica, in particolare quella sui processi cognitivi umani, si sta orientando in modo sempre più netto verso l'utilizzo della simulazione con modelli computazionali come principale strumento per lo sviluppo teorico. In questo articolo descrivo tre fasi fondamentali della modellizzazione e mi soffermo sui criteri di valutazione di un modello e in particolare sul problema di aggiudicare tra teorie in competizione. Attraverso gli esempi di modelli computazionali basati sulle reti neurali (modelli connessionisti) sviluppati nel mio laboratorio, discuto una serie di fattori che determinano il successo o il fallimento di un modello e la necessità di esplicitare e di valutare in modo sistematico le assunzioni contenute rispetto alla natura dei processi, dell'apprendimento, delle rappresentazioni e dell'architettura. Nella parte finale sostengo la necessità di aumentare la plausibilità biologica dei modelli allo scopo di ottenere una migliore corrispondenza tra dati simulati e dati empirici a diversi livelli di descrizione, da quello comportamentale (psicologia) a quello neurale (neuroscienze).

L'interesse per la modellizzazione dei processi cognitivi e per la comprensione delle computazioni sottostanti alla cognizione umana si è fortemente sviluppato negli anni Settanta del secolo scorso con la psicologia cognitivista (o HIP, *Human Information Processing*), che ha tentato sistematicamente di descrivere questi processi in modo meccanicistico attraverso dei diagrammi di flusso. Nella tassonomia dei modelli cognitivi riportata in figura 1, i modelli a «frecce e scatole» (*box and arrows*) rientrano nella categoria dei modelli pre-quantitativi. Questi modelli «boxologici» hanno permesso notevoli progressi in psicologia cognitiva per la loro capacità di generare predizioni empiriche piuttosto precise ma comunque di natura qualitativa. I modelli quantitativi si dividono invece tra i modelli matematici, che nella forma più semplice possono consistere di un'unica equazione (come le ben note

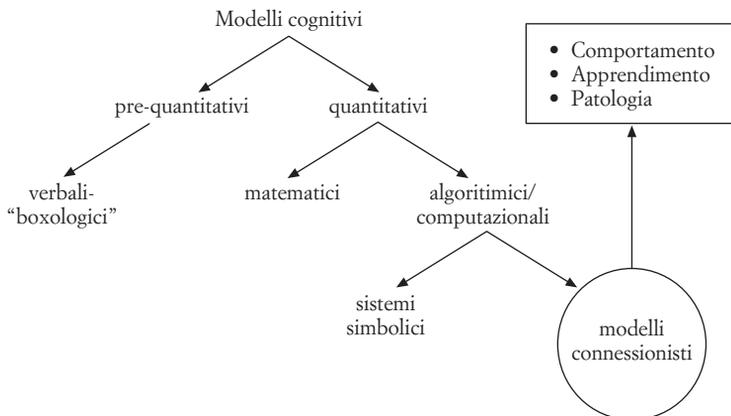


FIG. 1. Una tassonomia dei modelli cognitivi (adattata da Jacobs e Grainger, 1994).

leggi della psicofisica), e i *modelli computazionali*, che sono l'oggetto di questo articolo.

Un modello computazionale, oltre a descrivere in modo estremamente preciso e quantitativo il comportamento umano (capacità che può condividere con un modello matematico), offre una descrizione formale dei meccanismi sottostanti al comportamento in esame. L'approccio computazionale in psicologia parte dall'assunzione che i sistemi cognitivi esibiscano delle proprietà «mentali» in virtù della loro organizzazione causale e che questa organizzazione sia di importanza primaria nella spiegazione del comportamento. Una proprietà centrale della computazione è proprio quella di offrire una specificazione dell'organizzazione causale di un sistema.

Prenderò qui in considerazione una particolare classe di modelli computazionali, i *modelli connessionisti*, che sono sistemi di elaborazione dell'informazione ispirati al funzionamento del cervello e ai principi della computazione neurale. Una discussione dei meriti di questo approccio, e ancor più un confronto con l'approccio computazionale simbolico, va oltre gli scopi di questo articolo. Mi limiterò quindi a sottolineare che l'approccio connessionista si è progressivamente affermato nel corso degli ultimi due decenni perché si è dimostrato il più adatto ad affrontare il problema in una prospettiva che unifica lo studio di mente e cervello: nei modelli connessionisti, i processi cognitivi sono implementati in termini di complesse interazioni non-lineari tra un ampio numero di semplici unità di elaborazione che formano una rete neurale (per approfondimenti sul dibattito tra sistemi simbolici e sub-simbolici si veda, ad esempio, Rumelhart e McClelland, 1986 e Chalmers, French e Hofstadter, 1992). Come ve-

dremo in seguito, il connessionismo è l'unico paradigma computazionale che permette di unificare lo studio del comportamento, dell'apprendimento e della patologia.

#### LA SIMULAZIONE CON MODELLI COMPUTAZIONALI

I modelli computazionali sono caratterizzati dalla loro capacità di *simulare* un comportamento, o più in generale un insieme di dati empirici. I metodi di simulazione costituiscono un passo importante per avvicinare maggiormente la psicologia alle scienze esatte (si veda Parisi, 2001, per una discussione approfondita). Ma che cos'è una *simulazione*? E quali scopi ha? Partiamo da una premessa: la mente è un meccanismo non osservabile, non può essere «aperta» per studiarne il funzionamento. Gli psicologi possono utilizzare solo dei metodi indiretti (la psicofisica, la cronometria mentale, la neuropsicologia) per inferire il funzionamento di processi mentali che non sono direttamente osservabili.

Il metodo simulativo consiste nello sviluppare un modello che è esplicito dal punto di vista *computazionale*, tradotto cioè in un programma per computer che dovrebbe riprodurre in modo fedele il comportamento umano. Si tratta di un importante passo avanti rispetto ad una tradizionale teoria verbale. La simulazione non si limita a spiegare il comportamento ma piuttosto lo riproduce. Spesso una teoria verbale è il punto di partenza per sviluppare un modello computazionale, però quest'ultimo diventa molto più esplicito e preciso della teoria da cui deriva. La simulazione permette quindi di valutare in modo puntuale l'adeguatezza di una teoria (cioè quanto bene possa spiegare i dati sperimentali) e di ottenere predizioni che potranno essere verificate attraverso nuovi esperimenti. Può essere considerata come un laboratorio sperimentale virtuale, nel quale si possono osservare i fenomeni (simulati) e si possono manipolare le variabili per osservarne gli effetti. In alcuni casi, come vedremo in seguito, la simulazione può servire a testare teorie alternative.

Nel caso dei processi cognitivi, le simulazioni possono riguardare tre diversi ambiti:

- il comportamento di soggetti adulti in un compito cognitivo (documentato da accuratezza e/o tempi di reazione in esperimenti di psicologia sperimentale);
- l'acquisizione di una data funzione o abilità cognitiva (descritta da esperimenti su soggetti in età evolutiva);
- il danneggiamento di una funzione cognitiva in seguito a una lesione cerebrale (riportato da studi neuropsicologici su pazienti).

Sviluppare un modello computazionale è un processo lungo e complesso. È utile quindi individuare tre fasi fondamentali della modellizzazione.

1. *Dalla teoria al modello computazionale.* Un modello computazionale viene realizzato a partire da una teoria di riferimento. Il modello è un programma per computer che non «gira» se non è pienamente specificato. Quindi una teoria non diventa un programma funzionante, ovvero un modello computazionale, se non è pienamente specificata. Formulare una teoria sui processi cognitivi attraverso un modello computazionale rivela immediatamente molti modi in cui la teoria è incompleta o sotto-specificata.

2. *La valutazione del modello computazionale.* Quando il programma «gira» possiamo valutare in modo rigoroso, attraverso la simulazione, l'adeguatezza della teoria: troviamo nel comportamento del modello gli stessi effetti che osserviamo nel comportamento umano?

3. *La discrepanza tra simulazione e sperimentazione.* La discrepanza tra dati reali e dati simulati ci rivela in che modo la teoria di partenza è sbagliata. A volte una parziale riformulazione di aspetti specifici della teoria possono eliminare le discrepanze – in questo caso la simulazione ha portato ad una migliore teoria. Altre volte le discrepanze sono così fondamentali da non poter essere eliminate attraverso modificazioni minori della teoria – in questo caso la simulazione ha portato a rigettare la teoria.

È utile soffermarsi sui criteri di valutazione perché da questi dipenderà il nostro giudizio sull'adeguatezza del modello. Descriverò qui di seguito tre criteri: l'adeguatezza descrittiva, la semplicità e la generalità.

#### ADEGUATEZZA DESCRITTIVA

Il primo e fondamentale criterio di valutazione di un modello è quello della sua *adeguatezza descrittiva*. Con questo termine si indica il grado di accuratezza con cui un modello descrive (simula) un insieme di dati sia a livello qualitativo che quantitativo (*goodness of fit*). Vi sono due modi di valutare la corrispondenza tra i dati di uno studio su soggetti umani e i dati simulati dal modello sottoposto agli stessi stimoli e alle stesse condizioni sperimentali:

– *Metodo fattoriale*. Viene condotta un’analisi della varianza sui tempi di reazione (TR) e/o sugli errori del modello utilizzando gli stessi fattori (variabili indipendenti) utilizzati nelle analisi dei dati umani. Dovremo quindi valutare se tutti gli effetti che risultano significativi nelle analisi dei dati umani lo sono anche nei dati del modello. Procederemo in modo analogo per gli effetti non significativi, allo scopo di verificare che il modello non produca effetti paradossali (ovvero che non sono osservati nel comportamento umano).

– *Metodo della regressione*. Viene condotta un’analisi della regressione lineare in cui i dati del modello (ad esempio i TR) vengono utilizzati come predittori dei dati umani. Ciò può essere fatto a livello dei singoli item (per ognuno dei diversi stimoli in un esperimento). La proporzione di varianza spiegata (R-quadrato) indica in modo preciso quanto buona è la corrispondenza.

L’adeguatezza descrittiva è il criterio principale con cui un modello viene valutato dalla comunità scientifica. Una valutazione complessiva terrà conto di quanto il modello raggiunge due obiettivi: la *completezza*, ovvero l’aver adeguatamente specificato tutti i processi in gioco, e la *sufficienza*, ovvero l’aver offerto una spiegazione di tutti i fenomeni empirici rilevanti. Tuttavia, è importante sottolineare che l’adeguatezza di un modello non garantisce affatto la sua correttezza, né quella della teoria a cui fa riferimento. È possibile infatti che una seconda teoria, implementata in un modello computazionale diverso, produca lo stesso risultato. Come vedremo in seguito, questa eventualità è tutt’altro che remota e ci porterà a discutere il problema di come aggiudicare tra modelli in competizione.

#### GENERALITÀ E SEMPLICITÀ

In aggiunta all’adeguatezza descrittiva, è possibile utilizzare altri due criteri di valutazione. Questi sono la *generalità* e la *semplicità* del modello. Il primo si riferisce alla possibilità di generalizzare il modello a stimoli diversi, compiti diversi, o tipi di risposta diversi. Un problema diffuso in psicologia cognitiva è infatti quello di costruire teorie che non si limitino a spiegare un determinato effetto, misurato da una particolare variabile dipendente, in un compito specifico. Il criterio della semplicità si riferisce principalmente al numero di assunzioni o di parametri che caratterizzano un modello. La semplicità può essere valutata più facilmente in termini relativi piuttosto che in termini assoluti. Un modello può essere più semplice dei modelli alternativi, oppure può essere relativamente semplice rispetto al grado di generalità e di adeguatezza descrittiva.

La generalità di un modello è collegata anche ad un principio che dovrebbe essere seguito nella fase di progettazione. In altre scienze, soprattutto quelle esatte, è procedura standard che un nuovo modello renda conto degli effetti critici spiegati dalla precedente generazione dello stesso modello o dei modelli alternativi. Questa strategia di modellizzazione «nidificata» o *nested modeling* (Jacobs e Grainger, 1994) richiede che un nuovo modello abbia una chiara relazione con il suo diretto precursore e che sia valutato rispetto ai set di dati che avevano motivato la costruzione del vecchio modello prima di essere testato rispetto a nuovi dati. Un nuovo modello dovrebbe quindi includere il vecchio come un caso speciale o fornire una dimostrazione formale di questa inclusione, oppure dovrebbe rifiutarlo dopo averne falsificato le assunzioni di base. Sfortunatamente, questo principio è stato poco utilizzato come principio guida della modellizzazione in psicologia.

#### LA COMPETIZIONE TRA MODELLI

In psicologia cognitiva esistono numerosi esempi di teorie in competizione. Per un qualsiasi fenomeno sperimentale, sarà facile trovare almeno due teorie alternative che sostengono di spiegarlo. Questo fenomeno, più diffuso nelle scienze umane che nelle scienze esatte, è parte integrante del progresso scientifico ma solleva in maniera forte il problema di come aggiudicare tra teorie in competizione. Uno dei motivi del successo dell'approccio computazionale è proprio quello di facilitare l'aggiudicazione. Se le teorie in competizione sono davvero diverse, lo saranno anche i rispettivi modelli computazionali. A questo punto, la competizione tra i due modelli può procedere in due fasi successive. Nella prima fase, eseguiremo una «valutazione comparativa» dell'adeguatezza descrittiva dei diversi modelli. Vi sono effetti critici spiegati da un modello ma non dall'altro? Un modello spiega una proporzione maggiore di varianza nei dati umani rispetto all'altro? Tuttavia, è possibile che due modelli siano molto simili dal punto di vista dell'adeguatezza descrittiva. In altre parole, tutti gli effetti simulati dal modello A sono simulati anche dal modello B, e viceversa. Se i dati non discriminano tra i modelli, come possiamo scegliere tra i due? A questo punto rimangono due possibili strade. La prima è quella di utilizzare il criterio della semplicità: scegliere il modello più semplice. Una seconda via è quella del *falsificazionismo e inferenza forte*. Questa è una strategia di ricerca basata sulle capacità di un modello di generare predizioni che possono essere falsificate. Se due modelli non sono solo delle varianti notazionali, un'analisi fine delle

loro differenze farà emergere modi in cui i modelli fanno predizioni diverse su esperimenti non ancora condotti. Fare questi esperimenti potrebbe quindi portare all'aggiudicare tra le due teorie (rigettando quella che ha fatto la predizione sbagliata), oltre a generare nuove scoperte sul dominio cognitivo in questione.

Nella parte restante di questo articolo, passando in rassegna alcune ricerche computazionali condotte nel mio laboratorio (*Computational Cognitive Neuroscience Lab*; <http://ccln.psy.unipd.it>), mostrerò quali fattori determinano il successo o il fallimento di un modello connessionista dei processi cognitivi. In particolare, dimostrerò che è necessario esplicitare e valutare le assunzioni contenute nel modello rispetto alla natura dei processi, dell'apprendimento, delle rappresentazioni e dell'architettura. Una questione aggiuntiva, che tratterò nella parte conclusiva ma che si interseca con tutte le altre, è quella della *plausibilità biologica* dei modelli.

#### I PROCESSI

Tra i più avanzati e paradigmatici esempi dell'approccio computazionale in psicologia cognitiva si possono citare i modelli proposti nell'ambito della lettura e del riconoscimento di parole. A partire dai lavori iniziali di Seidenberg e McClelland (1989), i modelli connessionisti della lettura hanno fornito un'opportunità unica per formulare dettagliate predizioni sperimentali (falsificabili), che hanno profondamente influenzato la ricerca empirica. Tali modelli sono molto più espliciti delle tradizionali teorie verbali e possono produrre simulazioni molto dettagliate delle prestazioni umane, sia dal punto di vista dei tempi di reazione che dell'accuratezza nella lettura di parole (Zorzi, 2005, per una rassegna). I modelli connessionisti riescono inoltre a simulare le prestazioni deficitarie dei pazienti che hanno subito lesioni cerebrali (ad es., Houghton e Zorzi, 2003; Zorzi, Houghton e Butterworth, 1998a) e lo sviluppo delle abilità di lettura nei bambini (Zorzi, Houghton e Butterworth, 1998b).

Tre diversi modelli computazionali si sono affermati in letteratura (vedi fig. 2): il modello DRC (*Dual Route Cascaded*) di Coltheart, Rastle, Perry, Langdon e Ziegler (2001), il modello *Triangle* (triangolo) di Plaut, McClelland, Seidenberg e Patterson (1996) e il modello CDP (*Connectionist Dual Process*) di Zorzi e collaboratori (1998a). Una caratteristica cruciale del modello DRC è che la sua via di lettura nonlessicale opera attraverso un processo seriale che effettua la conversione grafema-fonema prendendo una lettera per volta da sinistra verso destra. Gli altri due modelli, invece, funzionano competamente in modo parallelo come tipico nei modelli connessionisti. Questa dif-

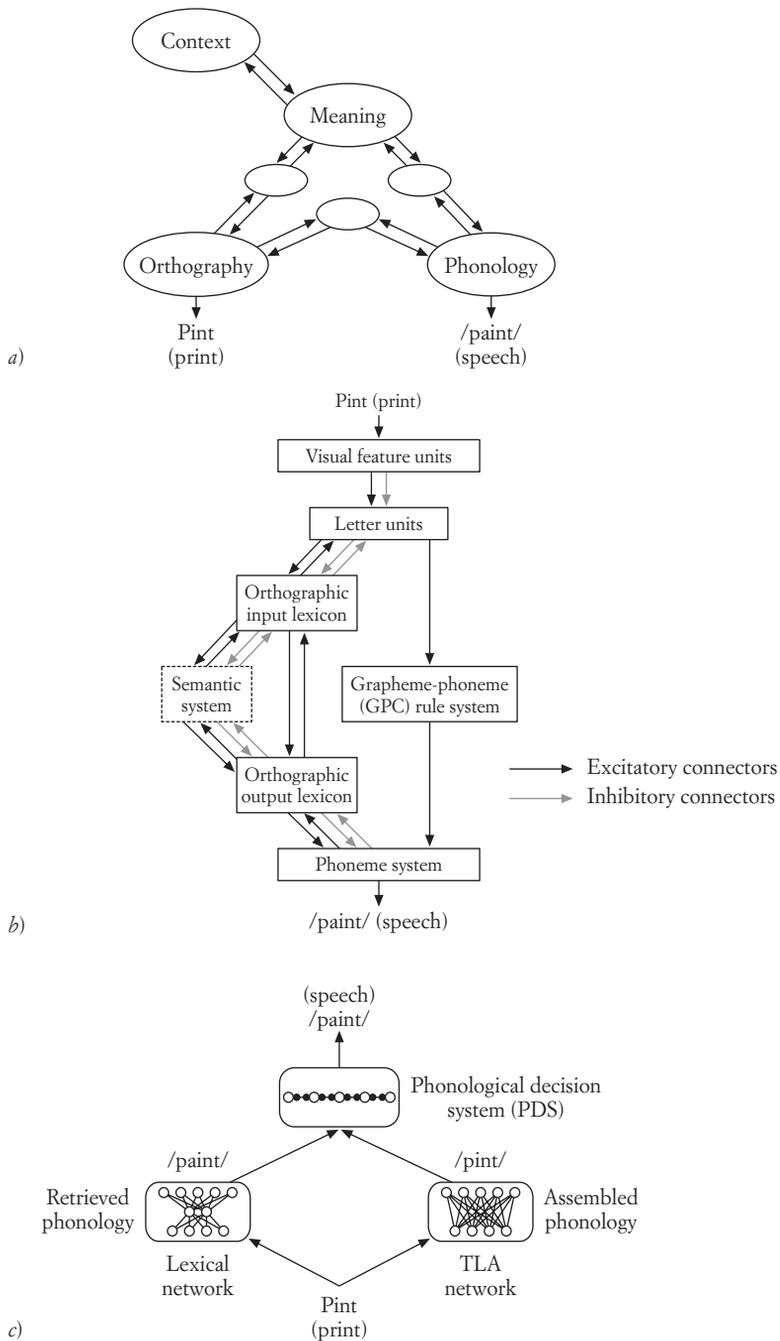


FIG. 2. Modelli computazionali della lettura di parole. **A:** Il modello «Triangle» di Plaut *et al.* (1996). **B:** Il modello DRC di Coltheart *et al.* (2001). **C:** Il modello CDP di Zorzi *et al.* (1998).

ferenza è rilevante dal punto di vista teorico e ha spinto alcuni ricercatori ad eseguire degli esperimenti che permettessero di aggiudicare tra i diversi modelli. In particolare, è stato descritto un fenomeno che è stato interpretato come *effetto seriale* e che confermerebbe le assunzioni teoriche del modello DRC. Questo è noto come «effetto di posizione dell'irregolarità»: il costo dell'irregolarità (differenza di TR tra parole irregolari e regolari) nelle parole a bassa frequenza è modulato dalla posizione nella parola della corrispondenza irregolare tra grafema e fonema. Ad esempio, la corrispondenza irregolare è nella prima posizione in CHEF, nella seconda posizione in TOMB, e nella terza posizione in GLOW. Rastle e Coltheart (1999), gli scopritori di questo effetto, concludono che «I modelli che operano solo in parallelo non predicano questo effetto seriale». Si noti che questo è un esempio di falsificazionismo e inferenza forte. Dimostrare nei dati umani la presenza di un effetto predetto dal modello DRC li ha portati a rigettare tutti gli altri modelli che non prevedono un processo seriale.

Tuttavia, Rastle e Coltheart (1999) hanno commesso un errore fondamentale: non hanno infatti verificato con una simulazione le effettive predizioni degli altri modelli computazionali. Infatti, il modello CDP può simulare questi fenomeni «seriali» nonostante la sua architettura parallela (Zorzi, 2000; si veda la fig. 3). L'analisi del modello ha poi permesso di scoprire che l'effetto «seriale» dipende da una variabile non controllata negli stimoli sperimentali (si veda Zorzi, 2000, per ulteriori dettagli). Più in generale, questo esempio dimostra la fallacia nell'attribuire un fenomeno che si suppone seriale ad una

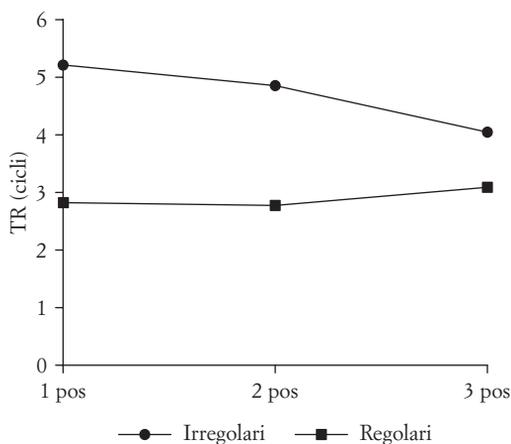


FIG. 3. L'effetto di posizione dell'irregolarità nel modello CDP (Zorzi, 2000). Si noti come la differenza nei tempi di reazione (TR) del modello tra parole irregolari e regolari è modulato dalla posizione del grafema irregolare.

specifica assunzione riguardo all'architettura del modello. Inoltre, tentare di derivare predizioni da complessi modelli computazionali senza effettuare le simulazioni può portare ad inferenze errate.

#### L'APPRENDIMENTO

Un punto di forza dei modelli connessionisti è la loro capacità di apprendere. Quanto è simile l'apprendimento umano a quello che avviene in una rete neurale? La risposta a questa domanda è molto sfaccettata. Ad esempio, quando una rete deve apprendere più liste di stimoli si verifica una «interferenza catastrofica» che non si osserva nei soggetti umani. Tuttavia, queste differenze scompaiono in modelli neurali più realistici dell'apprendimento che utilizzano due sistemi complementari di memoria, l'ippocampo e la neocorteccia (per una rassegna si veda O'Reilly, 2005).

L'idea che i modelli connessionisti simulino bene l'apprendimento umano è comunque piuttosto consolidata in psicologia ed è collegata ai risultati di numerose ricerche che dimostrano fenomeni di «apprendimento statistico» nell'uomo, perfino in bambini di pochi mesi (ad esempio, Saffran, Aslin e Newport, 1996). Hutzler, Ziegler, Perry, Wimmer e Zorzi (2004) si sono chiesti se l'apprendimento nei modelli connessionisti della lettura descriva in modo adeguato lo sviluppo delle abilità di lettura nei bambini. Hutzler e colleghi si sono focalizzati su un fenomeno che hanno denominato «learning rate effect» (effetto della velocità di apprendimento): imparare a leggere un'ortografia relativamente irregolare come l'inglese è più difficile e più lento rispetto ad un'ortografia relativamente regolare come il tedesco (si veda la figura 4). Le reti neurali sono sensibili alle regolarità statistiche presenti nel set di stimoli utilizzati per l'apprendimento; è dunque lecito chiedersi se ciò può rendere conto delle differenze cross-linguistiche osservate nei bambini che imparano a leggere. Nel confronto tra modelli connessionisti che apprendevano a leggere l'inglese e il tedesco, Hutzler e colleghi hanno osservato una discrepanza fondamentale tra le simulazioni e i dati sui bambini. Ciò ha permesso di determinare che per catturare in modo adeguato il *pattern* di acquisizione delle capacità di lettura i modelli connessionisti devono essere sensibili non solo alle regolarità statistiche delle relazioni scritto-suono ma anche al modo in cui la lettura viene insegnata nei diversi paesi (vedi fig. 4).

Più in generale, questa ricerca dimostra che l'apprendimento statistico nelle reti neurali può non essere sufficiente a rendere conto dell'apprendimento umano. È importante introdurre nelle simulazioni un maggiore realismo riguardo ai fattori ambientali e culturali che possono influenzare l'apprendimento.

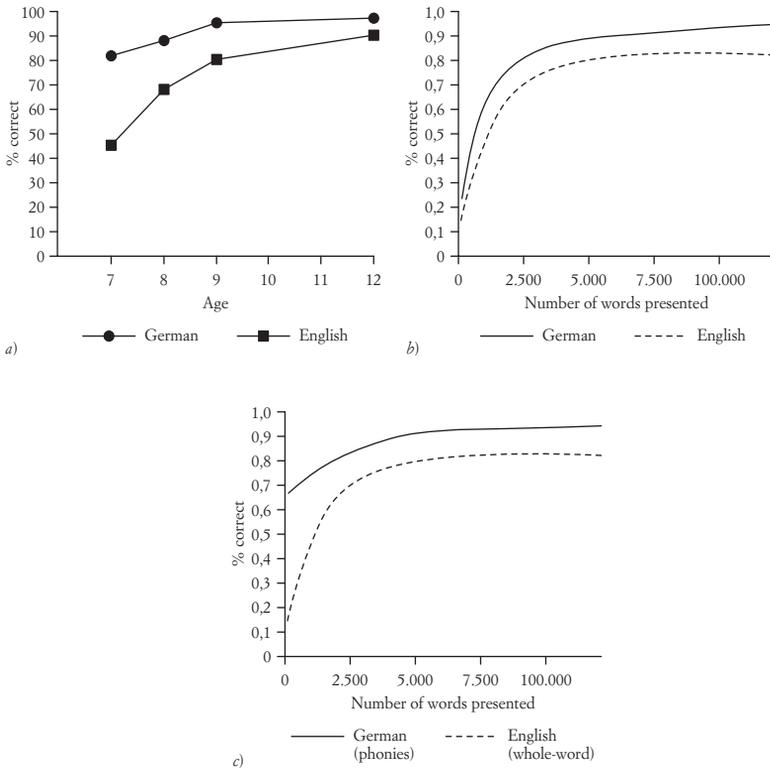


FIG. 4. Il *learning rate effect* (Hutzler *et al.*, 2004). **A**: Accuratezza nella lettura di parole in bambini inglesi e tedeschi. **B**: Accuratezza nei modelli di lettura inglese e tedesco durante un apprendimento standard. **C**: Accuratezza nei modelli di lettura inglese e tedesco quando l'apprendimento prende in considerazione anche i diversi metodi di insegnamento.

#### LE RAPPRESENTAZIONI

Uno dei fenomeni più noti e studiati nel dominio della cognizione numerica è l'effetto della grandezza del problema nell'aritmetica mentale (vedi Zorzi, 2004, per una rassegna): nella risoluzione di problemi aritmetici semplici (le cosiddette tabelline, ovvero le moltiplicazioni o addizioni con numeri a una cifra) le latenze nella risposta e gli errori dei soggetti aumentano in funzione della grandezza del problema. Tuttavia, le numerose ricerche non hanno ancora permesso di chiarire se l'aritmetica mentale sia basata su codici verbali/fonologici (Dehaene, Spelke, Pinel, Stanescu e Tsivkin, 1999) o su codici semantici specifici per dominio (McCloskey, 1992). I numeri sembrano essere rappresentati su una linea numerica mentale spazialmente orientata (Zorzi, Prif-

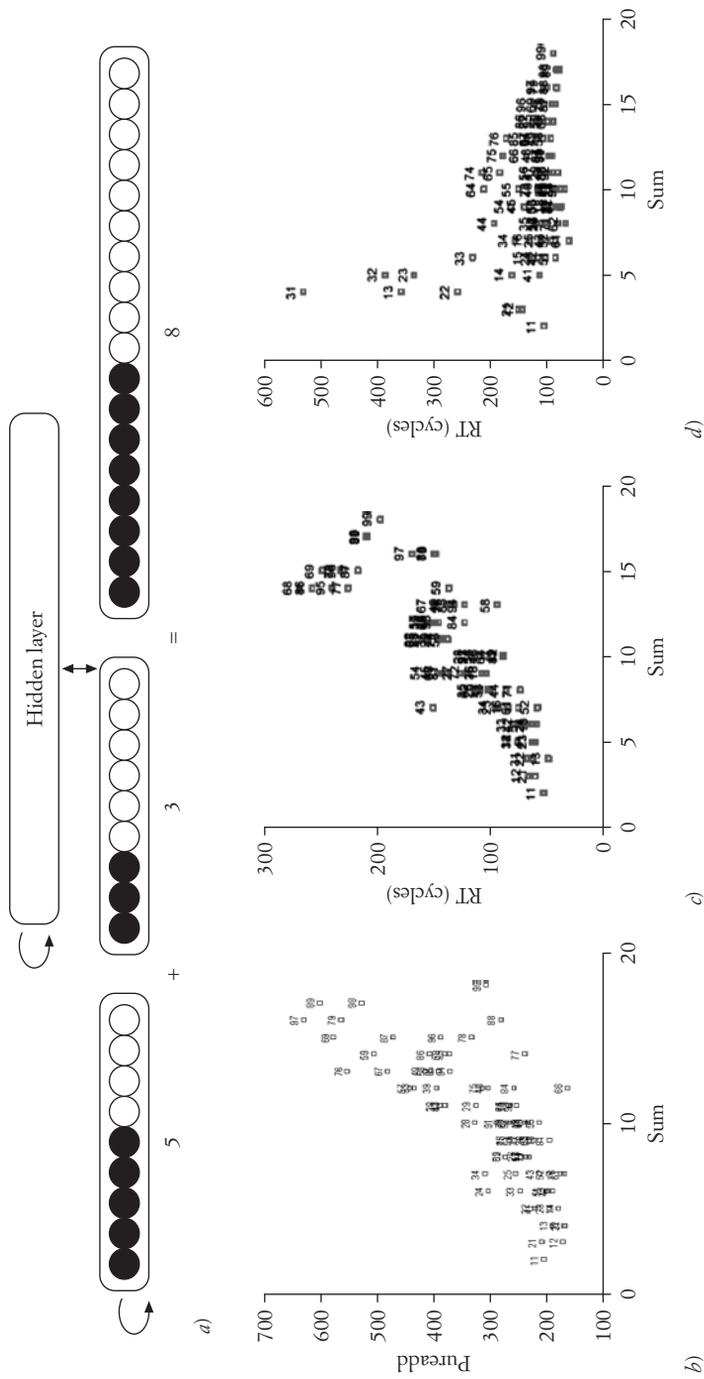


FIG. 5. L'effetto della grandezza del problema nell'aritmetica mentale. **A:** Architettura del modello connessionista (Zorzi *et al.*, 2005). Nella figura del modello i due operandi e il risultato (somma) sono codificati con una rappresentazione cardinale. **B:** Grafico della distribuzione dei tempi di risposta di soggetti adulti nell'esecuzione di addizioni con numeri ad una cifra, in funzione della grandezza del problema (somma). **C:** Risultati di una simulazione in cui il modello utilizza rappresentazioni dei numeri con una codifica cardinale. **D:** Risultati di una simulazione in cui il modello utilizza rappresentazioni dei numeri con una codifica posizionale.

tis e Umiltà, 2002), ma il formato specifico di queste rappresentazioni semantiche è controversa. Le diverse proposte vanno da quella di una compressione logaritmica dei numeri sulla linea numerica (Dehaene, 2003), a quella di una variabilità scalare che rende la rappresentazione «rumorosa» (Gallistel e Gelman, 1992), a quella di una codifica cardinale della numerosità (Zorzi e Butterworth, 1999).

In una serie di ricerche computazionali (per una rassegna si veda Zorzi, Stoianov e Umiltà, 2005) abbiamo quindi indagato quale fosse il formato delle rappresentazioni numeriche utilizzate per accedere ai fatti aritmetici immagazzinati nella memoria umana. Abbiamo quindi confrontato nelle simulazioni quattro diversi modelli di rappresentazione mentale dei numeri, utilizzando reti neurali che erano identiche in ogni dettaglio tranne che per il formato con cui venivano rappresentati i numeri. Le simulazioni hanno dimostrato che un solo tipo di rappresentazione, basata sulla codifica cardinale della numerosità, fa emergere l'effetto della grandezza del problema in un modello connessionista (vedi fig. 5).

Solo un modello tra quattro ha avuto successo nel descrivere accuratamente i dati umani. Queste ricerche dimostrano quindi l'importanza cruciale delle assunzioni sul formato delle rappresentazioni utilizzate in un modello computazionale.

#### L'ARCHITETTURA

Il termine *architettura* si riferisce allo schema di connettività che identifica l'eventuale presenza di gruppi di neuroni diversi ed il modo in cui i neuroni sono connessi tra di loro per formare una rete neurale. Discuteremo ora un fenomeno che è riconducibile a delle assunzioni di tipo architettonico: le asimmetrie emisferiche nella negligenza spaziale unilaterale (*neglect*). Il *neglect* è una sindrome neuropsicologica caratterizzata dalla mancata percezione ed esplorazione dello spazio controlaterale alla lesione (per una rassegna, si veda Vallar, 1998). Si manifesta più frequentemente e con maggiore intensità dopo lesioni emisferiche destre ed in particolare alla corteccia parietale posteriore (PPC), un'area coinvolta nell'elaborazione di rappresentazioni spaziali che legano la sensazione all'azione (Colby e Goldberg, 1999).

Esistono varie ipotesi sui meccanismi che determinano le asimmetrie emisferiche nel *neglect*, ma la loro formulazione verbale non permette un confronto sistematico e quantitativo. Inoltre, la questione non può essere affrontata studiando il cervello dei primati non umani attraverso tecniche neurofisiologiche, perché l'asimmetria è presente solo nell'uomo. Di Ferdinando, Casarotti, Vallar e Zorzi (2005) hanno

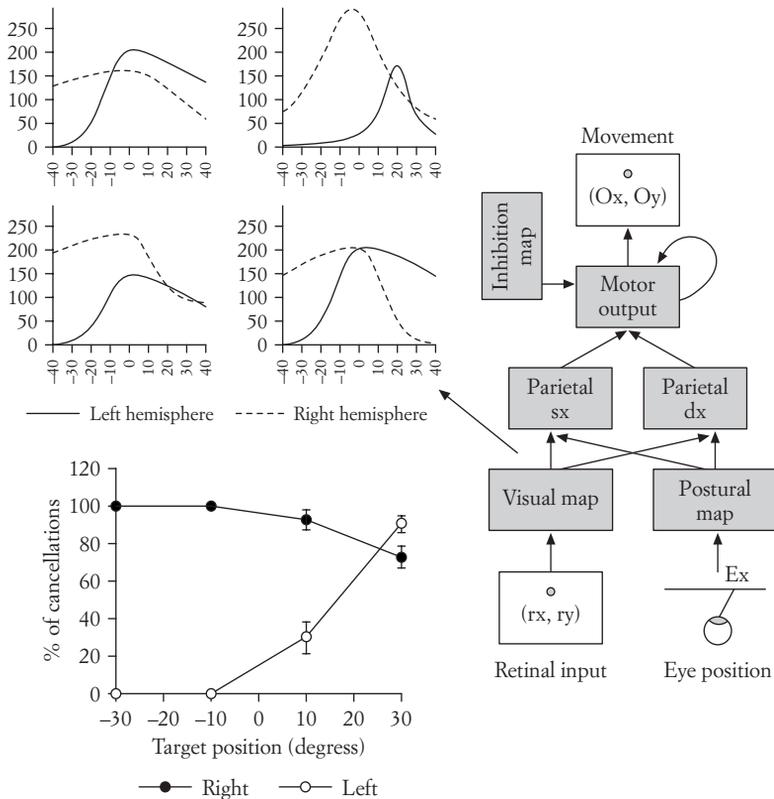


FIG. 6. Modello neurocomputazionale delle trasformazioni sensorimotorie (Di Ferdinando *et al.*, 2005). L'architettura del modello è visibile nella parte destra della figura, mentre nel riquadro in alto a sinistra sono mostrati quattro diversi pattern di connettività tra l'input visivo e le mappe spaziali nei lobi parietali. Il riquadro in basso a sinistra mostra la prestazione del modello rivelandosi più adeguato nel simulare i sintomi del neglect dopo una lesione parietale. La simulazione di un compito in cui è necessario marcare dei target disposti in modo casuale in tutto lo spazio visivo (test di Albert) mostra un comportamento marcatamente asimmetrico: dopo una lesione destra il modello omette sistematicamente i target posti nello spazio sinistro, mentre una lesione sinistra ha un effetto quasi nullo sulla prestazione.

sviluppato un modello computazionale neurofisiologicamente plausibile dei circuiti cerebrali e dei processi computazionali coinvolti nelle trasformazioni sensorimotorie, con il quale poter implementare varie teorie del *neglect* ed in particolar modo della sua natura asimmetrica. Sono state condotte una serie di simulazioni usando reti neurali artificiali che pur condividendo lo stesso scenario simulativo si proponevano di implementare quattro diverse teorie del *neglect* (vedi figura 6). Le diverse assunzioni teoriche riguardavano il gradiente di connet-

tività tra le mappe visive e le mappe spaziali: ad esempio, una delle teorie più note (Heilman, Bowers, Valenstein e Watson, 1987) sostiene che la corteccia parietale destra contiene una rappresentazione dello spazio bilaterale, mentre quella sinistra rappresenta solo lo spazio controlaterale. Il comportamento delle diverse reti neurali è stato quindi confrontato in due classici test clinici usati per i pazienti con *neglect*, ossia il compito di bisezione di linee e quello di cancellazione. Il confronto è stato fatto sia per reti «normali», ossia non lesionate, sia per reti in cui era stata lesionata una delle due aree parietali, per verificare in quali condizioni emerge un'asimmetria tra lesione sinistra e lesione destra. Dalle analisi condotte è emerso che solo una delle teorie è in grado di render conto in modo adeguato sia della prestazione deficitaria dei pazienti che delle asimmetrie tra le lesioni ai due diversi emisferi. In particolare, le simulazioni indicano che la variabile più importante per spiegare le asimmetrie nel *neglect* è la dominanza dell'emisfero destro nella rappresentazione dello spazio, espressa in termini di un maggior numero di neuroni coinvolti nei compiti spaziali.

In generale, il modello computazionale di Di Ferdinando e colleghi (2005) ha permesso di analizzare in modo rigoroso il ruolo dei diversi *pattern* di connettività nell'asimmetria che caratterizza il *neglect* e di aggudicare quindi tra le diverse teorie in competizione.

#### LA PLAUSIBILITÀ BIOLOGICA

L'approccio simulativo in psicologia si basa sull'assunzione che tutti gli aspetti importanti della cognizione umana possano essere catturati in un modello computazionale. A sua volta, questa assunzione si basa sulla convinzione che, oltre un certo livello, i dettagli dell'implementazione siano irrilevanti per la cognizione. Per i ricercatori che si rifanno all'approccio dei sistemi simbolici (Newell e Simon, 1976), il cervello e i neuroni non hanno alcun status preferenziale come substrato della mente. Il connessionismo pone l'accento su alcuni principi della computazione neurale (O'Reilly, 1998), ma ne trascura altri che potrebbero rivelarsi di importanza fondamentale.

I modelli connessionisti «classici» si rivelano spesso implausibili sia per quanto riguarda l'architettura neurale che per le procedure di apprendimento utilizzate. Ad esempio, la grande maggioranza dei modelli sviluppati nel corso degli anni Ottanta e Novanta del secolo scorso utilizzano un'architettura di tipo *feed-forward* (l'attivazione si propaga solo dall'input verso l'output), che permette un'elaborazione dell'informazione esclusivamente dal basso verso l'alto (*bottom-up*). Inoltre, la maggior parte dei modelli connessionisti dei processi co-

gnitivi sono basati su reti neurali addestrate attraverso varianti della *back-propagation* (Rumelhart, Hinton e Williams, 1986), un algoritmo di apprendimento supervisionato molto potente ma biologicamente implausibile perché richiede che il segnale d'errore sia propagato all'indietro, dal dendrite del neurone ricevente, attraverso la sinapsi, verso l'assone del neurone trasmittente, e successivamente integrato e moltiplicato per la forza della sinapsi e la derivata della funzione di output (O'Reilly e Munakata, 2000). Molti ricercatori hanno scelto di ignorare questo problema, considerandolo un compromesso accettabile per ottenere un apprendimento efficiente di compiti cognitivi complessi. Un ulteriore punto debole è che l'algoritmo *back-propagation* richiede sempre la presenza di un insegnante esterno, ovvero di un vettore di output desiderato per il calcolo del segnale di errore, che è utilizzato per aggiustare gradualmente i pesi fino a quando le prestazioni della rete raggiungono un livello considerato adeguato in base a qualche criterio esterno (Hinton, 1999). Si noti che l'apprendimento supervisionato ripropone una visione sostanzialmente associazionista (stimolo-risposta) dell'apprendimento.

Esistono tuttavia dei modelli neurali caratterizzati da una maggiore aderenza ai principi della computazione corticale e che apprendono a «rappresentare il mondo» non attraverso catene di associazioni stimolo-risposta ma «osservandolo» e costruendone dei *modelli interni*. Questi modelli, chiamati *modelli generativi*, oltre a rappresentare lo stato dell'arte delle ricerche di *machine learning*, rappresentano modelli plausibili dell'apprendimento corticale e danno forti indicazioni sul ruolo delle *connessioni a feedback* (o *top-down*) e delle *connessioni laterali* nella corteccia. La formulazione dei modelli generativi proviene da un approccio di tipo statistico, che fonda le sue radici nel teorema di Bayes. Un modello generativo avanza infatti delle ipotesi sui dati sensoriali (l'input), per poter modificare i parametri che lo caratterizzano (si veda Zorzi, 2006, per una discussione più approfondita). Il collegamento tra modelli generativi (nella loro accezione statistica) e il connessionismo deriva dallo sviluppo di reti neurali che apprendono senza supervisione (Hinton e Sejnowski, 1999). Un modello neurale generativo è un modello probabilistico di come i dati sensoriali siano causati da sottostanti proprietà fisiche del mondo, e rappresenta un'utile *funzione oggettiva* per l'apprendimento non supervisionato. L'apprendimento può essere visto come la massimizzazione della verosimiglianza dei dati osservati rispetto al modello generativo, il che equivale a scoprire modi efficienti di codificare l'informazione sensoriale (Hinton e Ghahramani, 1997). Il concetto di apprendimento non supervisionato non è certamente nuovo, ma gli algoritmi classici (basati su semplici varianti della regola di Hebb) sono stati raramente utilizzati per la simulazione di funzioni cognitive complesse a causa

dei loro limiti intrinseci. Alcuni algoritmi sviluppano rappresentazioni distribuite ma lineari (ad esempio, le componenti principali nei dati); la linearità implica una capacità di catturare l'informazione relativa alla covarianza *pairwise* delle variabili di input, mentre l'informazione di ordine superiore è invisibile alla rete. Altri sviluppano rappresentazioni non-lineari ma localistiche (ad esempio, i raggruppamenti nei dati scoperti dall'apprendimento competitivo) in cui si assume che ogni osservazione è generata da una singola unità (o variabile) nascosta. Gli sviluppi più importanti per l'apprendimento non supervisionato riguardano proprio la possibilità di ottenere modelli generativi probabilistici che scoprono rappresentazioni distribuite e non lineari dei dati di input.

I modelli neurali generativi vengono utilizzati da alcuni anni nel mio laboratorio per le simulazioni nell'ambito delle neuroscienze cognitive (per una rassegna si veda Zorzi, 2006). Durante l'apprendimento i pesi della rete neurale vengono modificati per minimizzare la discrepanza tra i dati «osservati» (input presentato alla rete) e i dati «ricostruiti» (corrispondenti all'input ma generati dal modello interno). La possibilità di ricostruire i *pattern* osservati si basa sulla presenza di connessioni a *feedback*, che permettono ai segnali delle unità nascoste di attivare (in modo top-down) le unità che codificano l'informazione sensoriale (input). In questi modelli, come in ogni rete neurale completamente ricorrente (vedi Hopfield, 1982), i *pattern* appresi corrispondono a dei minimi energetici denominati *attrattori*. Le proprietà dinamiche della rete determinano una propagazione delle attivazioni tra le unità della rete attraverso vari cicli di elaborazione in cui queste si modificano fino ad arrivare ad uno stato stabile (*convergenza* verso l'attrattore più vicino). Il numero di cicli necessario per raggiungere uno stato stabile fornisce un'utile analogia con i tempi di reazione, permettendo un confronto diretto con i dati comportamentali.

All'interno di questa cornice teorica è possibile sviluppare modelli neurali dei processi cognitivi che permettono l'integrazione di vincoli funzionali (livello comportamentale) e strutturali (livello neurale). Ad esempio, in un modello che apprende le trasformazioni sensomotorie necessarie per la programmazione di movimenti oculari saccadici e di movimenti di raggiungimento verso bersagli visivi, le rappresentazioni interne sviluppate dalle unità nascoste della rete mostrano proprietà simili a quelle dei neuroni reali riportate in studi neurofisiologici (Zorzi, Casarotti e Cutini, 2004). La risposta di un neurone (reale o simulato) ad uno stimolo visivo che cade nel suo campo recettivo è modulata da segnali che trasmettono informazioni posturali, come la posizione dell'occhio e del capo. Confrontando l'apprendimento di trasformazioni sensomotorie in modelli generativi e in modelli basati

sull'algoritmo di apprendimento *back-propagation* emergono delle importanti differenze qualitative: i modelli generativi sviluppano campi recettivi di forma più regolare e maggiormente sensibili alla modulazione da parte di segnali posturali. Si noti infine che un'architettura ricorrente con connessioni a *feedback* produce sistematici effetti top-down. Nel caso di compiti sensorimotori, un programma motorio dovrebbe produrre un *pattern* d'attivazione nelle unità sensoriali della rete. Ciò si traduce in fenomeni di priming attentivo a partire dalla programmazione del movimento, con una precisa relazione topografica tra le regioni dello spazio che rappresentano i bersagli della programmazione motoria e le corrispondenti aree retiniche, in linea con le predizioni della teoria premotoria dell'attenzione (Casarotti, Zorzi e Umiltà, 2005).

I modelli generativi possono quindi essere utilizzati con successo nella simulazione dei processi cognitivi e permettono l'integrazione di vincoli funzionali (livello comportamentale) e strutturali (livello neurale). Un maggiore grado di plausibilità biologica dei modelli connessionisti trova riscontro nella buona corrispondenza tra i dati simulati ed i dati empirici a diversi livelli di descrizione, dalla psicologia alle neuroscienze.

#### CONCLUSIONI

La simulazione dei processi cognitivi umani con modelli computazionali è un approccio che sta assumendo un peso sempre maggiore nella ricerca psicologica. Nella prima parte di questo lavoro ho descritto le fasi fondamentali della modellizzazione e i criteri di valutazione di un modello. Nella seconda parte invece ho cercato di dimostrare la necessità di valutare in modo stringente (e comparativo) il ruolo dei principi e delle assunzioni incorporate nei modelli riguardo ai processi, all'apprendimento, alle rappresentazioni e all'architettura. Tutti questi fattori si intersecano con quello della plausibilità biologica, ovvero il grado di maggiore o minore aderenza del modello ai principi della computazione corticale.

Per quanto riguarda il rapporto tra l'approccio computazionale in psicologia e la ricerca sperimentale è utile sottolineare l'importanza di identificare gli effetti critici che permettono di dissociare tra classi di modelli. Questo può portare a nuovi esperimenti che mettano alla prova i modelli, in uno spirito falsificazionista e di inferenza forte. Va tuttavia ricordata la necessità di eseguire le simulazioni per ricavare predizioni teoriche accurate dai modelli (che sono spesso di natura complessa e non-lineare) e per poter quindi aggiudicare tra teorie in competizione.

Le ricerche qui descritte sono state finanziate da Università di Padova (Progetto di Ateneo, 2003), MIUR (PRIN, 2004) e Commissione Europea (MRTN-CT-2003-504927 NUMBRA).

## BIBLIOGRAFIA

- CASAROTTI M., ZORZI M., UMITÀ C. (2005). A recurrent model for spatial attention. In B.G. Bara, L. Barsalou, M. Bucciarelli (eds.), *Proceedings of the Twenty-seventh Annual Meeting of the Cognitive Science Society*. Mahwah, N.J.: Erlbaum.
- CHALMERS D.J., FRENCH R.M., HOFSTADTER D.R. (1992). High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental & Theoretical Artificial Intelligence*, 4, 185-211.
- COLBY C.L., GOLDBERG M.E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22, 319-349.
- COLTHEART M., RASTLE K., PERRY C., LANGDON R., ZIEGLER J.C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204-256.
- Dehaene S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, 7, 145-147.
- DEHAENE S., SPELKE E., PINEL P., STANESCU R., TSIVKIN S. (1999). Sources of mathematical thinking: Behavior and brain-imaging evidence. *Science*, 284, 970-974.
- DI FERDINANDO A., CASAROTTI M., VALLAR G., ZORZI M. (2005). Hemispheric asymmetries in the neglect syndrome: A computational study. In A. Cangelosi, G. Bugmann, R. Borisjuk (eds.), *Modeling language, cognition and action*. Singapore: World Scientific.
- GALLISTEL C.R., GELMAN R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44, 43-74
- HEILMAN K.M., BOWERS D., VALENSTEIN E., WATSON R.T. (1987). Hemisphere and hemispatial neglect. In M. Jeannerod (ed.), *Neurophysiological and neuropsychological aspects of spatial neglect*. Amsterdam: North-Holland, pp. 115-150.
- HINTON G.E. (1999). Supervised learning in multilayer neural networks. In R.A. Wilson, F.C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.
- HINTON G.E., GHARAHAMI Z. (1997). Generative models for discovering sparse distributed representations. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 352, 1177-1190.
- HINTON G.E., SEJNOWSKI T. (1999). *Unsupervised learning: Foundations of neural computation*. Cambridge, MA: MIT Press.
- HOPFIELD J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA*, 79, 3088-3092.
- HOUGHTON G., ZORZI M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology*, 20, 115-162.
- HUTZLER F., ZIEGLER J.C., PERRY C., WIMMER H., ZORZI M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition*, 91, 273-296.

- JACOBS A.M., GRAINGER J. (1994). Models of visual word recognition – Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1311-1334.
- MCCLOSKEY M. (1992). Cognitive mechanisms in number processing: Evidence from acquired dyscalculia. *Cognition*, 44, 107-157.
- NEWELL A., SIMON H.A. (1976). Computer science as empirical inquiry: Symbols and search. Reprinted in J.L. Garfield (ed.), *Foundations of cognitive science: The essential readings*. New York: Paragon House, 1990, pp. 113-138.
- O'REILLY R.C. (1998). Six principles for biologically-based computational models of cortical cognition. *Trends in Cognitive Sciences*, 2, 455-462.
- O'REILLY R.C. (2005). The division of labor between the neocortex and hippocampus. In G. Houghton (ed.), *Connectionist modeling in cognitive psychology*. London: Psychology Press.
- O'REILLY R.C., MUNAKATA Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.
- PARISI D. (2001). *Simulazioni. La realtà rifatta nel computer*. Bologna: Il Mulino.
- PLAUT D.C., MCCLELLAND J.L., SEIDENBERG M.S., PATTERSON K.E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domain. *Psychological Review*, 103, 56-115.
- RASTLE K., COLHEART M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 482-503.
- RUMELHART D.E., MCCLELLAND J.L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.
- RUMELHART D.E., HINTON G.E., WILLIAMS R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart, J.L. McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press, pp. 318-362.
- SAFFRAN J.R., ASLIN R.N., NEWPORT E.L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- SEIDENBERG M.S., MCCLELLAND J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- VALLAR G. (1998). Spatial hemineglect in humans. *Trends in Cognitive Sciences*, 2, 87-97.
- ZORZI M. (2000). Serial processing in reading aloud: No challenge for a parallel model. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 847-856.
- ZORZI M. (2004). Pensare con i numeri. In M. Zorzi, V. Girotto (a cura di), *Fondamenti di Psicologia Generale*. Bologna: Il Mulino, pp. 319-329.
- ZORZI M. (2005). Computational models of reading. In G. Houghton (ed.), *Connectionist models in cognitive psychology*. London: Psychology Press.
- ZORZI M. (in corso di stampa). Dai neuroni al comportamento: la simulazione dei processi cognitivi con modelli generativi. *Sistemi Intelligenti*.
- ZORZI M., BUTTERWORTH B. (1999). A computational model of number comparison. In M. Hahn, S.C. Stoness (eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*. Mahwah, N.J.: Erlbaum, pp. 778-783.
- ZORZI M., CASAROTTI M., CUTINI S. (2004). La programmazione del movimento nella corteccia parietale posteriore: un approccio computazionale.

*Atti del Congresso Nazionale dell'Associazione Italiana di Psicologia (Sezione di Psicologia Sperimentale)*, Sciacca (AG).

- ZORZI M., HOUGHTON G., BUTTERWORTH B. (1998a). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1131-1161.
- ZORZI M., HOUGHTON G., BUTTERWORTH B. (1998b). The development of spelling-sound relationships in a model of phonological reading. *Language and Cognitive Processes*, 13, 337-371.
- ZORZI M., PRIFTIS K., UMILTÀ C. (2002). Neglect disrupts the mental number line. *Nature*, 417, 138-139.
- ZORZI M., STOIANOV I., UMILTÀ C. (2005). Computational modeling of numerical cognition. In J. Campbell (ed.), *Handbook of mathematical cognition*. New York: Psychology Press.

*La corrispondenza va inviata a Marco Zorzi, Dipartimento di Psicologia Generale, Via Venezia 8, 35131@ Padova, e-mail: marco.zorzi@unipd.it, Web: <http://ccnl.psy.unipd.it>.*

