

Dott.ssa Tiziana Giancesini

Dip di Scienze Biomediche e Biotecnologie

Università degli Studi di Brescia

Viale Europa, 11-25100-Brescia

giancesin@med.unibs.it

DISPENSA PER IL CORSO DI PSICOLOGIA DELLA PERCEZIONE

LEZIONE DEL 25-05-06

La Ricerca visiva

Trovare un oggetto specifico fra molti altri nell'ambiente è una delle azioni più comuni della nostra vita quotidiana.

In laboratorio un compito di ricerca visiva richiede tipicamente ai soggetti di cercare uno stimolo bersaglio (target) fra un certo numero di stimoli distrattori.

Questo paradigma semplice permette ai ricercatori di esaminare come gli stimoli visivi si differenzino tra loro, che proprietà di uno stimolo attraggono l'attenzione, come l'attenzione si sposti da un oggetto al seguente e così via. Il paradigma di ricerca visiva è stato usato estesamente. In laboratorio si usano tipicamente stimoli artificiali (linea colorata segmenti, lettere, ecc). Eppure, questi compiti si approssimano alla ricerche visive che si fanno tutti i giorni, ad esempio la ricerca di un dente di leone in mezzo ad un prato di fiori gialli o la ricerca di una via alla guida dell'auto in una zona non conosciuta nel buio della notte. Un esempio di utilizzato in laboratorio è mostrato nella figura seguente.

X X X X X X

X X X X X X

X X X X X X

X X X X X X

Il compito del soggetto è rilevare la presenza o meno di una X verde all'interno della configurazione. Dovreste in primo luogo notare la X verde che sembra "saltare fuori" (POP OUT) dall'allineamento. Questo è un esempio di una ricerca facile ed efficiente.

Ora si provi ad individuare di nuovo la X verde nella figura seguente. È un esempio di un tipo più difficile e più inefficiente di ricerca.

X X O X X X

X O X O X O

X O X X X X

O X O X O X

Il numero totale di elementi nell'esposizione è conosciuto come la numerosità dell'insieme (set size). Il target è presentato su una certa percentuale delle prove, in genere 50%. I soggetti premono un tasto se il target è presente e un altro tasto se compaiono soltanto i distrattori.

Il soggetto deve rispondere il più rapidamente e accuratamente possibile.

Viene quindi utilizzato il tempo di reazione (reaction time, RT) sia l'accuratezza per misurare la prestazione del soggetto nel rilevare la presenza o l'assenza di un certo target.

Nei compiti in cui vengono utilizzati i RT, l'esposizione è solitamente presente fino all'emissione di una risposta. Nei compiti in cui viene utilizzata l'accuratezza, l'esposizione è solitamente molto breve, seguito da una mascherina visiva interferente. A seconda dei casi gli stimoli utilizzati possono essere estremamente semplici, come brevi barrette luminose di orientamento variabile, o più complessi, come lettere dell'alfabeto, numeri o addirittura piccole immagini di oggetti reali. Quando gli stimoli utilizzati

rendono sufficientemente difficile l'esecuzione del compito, si osserva che il tempo necessario al soggetto per fornire la risposta cresce in modo pressoché lineare al crescere del numero totale degli stimoli. Questo risultato è stato tradizionalmente interpretato in base ad un modello seriale, cioè in base all'idea che un "faro" attenzionale debba ispezionare uno alla volta ciascuno stimolo presente. A sua volta questo potrebbe dipendere dal fatto che la corretta identificazione di certi stimoli richiede l'intervento dell'attenzione focale, ovvero l'utilizzazione di una notevole quota delle risorse computazionali (cognitive) disponibili su una porzione limitata del campo visivo. In alternativa, altri studiosi del problema hanno proposto che l'aumento lineare del tempo di risposta col numero degli stimoli presenti può essere spiegato nei termini di un modello parallelo competitivo. In breve, secondo un modello di questo tipo tutti gli stimoli presenti nella scena verrebbero elaborati simultaneamente. Tuttavia, all'aumentare del numero degli stimoli si avrebbe un frazionamento sempre maggiore delle risorse computazionali totali e una conseguente diminuzione delle risorse assegnate a ciascuno stimolo. Questo a sua volta porterebbe a un rallentamento del tempo medio necessario a completare l'analisi percettiva di tutti gli stimoli, così da ritardare il momento in cui il soggetto può esprimere la sua decisione.

I meccanismi della ricerca visiva e dell'attenzione più in generale possono essere ottenute esaminando l'efficienza dei compiti di ricerca. Ci sono diversi modi per misurare l'efficienza di ricerca. Il metodo più comune è variare il numero di elementi nell'esposizione (variare il set size) e nel RT in funzione del formato dell'insieme. La pendenza (slope) delle funzioni stabilite dai RTs è una misura di efficienza di ricerca. Una pendenza di msec/item zero indica che il target, quando presente, è rilevato senza interferenza da parte dei distrattori. Le pendenze più ripide indicano che la ricerca è meno efficiente e che viene "pagato un costo" per ogni distrattore supplementare. Per i compiti di ricerca più efficienti le pendenze tendono a variare da 0 msec/item (per esempio una ricerca di un target rosso fra distrattori verdi) per elemento fino a 20-30 millisecondi per elemento (nelle prove target presenti), dove per convenzione cominciano ad essere definite ricerche inefficienti (per esempio una ricerca di una vocale fra le consonanti).

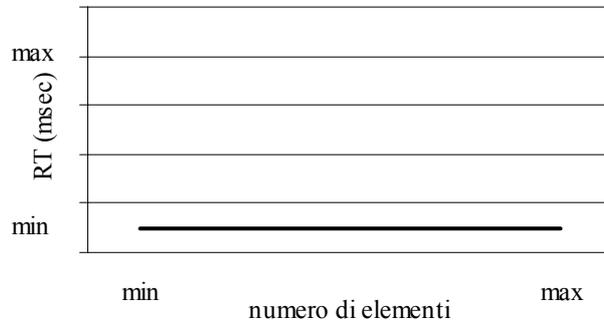


Grafico di un tipico risultato in un esperimento di numerosità con compito di pop-out, cioè con target semplice da rilevare. La funzione risulta piatta, cioè il TR (in ms) non risente del numero di elementi nello schermo. Ciò indica che il meccanismo implicato è preattentivo, dunque in parallelo.

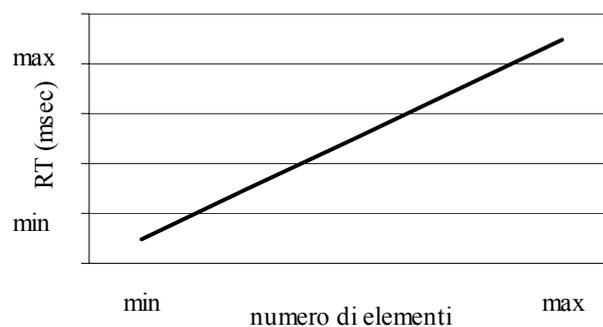


Grafico di un tipico risultato in un esperimento di numerosità con compito di conjunction, cioè con target difficile da rilevare. La funzione risulta inclinata, cioè il TR (in ms) aumenta all'aumentare del numero degli stimoli. Ciò indica che il meccanismo implicato è attenzionale, ma alcuni autori lo considerano seriale, altri in parallelo a risorse limitate.

Si possono trovare pendenze molto più ripide per compiti più complessi (per esempio si immagini di provare a trovare una serie di matrici di 16 puntini fra delle serie di matrici di 17 puntini). Le misure di accuratezza sono il secondo metodo comune per la misura delle prestazioni di ricerca. Le ricerche efficienti producono livelli elevati di accuratezza indipendentemente dalla numerosità anche quando l'esposizione è molto breve. Per i compiti meno efficienti l'accuratezza scende all'aumentare della numerosità (a meno che il tempo di esposizione aumenti).

Che cosa determina l'efficienza della ricerca visiva? Esiste una differenza qualitativa o soltanto quantitativa fra ricerca efficiente e ricerca inefficiente? Esistono vaste revisioni in letteratura ma gli scopi attuali, alcuni principi di base basteranno.

La teoria dell'integrazione delle caratteristiche di Treisman (Feature Integration Theory, FIT) (Treisman, 1988; Treisman & Gelade, 1980; Treisman & Sato, 1990)

La FIT di Treisman prevede: un primo stadio, *preattentivo*, nel quale le caratteristiche elementari (*features*) di un oggetto sono registrate *in parallelo* su *tutta* la scena visiva, senza l'uso dell'attenzione ed un secondo stadio, *attentivo*, nel quale tali caratteristiche sono combinate *in serie*, in una *parte* soltanto della scena, tramite l'attenzione focale. L'attenzione interviene per legare (*conjoin*) insieme le differenti caratteristiche (es colore e forma) in un singolo oggetto. Senza l'attenzione, esse verrebbero elaborate lo stesso, ma non potrebbero in nessun modo portare alla rappresentazione unitaria dell'oggetto.

I due stadi appena menzionati si possono rilevare rispettivamente con due tipi di compito: un primo compito, detto di *pop-out*, nel quale il target è definito da una singola caratteristica, es. rilevare un cerchio rosso tra cerchi verdi; un secondo compito, detto di *conjunction search* o *serial search*, nel quale il target è definito dalla combinazione di più caratteristiche, es. rilevare

una barra verticale rossa tra barre verticali blu e barre orizzontali rosse (fig.2).

La FIT prevede che la funzione di ricerca per le prove target assenti e per le prove target presenti abbia un rapporto 2:1

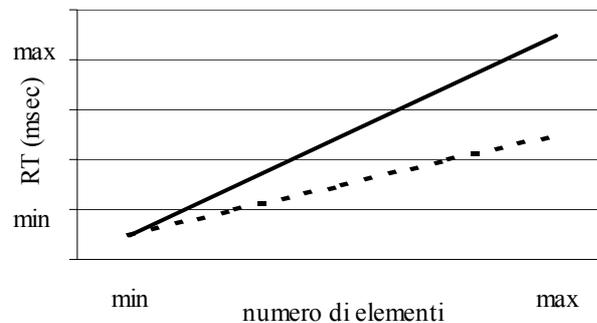


Grafico che mostra il rapporto tra pendenze 2:1, ovvero una funzione inclinata nel caso di target assente e una funzione inclinata della metà nel caso di target presente.

Questo è la logica dell'ipotesi di ricerca seriale *autoterminante*: la ricerca viene interrotta non appena trovato il target; dunque, se il target è presente lo si individua *in media* dopo aver controllato la metà degli elementi, mentre se è assente bisogna sempre controllarli tutti.

Anche se la FIT ha una struttura elegante che ha stimolato molto lavoro nel campo, la dicotomia rigorosa fra compiti paralleli e di ricerca seriale non è stata sostenuta chiaramente nei dati raccolti successivamente.

In alternativa, altri studiosi del problema hanno abbandonato la distinzione di serial/parallelo e hanno proposto che l'aumento lineare del tempo di risposta col numero degli stimoli presenti può essere spiegato nei termini di un modello parallelo competitivo. In breve, secondo un modello di questo tipo (limited capacity) tutti gli stimoli presenti nella scena verrebbero elaborati simultaneamente. Tuttavia, all'aumentare del numero degli stimoli si avrebbe un frazionamento sempre maggiore delle risorse computazionali totali e una conseguente diminuzione delle risorse assegnate a ciascuno stimolo. Questo a sua volta porterebbe a un rallentamento del tempo medio

necessario a completare l'analisi percettiva di tutti gli stimoli, così da ritardare il momento in cui il soggetto può esprimere la sua decisione.

modello limited capacity (Duncan e Humphreys 1989)

Questo modello riesce a spiegare la variazione in modo continuo della pendenza, che varia da poche unità a centinaia di ms per elemento in base alla semplice combinazione di tre fattori: la somiglianza tra distrattori (N-N); la somiglianza dei distrattori rispetto al target (N-T); il numero totale di elementi. Ad un estremo, se i distrattori sono uguali tra loro e diversi dal target, il numero di elementi non conta; si tratta del compito pop-out. All'estremo opposto, se i distrattori sono diversi tra loro e simili al target, il numero di elementi conta perché prolunga la competizione; si tratta del compito di conjunction.

Anziché supporre una dicotomia tra ricerca parallela e seriale, il modello spiega quell'ampia e continua variabilità nell'efficienza di ricerca a seconda del grado di somiglianza tra stimoli.

Ad uno stadio parallelo di raggruppamento percettivo (*grouping*) e di descrizione degli elementi, segue un'interazione competitiva tra input la quale guida l'accesso selettivo alla coscienza e poi all'azione. In generale, in entrambi i compiti, un input guadagna attivazione a seconda del grado di somiglianza rispetto al target, secondo un confronto tra la descrizione dell'informazione in entrata e l'informazione interna. Per via del *grouping*, conta molto di più la somiglianza tra N-N anziché la differenza tra N-T, nel senso che solo nel primo caso gli stimoli tendono a venir raggruppati percettivamente nello schermo.

Detto in altri termini, la prestazione facilitata nel compito pop-out dipende dal fatto che le singole features specificano un sottoinsieme di ricerca, cosa che velocizza di gran lunga la ricerca stessa. la ricerca visiva agisce in parallelo attraverso una finestra che definisce il campo attentivo e che raggruppa gli elementi in base ai semplici principi di somiglianza e prossimità spaziale. In seguito, la ricerca elimina in modo ricorsivo le porzioni di campo dove è già stato effettuato il *grouping* dei distrattori. Secondo il modello dunque, la performance non è influenzata dal numero di distrattori, bensì dal numero di gruppi di distrattori che si devono formare.

Attraverso l'implementazione del modello, gli autori sono riusciti a dimostrare tre punti. In primo luogo, che appunto si può implementare un modello computazionale della ricerca visiva in grado di codificare compiti di conjunction in parallelo anziché in serie; in grado inoltre di generare funzioni sia piatte sia inclinate e, per queste ultime, nel classico rapporto 2:1 a seconda che il target sia assente o presente rispettivamente. In secondo luogo, gli autori hanno evidenziato che la pendenza non è un ragionevole indice di ricerca seriale, poiché non contano gli stimoli ma i gruppi che possono formarsi, gruppi poi selezionati come un singolo percetto. In terzo luogo, benché secondo il modello la ricerca non sia seriale, il TR aumenta a seconda del numero di gruppi ricorsivamente selezionati ed eliminati; inoltre aumenta se occorre un controllo reiterativo dei gruppi che competono per la selezione. Durante il raggruppamento e il relativo controllo il modello ammette la possibilità di commettere errori.

In conclusione, secondo il modello in parallelo della competizione sbilanciata, anche nel caso di una funzione crescente, non serve ipotizzare meccanismi in serie; basta ipotizzare un meccanismo in parallelo a risorse limitate, che risente dell'aumento del numero degli elementi.

Il modello della ricerca guidata (Guided Search, Wolfe, 1989)

È un ibrido che deriva dalla teoria dell'integrazione delle features e dai modelli in parallelo. Il modello della ricerca guidata asserisce che i processi paralleli possono guidare l'attenzione focale seriale verso le posizioni dei probabili target mediante mappe di *features* semplici (come quelle di Treisman) e mappe d'attivazione complesse (come la master map di Treisman). Le mappe di *features* presentano un'attivazione basata su differenze locali (*bottom-up*) oppure su differenze specificate dal compito (*top-down*) che riguardano le somiglianze N-N o T-N. Le mappe di attivazione presentano invece un'attivazione che corrisponde alla somma pesata di quelle delle mappe di *features*. Ogni volta che lo stadio seriale comincia un nuovo ciclo di processamento, lo comincia dall'elemento con maggiore attivazione nella mappa d'attivazione e nel caso concluda che non è un target lo esclude da ulteriori cicli.

Nella ricerca visiva, cercando di sopperire alla limitazione delle risorse e sfruttando i segnali *top-down*, l'attenzione tenta d'individuare il più probabile target, diminuendo di conseguenza l'attivazione generale ed aumentando quella specifica per il target in questione; per esempio nella ricerca di una linea verticale rossa, i processi top-down attivano tutte le linee verticali e tutte le linee rosse. Quindi solo la posizione corrispondente al target nella mappa d'attivazione avrà una doppia attivazione. Ma a questo si aggiunge spesso un rumore di fondo che non corrisponde alla posizione del target e che richiederà una ricerca seriale di controllo. Il numero di passi sarà proporzionale alla numerosità dell'insieme.

Secondo il modello, la ricerca parallela, seriale e guidata non differiscono qualitativamente. I processi in parallelo agiscono sempre, ma nei compiti di pop-out il segnale proveniente dalle mappe di *features* è molto forte e l'attenzione viene subito diretta al target, mentre nei compiti di *conjunction* il segnale è debole e l'attenzione viene diretta a più di una posizione.

Ecco spiegato perché il modello assume uno stadio seriale e un rumore di fondo: per spiegare la seppur minima inclinazione della funzione nel compito di *conjunction*.

I processi in parallelo individuano gli stimoli potenzialmente simili al target, dopodiché i processi seriali li confrontano; i primi dividono l'insieme degli elementi in distrattori e in candidati target sulla base delle *features*, guidando così i secondi nel dirigere l'attenzione verso la combinazione che caratterizza il target.

Il modello spiega così quei dati che la teoria dell'integrazione di *features* non poteva spiegare: una pendenza molto bassa in certi compiti di *conjunction*; una ricerca in parallelo di *features* che restringe la ricerca generale a sottoinsiemi di elementi; e una prestazione migliore per tre *conjunction* (per esempio colore, forma e grandezza) rispetto a due. Nel caso della combinazione di tre *features*, la ricerca è più facile perché nel guidare l'attenzione ci sono tre processi in parallelo anziché due.

Dunque, questi dati potrebbero essere spiegati dalla teoria dell'integrazione di *features* qualora assumesse che i processi in parallelo generano informazioni utili per guidare gli spostamenti dell'attenzione.

A questo punto vale la pena di approfondire brevemente la tematica dei processi in parallelo così come vengono visti dalla teoria dell'integrazione di features e dal modello della ricerca guidata.

Nella teoria dell'integrazione di *features*, attraverso lo stadio parallelo comincia a cercare una caratteristica del target. Se il target è definito solo da una caratteristica, il soggetto dà subito una risposta; se invece è definito da una combinazione di caratteristiche, il soggetto non dà subito una risposta perché il solo stadio in parallelo non è sufficiente all'individuazione del target, per cui subentra lo stadio seriale.

D'altro canto, nel modello della ricerca guidata, lo stadio seriale consente di trovare il target molto più velocemente di quanto previsto dall'altra teoria proprio perché è guidato dallo stadio parallelo. Naturalmente le informazioni fornite dallo stadio parallelo non sono perfette, cioè esiste una variabilità nell'efficienza di tale stadio, ed è per questo che spesso il target non viene trovato immediatamente.

La ricerca di features è efficiente quando ci sono grandi differenze target-distrattori, meno efficiente quando tali differenze sono minori o quando ci sono differenze tra distrattori, ma in quest'ultimo caso ritorna efficiente se il target è definito da una sola caratteristica. Di diverso dagli altri modelli c'è che non occorre assumere per forza un processo seriale al fine di spiegare la ricerca del target nel compito di conjunction.

È chiaro che questo modello costituisce una sintesi dei modelli finora esposti, combinando in un'unica prospettiva i punti della teoria dell'integrazione di features e del modello in parallelo.