



**UNIVERSITÀ
DI PARMA**

DIPARTIMENTO DI MEDICINA E CHIRURGIA

CORSO DI LAUREA MAGISTRALE IN

PSICOBIOLOGIA E NEUROSCIENZE COGNITIVE

**MEMORIA A LUNGO TERMINE PER SCENE NATURALI: CONFRONTO TRA
DIVERSE STRATEGIE DI CONSOLIDAMENTO MNESTICO**

Relatore:

Chiar.ma Prof.ssa VERA FERRARI

Controrelatore:

Chiar.ma Prof.ssa ANNALISA PELOSI

Laureando:

DANIELE SPICA

ANNO ACCADEMICO 2022 - 2023

INDICE

ABSTRACT.....	5
INTRODUZIONE.....	6
1. LIVELLI DI ELABORAZIONE E GIUDIZIO DI PIACEVOLEZZA.....	7
1.1. Livelli di elaborazione.....	7
1.2. Livelli di elaborazione e materiale di studio verbale: giudizio di piacevolezza.....	13
1.3. Livelli di elaborazione e materiale di studio visivo.....	16
1.3.1. Livelli di elaborazione e stimoli visivi: volti	16
1.3.2. Livelli di elaborazione e stimoli visivi: scene naturali.....	18
2. TESTING EFFECT	22
2.1. Testing Effect: prospettiva storica e paradigmi.....	23
2.2. Testing Effect e materiale di studio visivo.....	26
2.3. Cornici teoriche proposte per la spiegazione del fenomeno.....	28
2.3.1. Elaborazione appropriata al trasferimento.....	28
2.3.2. Difficoltà desiderabile e sforzo di recupero.....	30
2.3.3. Il modello della biforcazione.....	31
3. LO STUDIO ATTUALE.....	33
3.1. Materiali e metodi	37
3.1.1. Partecipanti.....	37
3.1.2. Materiali e disegno.....	38
3.1.3. Procedura.....	40
3.1.4. Analisi statistica dei dati comportamentali.....	47
4. RISULTATI.....	47
4.1. Fase intermedia.....	47
4.2. Test finale.....	49

4.2.1. Accuratezza nel 2AFC rispetto al compito svolto in fase intermedia.....	49
4.2.2. Rating di fiducia nella fase di test finale.....	51
4.3. Effetti legati alla categoria dell'immagine.....	53
5. DISCUSSIONE.....	56
BIBLIOGRAFIA.....	61

ABSTRACT

La possibilità di modulare la memoria a lungo termine visiva è stata esplorata tramite una versione modificata del paradigma tipicamente utilizzato nell'indagine del "testing effect" (Roediger e Karpicke, 2006b). È stato manipolato il compito svolto in una fase intermedia sulle scene naturali precedentemente codificate, utilizzando due compiti appaiati in termini di "profondità di elaborazione" (Craik & Lockhart, 1972) ma differenti rispetto ai processi che coinvolgono: un rating di familiarità, che richiede un recupero episodico attivo della traccia di memoria e un rating di piacevolezza, che non richiede questo tipo di processo. La teoria dei livelli di elaborazione sostiene che la traccia di memoria vari in funzione della profondità di elaborazione, dove una maggiore profondità implica un maggior grado di analisi semantica e cognitiva (Craik & Lockhart, 1972). Il compito di piacevolezza è uno dei più utilizzati per garantire una codifica profonda in termini di analisi semantica e un miglioramento nel mantenimento di materiale di studio sia verbale che visivo. La letteratura sul testing effect mostra come lo svolgere un test vada a beneficio del mantenimento a lungo termine rispetto al ristudio del materiale (Roediger & Karpicke, 2006b) sia nel caso di stimoli verbali che visivi, con questo vantaggio tipicamente attribuito al recupero del materiale (Karpicke, 2017). Sono state presentate scene naturali sottoforma di immagini dinamiche in una fase di Encoding, con alcune di esse presentate nuovamente in una Fase intermedia e sottoposte ai due compiti, diversamente dalle rimanenti che venivano ripresentate solo nel test finale (*2AFC*) una settimana dopo. È stato rilevato un testing effect per le scene naturali: svolgere un compito intermedio sulle immagini conferiva un vantaggio mnestico nel test finale rispetto alle immagini non sottoposte ad alcun compito. Tuttavia, non sono state rilevate differenze nella performance di memoria finale legate all'aver svolto quello di familiarità o quello di piacevolezza. I risultati ottenuti sono coerenti con l'interpretazione per cui il vantaggio mnestico conferito dallo svolgere un compito intermedio sul materiale di studio non sia legato al suo recupero dalla memoria quanto alla profondità della ricodifica che avviene nella fase intermedia.

INTRODUZIONE

Il termine memoria viene utilizzato per designare diversi concetti. Tra i significati più frequentemente ricorrenti del termine vi è quello di una “capacità neurocognitiva di codificare, immagazzinare e recuperare l’informazione” (Tulving, 2000, p. 34).

È possibile approcciarsi alla memoria in diversi modi. Uno di questi è di concepirla come costituita da diverse componenti, in maniera coerente con il modello modale di Atkinson e Shiffrin (1968), secondo cui il sistema di memoria si declina in tre strutture distinte, con il flusso di informazioni che procede tra di esse in maniera seriale. Un’alternativa è quella di focalizzarsi sulle operazioni in essa coinvolte. In questa seconda visione rientrano la teoria dei livelli di elaborazione proposta da Craik e Lockhart (1972), in cui viene sottolineata la centralità della profondità a cui il materiale di studio viene elaborato per il mantenimento della traccia di memoria, e il concetto di testing effect (Roediger & Karpicke, 2006b) in cui viene posto l’accento sull’importanza del processo di recupero dell’informazione.

In questa tesi verrà presentato uno studio sulla memoria a lungo termine per scene naturali in cui è stato sfruttato uno strumento spesso utilizzato per garantire un’elaborazione profonda del materiale da ricordare, il giudizio di piacevolezza, all’interno di un paradigma utilizzato negli studi sul testing effect, al fine di confrontare la performance di memoria in un test finale legata al recupero del materiale di studio in un compito intermedio di familiarità con quella associata ad un compito intermedio che garantisca un’elaborazione ugualmente “profonda”, ovvero la valutazione di piacevolezza.

Prima di presentare lo studio, sarà introdotta la teoria dei livelli di elaborazione (Craik & Lockhart, 1972; Craik & Tulving, 1975) assieme ad alcuni risultati empirici di come il giudizio di piacevolezza costituisca un efficace strumento di memoria. Successivamente verrà affrontato il concetto di testing effect, in particolare in riferimento a materiale di studio visivo.

1. LIVELLI DI ELABORAZIONE E GIUDIZIO DI PIACEVOLEZZA

1.1. Livelli di elaborazione

La teoria dei livelli di elaborazione si deve a Craik e Lockhart (1972). Ciò che è stato proposto, è che nella percezione l'analisi dello stimolo prosegue lungo una serie di stadi. Quelli preliminari sarebbero coinvolti nell'analisi delle caratteristiche fisiche o sensoriali dello stimolo mentre quelli successivi contribuirebbero al riconoscimento e all'estrazione del significato. A seguito del riconoscimento, lo stimolo può andare incontro ad un ulteriore arricchimento, evocando ad esempio associazioni o immagini in maniera coerente con l'esperienza passata dei soggetti con il mondo. In questa gerarchia di stadi di elaborazione, alla quale ci si può riferire con i termini di "profondità di elaborazione", una maggiore profondità implicherebbe un maggior grado di analisi semantica o cognitiva. Secondo gli autori, essendo la traccia di memoria il risultato di questa analisi percettiva, la sua persistenza varierebbe in funzione della sua profondità, con livelli più profondi di analisi associati a tracce più elaborate, durature e forti. La memoria, quindi, sarebbe legata ai diversi livelli di analisi percettiva, con questi livelli che possono essere pensati come un continuum, dal prodotto dell'elaborazione sensoriale a quello di operazioni semantico-associative. Tra i vari fattori che possono influenzare la profondità di elaborazione vi sarebbero il tempo di elaborazione disponibile, il grado di attenzione rivolto allo stimolo ma anche la sua compatibilità con le strutture d'analisi. Infatti, stimoli molto familiari e significativi compatibili con le strutture cognitive esistenti, saranno processati ad un livello profondo più rapidamente rispetto a stimoli meno significativi e verranno ben mantenuti in memoria.

Qualche anno dopo la pubblicazione di questo articolo, Craik e Tulving (1975) hanno realizzato una serie di esperimenti volti a mettere alla prova la cornice teorica della profondità di elaborazione. Di seguito questi verranno presentati assieme alle loro implicazioni per la teoria proposta. In un primo esperimento, gli autori hanno mostrato ai loro soggetti una serie di parole,

ciascuna presentata per 200 millisecondi tramite un tachistoscopio e preceduta da diversi tipi di domande volte ad indurre o un'analisi superficiale delle parole, o un livello di analisi fonemico oppure un livello di analisi semantico, richiedendo ai soggetti, rispettivamente, di rilevare se fosse presente o meno una parola o di focalizzarsi sulle sue caratteristiche fisiche per determinare se fosse scritta in maiuscolo o minuscolo, se la parola presentata rimasse come un'altra parola, o facendo svolgere un compito di categorizzazione in cui dovevano indicare se la parola appartenesse ad una certa categoria o un compito di completamento della frase, in cui veniva presentata una frase con una parola mancante e chiesto se la parola presentata fosse appropriata a completarla. Dopo una serie di prove, ai soggetti fu fatto eseguire un test di riconoscimento per le parole presentate, del quale non erano stati informati, consegnando loro un foglio che conteneva quelle già incontrate assieme a delle parole simili che agivano da distrattori e chiedendo loro di indicare tutte quelle viste in precedenza. La performance dei soggetti nel compito di memoria risultò direttamente proporzionale alla profondità di elaborazione delle parole, con quelle sottoposte al livello di elaborazione più profondo, ovvero al compito semantico, associate ad una miglior performance, seguite da quelle sottoposte ad una elaborazione meno profonda, quella basata sulle caratteristiche fonemiche, mentre le parole sottoposte al livello di elaborazione più superficiale, in cui era sufficiente focalizzarsi sulle caratteristiche fisiche delle parole, risultarono quelle associate alla performance di memoria peggiore. Lo stesso pattern di risultati fu rilevato anche negli esperimenti successivamente svolti: nell'esperimento 2 incrementando il numero di distrattori nel test di riconoscimento, nell'esperimento 3 sostituendo il test di riconoscimento con un test di richiamo libero e nell'esperimento 4 facendo svolgere il test di richiamo libero a soggetti informati del successivo compito di memoria e quindi anche in condizioni di apprendimento intenzionale. In questi quattro esperimenti, e in quelli successivamente svolti, fu anche rilevato come la latenza di risposta aumentasse in maniera lineare all'aumentare della profondità di elaborazione. I soggetti, infatti, per eseguire il compito fonologico impiegavano più tempo rispetto a quello speso per il compito strutturale, ma meno tempo rispetto a quello richiesto dal compito semantico. Per escludere

che la performance di memoria potesse semplicemente riflettere queste differenze nel tempo di elaborazione, indipendentemente dalla profondità dell'analisi, fu quindi realizzato un quinto esperimento. In esso ai partecipanti furono presentate nuovamente delle parole sulle quali svolgere o il compito semantico di completamento della frase oppure un compito strutturale più complesso consistente nell'indicare se la parola presentata rispecchiasse un certo pattern di vocali e consonanti che l'aveva preceduta. Per esempio, la parola "rabbit" poteva essere preceduta dal pattern "CVCCVC", dove C indica una consonante e V una vocale. In questo caso la risposta da fornire era positiva. Successivamente i soggetti venivano sottoposti a un test di riconoscimento per il quale nuovamente non erano stati informati. Sebbene il compito strutturale richiese più tempo per essere svolto rispetto a quello semantico, fu nuovamente rilevata una migliore performance di memoria per le parole sottoposte al secondo compito rispetto al primo, consentendo così di concludere come sia la natura qualitativa del compito, e quindi la profondità di elaborazione, a determinare la performance di memoria e non il solo tempo di elaborazione. Un'altra ipotesi che è stata esclusa è quella secondo cui il vantaggio del compito semantico è in misura minore di quello fonologico sul compito strutturale potesse basarsi sulla similarità delle operazioni di codifica che caratterizzano quest'ultimo tipo di compito. Nel compito strutturale, infatti, ciascuna parola è preceduta dallo stesso tipo di domanda, a differenza degli altri due compiti in cui ogni parola è preceduta da una domanda diversa che potrebbe contribuire ad una codifica unica delle parole. L'ipotesi era che i soggetti potessero utilizzare ciò che ricordavano della domanda di codifica come un aiuto per recuperare la parola target e che la stessa domanda di codifica utilizzata per molte parole target come avviene nel compito strutturale, sarebbe stata meno efficace come indizio di recupero dal momento che non andava a specificare in maniera unica un evento codificato in memoria. L'unicità delle operazioni di codifica è stata variata nell'esperimento 8, manipolando la dimensione del set di domande di ciascun compito tra tre gruppi di partecipanti: a un primo gruppo furono fatte svolgere 4 domande nel compito strutturale, 16 in quello fonologico e 40 in quello semantico; ad un secondo 16 domande nel compito strutturale, 40 in quello fonologico e 4 in quello semantico; il terzo gruppo

rispose a 40 domande nel compito strutturale, a 4 in quello fonologico e a 16 in quello semantico. L'ipotesi prevedeva che, codifiche simili come quelle che caratterizzano il compito strutturale, ridotte in numero, dovessero essere codificate a loro volta in maniera unica e di conseguenza ricordate meglio. Di conseguenza, la performance di riconoscimento sarebbe migliorata utilizzando set di domande di dimensioni ridotte. Tuttavia, anche in questo esperimento fu rilevato nuovamente lo stesso pattern di risultati in tutti i gruppi di soggetti, pertanto la manipolazione della dimensione del set non andò ad eliminare le differenze legate alla profondità di elaborazione. Altri esperimenti furono realizzati per generalizzare i risultati ottenuti anche ad altre condizioni, tra cui un contesto meno controllato. Pertanto, nell'esperimento 9 fu presentata a soggetti frequentanti un corso di apprendimento e memoria una lista di parole, con il tempo di esposizione esteso a un secondo per ciascuna parola e con i soggetti informati del successivo test di riconoscimento. Nuovamente, per ciascuno stimolo i soggetti dovevano svolgere o il compito strutturale, o quello fonemico o quello semantico di categorizzazione. In maniera consistente, anche in queste condizioni meno controllate fu rilevato lo stesso pattern di risultati, con la performance di riconoscimento migliore per le parole sottoposte al compito semantico, seguite da quelle sottoposte al compito fonemico mentre la performance peggiore fu rilevata per le parole sottoposte al compito strutturale. Risultati identici emersero anche nelle stesse condizioni ma con un diverso gruppo di studenti. Nello stesso contesto di classe fu eseguita anche un'ulteriore manipolazione volta ad escludere una diversa ipotesi, ovvero che questi risultati potessero essere legati a fattori motivazionali e a verificare se i soggetti potessero raggiungere un'alta performance di riconoscimento per le parole sottoposte al compito strutturale se motivati da un forte incentivo. L'esperimento 10 fu quindi realizzato nelle stesse condizioni dell'esperimento 9, con la differenza che i soggetti venivano informati che gli sarebbe stata fornita una ricompensa per ogni parola correttamente riconosciuta. L'esperimento prevedeva tre gruppi, i quali differivano per il valore associato a ciascuna classe di parole: il gruppo 1 veniva pagato 1 centesimo per ogni parola riconosciuta sottoposta al compito strutturale, 6 centesimi per quelle riconosciute e sottoposte al compito fonologico, 3 centesimi per quelle sottoposte al compito

semantico. Diversamente il gruppo 2 veniva ricompensato con 3, 1 e 6 centesimi per ogni parola riconosciuta e sottoposta, rispettivamente, al compito strutturale, fonologico e semantico mentre nel gruppo 3 il valore monetario per ogni parola riconosciuta e sottoposta a questi tre compiti ammontava a 6, 3 e 1 centesimi rispettivamente. A ribadire la robustezza degli effetti precedentemente rilevati, anche in questo caso fu rilevato lo stesso pattern di risultati, ovvero la miglior performance di riconoscimento per le parole sottoposte al compito semantico, seguite da quelle sottoposte al compito fonologico, mentre quelle ricordate peggio furono nuovamente le parole sottoposte al compito strutturale. Se gli esperimenti riportati fino ad ora si dimostrano consistenti con il vantaggio teorizzato e attribuito alla profondità di elaborazione, sono stati anche rilevati dei dati che sembravano andare oltre quanto inizialmente proposto da Craik e Lockhart (1972). Da un lato, Craik e Tulving (1975) riscontrarono come le parole associate a risposte positive venissero generalmente ricordate meglio di quelle associate a risposte negative, per lo meno nel compito semantico e fonologico. Per spiegare questo dato, gli autori hanno fatto riferimento al concetto di integrazione o *congruenza* proposto da Schulman (1974). Egli presentò ai suoi soggetti una serie di domande alle quali rispondere sì o no e in ciascuna delle quali erano presenti una parola chiave e un descrittore che potevano formare una frase congrua o incongrua. Nel primo caso, i termini usati come parole chiave o descrittori erano appropriati e tra di loro vi era una relazione, pertanto essi conducevano ad una risposta positiva. “Una bolla è una sfera?” è un esempio di domanda congrua. Nel secondo caso, i termini erano utilizzati in maniera inappropriata come parole chiave o descrittori e pertanto conducevano ad una risposta negativa. Un esempio di domanda incongrua è “Un capitolo è scivoloso?”. Successivamente all’esecuzione di questo compito i soggetti venivano testati in maniera inaspettata rispetto alla loro capacità di ricordare o le parole chiave o i descrittori. Schulman rilevò un ampio vantaggio mnestico per i termini compresi in domande congrue, attribuendolo al fatto che questo tipo di domande andassero a favorire una codifica relazionale tra i due termini a differenza delle domande incongrue che favorirebbero una loro codifica indipendente, come unità semantiche non connesse. Rispetto ai loro risultati, Craik e

Tulving (1975) hanno quindi ipotizzato che nelle domande alle quali veniva fornita una risposta positiva, la domanda e la parola target potessero andare a formare una unità coerente ed integrata, favorendo una codifica “ricca” ed “elaborata”, diversamente dalle domande alle quali viene fornita una risposta negativa, in cui questa integrazione sarebbe meno probabile. Ciò varrebbe specialmente per il compito semantico ma si verificherebbe anche nel compito fonologico. Diversamente, nel compito strutturale, dovendo decidere se una parola è scritta in maiuscolo, presentarne una scritta in tal modo non condurrebbe ad una codifica più “elaborata” rispetto al presentarne una in lettere minuscole. È stato pertanto ipotizzato che domande capaci di indurre una codifica ugualmente elaborata nel caso di risposte positive e negative avrebbero condotto a livelli equivalenti di mantenimento per questi due tipi di risposte. Questa possibilità è stata esplorata nell’esperimento 6, in cui gli autori hanno indotto una codifica ugualmente elaborata per le parole associate a risposte negative o positive facendo svolgere ai soggetti un compito in cui ciascuna domanda implicava un confronto della parola target con un termine di riferimento rispetto a varie dimensioni, prima di eseguire un test di richiamo. Per esempio, rispetto alla domanda “è più grande di una sedia?”, sia parole associate a risposte positive, come “casa”, che parole associate a risposte negative come “spillo” sarebbero state codificate con gradi equivalenti di elaborazione. In maniera coerente con la previsione, dai risultati non emerse nessuna differenza significativa nella performance di memoria per le parole associate a risposte positive e negative, suggerendo come effettivamente anche la ricchezza o elaboratezza della codifica potesse aver contribuito alle differenze rilevate finora nella performance di memoria. Che quest’ultima dipenda dall’elaboratezza della codifica è stato rilevato anche nell’esperimento 7. In esso, ai soggetti è stato fatto svolgere solo il compito semantico di completamento della frase e il grado di elaboratezza della codifica è stato variato presentando frasi di tre livelli di complessità, da molto semplici a frasi di media complessità e molto elaborate che si supponeva avrebbero condotto ad una codifica a sua volta più elaborata e di conseguenza ad una miglior performance di memoria in un successivo compito di richiamo o richiamo guidato. A supporto della precedente interpretazione, solo nel caso delle

risposte positive, fu rilevato un miglior mantenimento delle parole presentate in frasi più complesse, in particolar modo nel compito di richiamo guidato quando venivano riproposte ai soggetti nuovamente le stesse cornici semantiche, ovvero le stesse frasi assieme a cui avevano codificato inizialmente le parole. È da notare come questi risultati furono ottenuti in un esperimento in cui i soggetti svolgevano un compito che assicurava un'elaborazione semantica per tutte le parole target. Ciò ha portato Craik e Tulving ad ipotizzare che anche altri fattori potessero partecipare alla performance di memoria oltre alla profondità inizialmente proposta. Tra questi, la complessità della domanda, la relazione tra la parola target e il contesto della domanda e la presenza o l'assenza di un indizio per il recupero della parola. In ogni caso, risulta chiaro dai vari esperimenti svolti dagli autori e sopraccitati come siano la natura qualitativa del compito e il tipo di operazioni svolte sul materiale da ricordare a determinarne il mantenimento (Craik & Tulving, 1975). Nel prossimo paragrafo saranno riportati alcuni studi che hanno utilizzato un particolare strumento per garantire un'elaborazione profonda del materiale di studio, il giudizio di piacevolezza.

1.2. Livelli di elaborazione e materiale di studio verbale: giudizio di piacevolezza.

Un compito spesso utilizzato in paradigmi di memoria incidentale è il giudizio di piacevolezza. Si tratta di un compito semantico (Hyde e Jenkins, 1969, 1973; Walsh & Jenkins, 1973) che può avere la forma o di una classificazione dello stimolo come *piacevole* o *non piacevole* (Baddeley & Hitch, 2017) o di una valutazione di questa dimensione lungo un range di valori (Packman & Battig, 1978). La superiorità mnestica per il materiale sottoposto al giudizio di piacevolezza rispetto a quello sottoposto a diversi compiti di natura non semantica è stata rilevata in diversi studi. Hyde e Jenkins (1969) hanno utilizzato questo compito per indagare il richiamo e l'organizzazione di una

lista di parole tra di loro associate tra diversi gruppi di soggetti. A un primo gruppo venivano semplicemente presentate le parole senza che essi svolgessero alcun tipo di compito sul materiale di studio e i soggetti venivano informati che avrebbero successivamente dovuto svolgere un test di richiamo su di esse. In una seconda condizione, i soggetti eseguivano uno di tre diversi compiti incidentali sul materiale di studio, senza essere informati del successivo test di richiamo. I tre compiti erano costituiti dal giudizio di piacevolezza, un compito in cui i soggetti dovevano rilevare la lettera E all'interno delle parole presentate e un compito in cui essi dovevano stimare il numero di lettere nelle parole presentate. La terza condizione era identica alla seconda, con la differenza che i soggetti venivano anche informati del successivo test di richiamo. In tutte le condizioni, rispetto alla media di parole richiamate, solo la performance dei soggetti che eseguirono il compito di piacevolezza risultò paragonabile a quella dei soggetti che eseguirono il solo test di richiamo. Tutti gli altri compiti svolti, condussero ad una performance di memoria peggiore, senza differenze nel mantenimento per le parole sottoposte al compito di rilevamento della "E" o al compito di stima delle lettere. Ciò fu interpretato in funzione del fatto che nel compito di piacevolezza le parole venissero utilizzate come unità semantiche, rispetto agli altri due compiti in cui esse venivano utilizzate come oggetti o insieme di lettere. Il vantaggio del compito di piacevolezza su altri compiti non semantici è stato rilevato anche in altri studi (Hyde & Jenkins, 1973; Walsh & Jenkins, 1973).

Una diversa interpretazione di questo vantaggio è quella fornita da Postman e Kruesi (1977). Nel confrontare il richiamo di parole sottoposte al giudizio di piacevolezza o ad un compito, a sua volta semantico, in cui i soggetti dovevano giudicare la frequenza d'uso delle parole presentate, è stato rilevato un vantaggio per le parole sottoposte al primo tipo di compito. Questo risultato è stato interpretato in termini di *ripetizione distribuita*, ovvero in funzione del fatto che giudicando gli stimoli su una singola scala soggettiva come nel caso della piacevolezza, è molto probabile che i soggetti confrontino i nuovi item con quelli precedenti per stabilire un giudizio relativo. Questi confronti comporterebbero una ripetizione distribuita che contribuirebbe alla formazione di associazioni tra gli item, ripetizione che diminuirebbe spostandosi dalla valutazione

di attributi puramente soggettivi a quella di attributi oggettivi ma non direttamente osservabili fino a quella di caratteristiche direttamente osservabili degli stimoli presentati. Sarà pertanto minore in un compito di frequenza in cui gli attributi da valutare sono oggettivi ma non direttamente osservabili, in cui i giudizi saranno più probabilmente ancorati a riferimenti esterni come parole prototipiche e sarà ulteriormente ridotta in altri compiti, come nel rilevamento del numero di volte in cui una particolare lettera compare all'interno di una parola, in cui l'attributo da discriminare è direttamente osservabile.

Il vantaggio del compito di piacevolezza rispetto ad altri tipi di processamento semantico è stato rilevato anche da Packman e Battig (1978) in un lavoro in cui i soggetti dovevano eseguire una valutazione delle parole presentate rispetto a sette dimensioni semantiche: concretezza, immaginazione, quanto fossero categorizzabili, significatività, familiarità, numero di attributi e piacevolezza. In ciascuna dimensione, la valutazione veniva fornita lungo un range di valori che andava da 1 -molto basso- a 7 -molto alto-. Successivamente a questi giudizi i soggetti eseguivano un compito di memoria per le parole presentate. I risultati mostrarono come la performance di memoria per le parole sottoposte al giudizio di piacevolezza fosse significativamente superiore rispetto a quella per le parole sottoposte a tutti gli altri tipi di giudizi semantici, tra i quali non vi erano differenze significative, con lo stesso pattern riscontrato in tutte le variabili di interesse misurate e sia in un test di richiamo libero che nel riconoscimento. Sulla base del dato per cui le valutazioni di piacevolezza eseguite dai soggetti fossero anche quelle meno correlate a tutti gli altri giudizi relativi alle altre 6 dimensioni, in maniera analoga a quanto precedentemente rilevato da Toggia & Battig (1978, citato in Packman & Battig, 1978), questo vantaggio è stato interpretato in funzione della maggiore distintività e unicità dei giudizi di piacevolezza rispetto alle altre sei dimensioni semantiche.

1.3. Livelli di elaborazione e materiale di studio visivo

Gli effetti dei livelli di elaborazione sul mantenimento del materiale di studio sono stati indagati anche rispetto a vari tipi di stimoli visivi e con risultati talvolta contraddittori. Nei prossimi paragrafi saranno presentati alcuni di questi studi a seconda degli stimoli visivi utilizzati, a partire dai volti fino alle scene naturali. Il riscontro di evidenze a favore del vantaggio di un'elaborazione profonda anche per le scene naturali è di particolare rilevanza dato che si tratta degli stimoli utilizzati nello studio presentato in questa tesi e che verrà discusso successivamente.

1.3.1 Livelli di elaborazione e stimoli visivi: volti

Come rilevato da Baddeley e Hitch (2017), l'indagine degli effetti legati ai livelli di elaborazione nel caso della memoria visiva si è focalizzata in larga parte sulla memoria per i volti. Rispetto a questo tipo di stimoli, Bower & Karlin (1974), hanno realizzato uno studio in cui è stata confrontata la performance mnestica per i volti sottoposti a un giudizio di simpatia o onestà rispetto ad un compito più superficiale in cui i soggetti dovevano giudicarne il sesso. Nello studio in questione, ai soggetti venivano presentate immagini di volti e rispetto a ciascuna di esse i partecipanti dovevano svolgere uno di tre compiti: giudicare il sesso del volto presentato, utilizzare qualsiasi criterio per giudicare la gentilezza della persona presentata in foto, oppure giudicarne l'onestà. La previsione era che il giudizio del sesso del volto avrebbe condotto ad un livello di elaborazione minore rispetto agli altri due compiti e di conseguenza ad una peggior performance in un test di riconoscimento che i soggetti svolgevano successivamente e del quale non erano stati informati. Ciò è esattamente quello che è stato rilevato: i volti giudicati rispetto al sesso furono ricordati peggio rispetto a quelli sottoposti al giudizio di gentilezza e a quello di onestà. Rispetto a questi due compiti, fu rilevato inoltre un vantaggio marginale a favore degli stimoli sottoposti al giudizio di onestà. Risultati simili

furono ottenuti anche in un secondo esperimento, identico al primo ma con la differenza che i soggetti venivano informati del successivo test di riconoscimento. Inoltre, un'altra differenza rilevata rispetto all'esperimento precedente fu un vantaggio per gli stimoli sottoposti al giudizio di gentilezza rispetto a quello di onestà in queste condizioni di apprendimento intenzionale. Le differenze rilevate furono imputate dagli autori al fatto che per eseguire il compito del giudizio del sesso del volto, questo poteva essere emesso semplicemente basandosi sulla presenza di una sua caratteristica saliente, ad esempio di sopracciglia folte. In maniera differente, un giudizio come quello di onestà potrebbe implicare diverse strategie, come il confronto tra una persona onesta conosciuta e lo stimolo test. Tali strategie richiederebbero di esaminare un maggior numero di caratteristiche dei volti presentati, di esaminarle per più tempo e un maggior utilizzo dello stimolo test dal momento che questo dovrebbe essere comparato con dei prototipi di onestà recuperati in memoria.

A differenza di Bower e Karlin, Patterson e Baddeley (1977) hanno utilizzato in un loro studio un diverso compito come mezzo per garantire un'elaborazione profonda dei volti, ovvero un giudizio delle caratteristiche di personalità, confrontandolo con un compito più superficiale che consisteva nella valutazione delle caratteristiche fisiche dei volti, rilevando un vantaggio in un successivo test di riconoscimento per gli stimoli sottoposti al primo tipo di giudizio.

È di rilievo enfatizzare come il giudizio di piacevolezza abbia rappresentato un utile strumento mnestico anche per questa categoria di stimoli. Ciò è stato rilevato da Warrington e Ackroyd (1975) in un esperimento in cui ai soggetti venivano presentate sia delle parole sia degli stimoli visivi costituiti da volti. Per ciascuna di questa classe di stimoli erano previste tre condizioni: una in cui ai soggetti venivano semplicemente presentati gli stimoli informandoli del fatto che avrebbero dovuto svolgere un test di riconoscimento rispetto ad essi, una seconda condizione in cui i soggetti categorizzavano ciascuno stimolo come piacevole o non piacevole prima di svolgere il test di riconoscimento del quale erano nuovamente stati informati, una terza condizione in cui i soggetti dovevano categorizzare ciascuna parola come rossa o verde o ciascun

volto come alto o basso prima del test di riconoscimento, del quale erano consapevoli. Sia per i volti che per le parole, fu rilevato un chiaro miglioramento della performance di riconoscimento per gli stimoli sottoposti al giudizio di piacevolezza rispetto a quella rilevata per gli stimoli sottoposti agli altri tipi o a nessun compito. Non furono invece riscontrate differenze tra gli stimoli sottoposti a nessun compito e quelli giudicati in base al colore o all'altezza.

1.3.2 Livelli di elaborazione e stimoli visivi: scene naturali

La generalizzazione degli effetti legati alla profondità di elaborazione anche alle scene naturali è di particolare rilevanza dal momento che si tratta della categoria di stimoli utilizzata nello studio presentato in questa tesi. Sebbene gli studi riportati nel paragrafo precedente mostrino come l'elaborazione approfondita possa favorire il mantenimento anche per stimoli visivi come i volti, come fatto notare da Intraub e Nicklos (1985) non è scontato che i risultati ottenuti su questo tipo di stimoli omogenei generalizzino anche ad altri stimoli caratterizzati da una ricchezza di dettagli visivi. Nello studio dei due autori gli effetti legati a differenti livelli di elaborazione dello stimolo sono stati indagati utilizzando come stimoli visivi delle scene naturali a colori, ciascuna delle quali era preceduta o da una domanda di tipo semantico che richiedeva di considerare il significato dell'immagine oppure da una domanda che richiedeva di basarsi sulle sue caratteristiche fisiche, come l'orientamento o il colore. Rispetto alla performance in un successivo compito di richiamo libero del quale non erano stati informati, in cui i soggetti dovevano scrivere una breve descrizione per ciascuna immagine richiamata in modo tale da consentire allo sperimentatore di riconoscere a quale si stessero referendo, fu effettivamente riscontrato anche per le scene naturali un effetto differente legato ai due tipi di domande, tuttavia opposto rispetto a quello tipicamente rilevato. Fu infatti osservato un vantaggio nella performance di memoria nel compito di richiamo per le immagini sottoposte al compito strutturale rispetto a quelle sottoposte al compito semantico. Lo stesso pattern di risultati è stato rilevato utilizzando domande semantiche specifiche per determinate

categorie, con differenti tempi di esposizione dell'immagine e utilizzando un test di richiamo guidato, e con fotografie di singoli oggetti. Dallo studio sopra riportato sembra che un'elaborazione profonda basata sul significato e le caratteristiche semantiche non conduca ad un migliore mantenimento di questa categoria di stimoli.

Tuttavia, un'analisi estensiva degli effetti legati ai diversi livelli di elaborazione utilizzando diversi set di oggetti e scene naturali come stimoli visivi è stata realizzata da Baddeley e Hitch (2017), i quali hanno rilevato un effetto modesto della profondità di elaborazione rispetto a quello ottenuto per diversi tipi di stimoli verbali, tuttavia consistente. Inoltre, nello studio in questione è stato utilizzato il giudizio di piacevolezza come strumento per favorire un'elaborazione profonda del materiale di studio. Data quindi la rilevanza per il presente lavoro, verranno ora discussi i vari esperimenti svolti da questi autori. Il primo esperimento prevedeva la presentazione di due categorie di stimoli ai partecipanti, stimoli visivi costituiti da immagini di porte e stimoli verbali costituiti da nomi di animali e oggetti domestici. Dopo la presentazione di ciascuno stimolo, rispetto ad essi i soggetti dovevano eseguire o un compito superficiale in cui dovevano giudicare se fosse marrone o non marrone oppure il giudizio di piacevolezza, indicando se lo stimolo fosse piacevole o non piacevole. Successivamente i soggetti venivano sottoposti ad un test di riconoscimento per gli stimoli presentati. Fu rilevato per entrambe le categorie di stimoli un effetto legato alla profondità di elaborazione. Infatti, sia le immagini che le parole sottoposte al giudizio di piacevolezza furono ricordate meglio degli stimoli sottoposti al compito superficiale. Tuttavia, l'effetto rilevato per le immagini era significativamente inferiore a quello rilevato per gli stimoli verbali. Il secondo esperimento prevedeva una procedura simile, con alcune differenze. Furono utilizzate porte appartenenti ad un diverso set come stimoli visivi, e nomi propri e di mestieri come stimoli verbali, andando a variare anche il compito superficiale, che in questo caso prevedeva di nominare il colore dello stimolo. Fu nuovamente rilevato un vantaggio mnestico per tutti gli stimoli sottoposti al giudizio di piacevolezza rispetto al compito superficiale, con l'effetto maggiore rilevato per i nomi di mestieri. Per le immagini di porte fu rilevato un vantaggio mnestico significativo a seguito

dell'elaborazione profonda, ma inferiore a quello rilevato per il materiale verbale. Tuttavia, l'effetto della profondità di elaborazione rilevato per gli stimoli visivi risultò comparabile a quello ottenuto con l'altra classe di stimoli verbali, ovvero i nomi propri. Gli esperimenti successivi furono eseguiti andando a variare nuovamente gli stimoli utilizzati e il compito superficiale. Uno di questi prevedeva la presentazione dello stesso set di porte dell'esperimento 2 ma anche di un'altra classe di stimoli visivi, costituiti da immagini di orologi, mentre gli stimoli verbali utilizzati erano costituiti da nomi di cibi. Rispetto al compito superficiale, questo richiedeva di giudicare la dimensione dello stimolo. A conferma dei risultati precedenti, l'elaborazione profonda legata al giudizio di piacevolezza rispetto all'esecuzione del compito superficiale portò ad un vantaggio nella performance di riconoscimento per tutti i tipi di stimoli, ma nuovamente fu rilevato un effetto modesto per entrambe le categorie di stimoli visivi, senza differenze tra di essi, rispetto a quanto rilevato per gli stimoli verbali. Risultati identici furono ottenuti in un successivo esperimento sostituendo alle immagini di orologi immagini di telefoni e ai nomi di cibi nomi di stazioni londinesi e al precedente compito superficiale l'istruzione di riportare il colore dominante dello stimolo. Tuttavia, nell'ultimo esperimento svolto fu riscontrata una differenza. Questo prevedeva la presentazione dello stesso set di porte come stimoli visivi e di scene naturali appartenenti a diverse categorie di ambienti interni ed esterni. Gli stimoli verbali presentati erano invece costituiti da nomi di animali appartenenti a diverse categorie che a loro volta presentavano diverse sottocategorie. Rispetto agli "insetti", i nomi presentati potevano ad esempio includere diverse specie di formiche, come le formiche di terra, formiche tagliafoglie ecc. Il compito superficiale utilizzato inoltre era costituito in questo caso dal giudicare se lo stimolo avesse un colore dominante o fosse multicolore. Anche in questo caso fu rilevato un miglioramento della performance di memoria per tutti gli stimoli sottoposti al giudizio di piacevolezza rispetto al compito superficiale, con questo effetto nuovamente modesto per gli stimoli visivi ma in questo caso non differente da quello rilevato per gli stimoli verbali. Dai vari esperimenti svolti, risulta quindi come l'elaborazione profonda abbia condotto ad un vantaggio modesto ma consistente nel mantenimento di tutti gli stimoli visivi

utilizzati mentre per i vari stimoli verbali impiegati i risultati furono più eterogenei. Per spiegare ciò, gli autori hanno fatto ricorso al concetto di affordance (Gibson, 1977), un termine che si riferisce alle possibilità d'azione offerte ad un animale dall'ambiente, in riferimento alle capacità d'azione dell'animale stesso (Osiurak et al., 2017). In maniera analoga, quello che viene proposto è che stimoli visivi come le porte domestiche utilizzate in questo esperimento, offrirebbero relativamente poco alla codifica in termini di potenziale elaborazione semantica. A meno che non si tratti di una porta familiare, nel caso di questi stimoli visivi il giudizio di piacevolezza si baserà su caratteristiche visive superficiali come la forma, la dimensione, tutte caratteristiche che verranno codificate rapidamente anche svolgendo compiti più superficiali come quelli in cui bisogna giudicare il colore dello stimolo. Ciò spiegherebbe il modesto effetto legato all'elaborazione profonda riscontrato per gli stimoli visivi utilizzati. Diversamente, gli stimoli verbali fornirebbero accesso durante la codifica ad un ampio set di associazioni semantiche. Nel caso del nome di un animale familiare, come quelli utilizzati nel primo esperimento svolto, la parola "leopardo" ad esempio potrebbe offrire accesso ad una rappresentazione visiva dell'animale, dell'habitat in cui vive e così via. Questo range di affordances sarebbe disponibile anche nel caso di stimoli come i nomi di cibi utilizzati nel secondo esperimento ma verrebbe meno nel caso dei nomi delle sottocategorie di animali dato che si tratta di stimoli non familiari e nel caso dei nomi propri, dal momento che questi ultimi sono arbitrari. Indipendentemente dall'interpretazione proposta, gli esperimenti svolti da Baddeley ed Hitch (2017) mostrano quindi come il giudizio di piacevolezza risulti efficace nel promuovere il mantenimento in memoria anche di stimoli visivi quali immagini di oggetti e scene naturali.

2. TESTING EFFECT

Riprendendo la definizione proposta inizialmente, di memoria come “capacità neurocognitiva di codificare, immagazzinare e recuperare l’informazione” (Tulving, 2000), se la teoria dei livelli di elaborazione pone l’accento sulla profondità della codifica per il successivo mantenimento della traccia di memoria (Craik & Lockhart, 1972; Craik & Tulving, 1975), il fenomeno del *testing effect* riflette l’importanza del processo di recupero dell’informazione nella promozione del suo mantenimento. Roediger e Karpicke (2006B) nel definire questo fenomeno sostengono come lo svolgere un test sul materiale di studio abbia un maggiore effetto positivo sul suo mantenimento rispetto allo spendere un ammontare equivalente di tempo nel ristudio del materiale, anche quando la performance nel test è lontana dall’essere perfetta e non viene fornito un feedback sull’informazione mancante. Questo fenomeno si basa sui processi di recupero che vengono intrapresi durante lo svolgimento di un test, ed è pertanto stato definito anche *retrieval-based learning* (Karpicke, 2017). L’importanza del recupero è ben nota: nel momento in cui un ricordo viene recuperato dalla memoria esso ritorna nello stato di fragilità che lo caratterizzava nel momento in cui si è formato, stato di fragilità in cui può essere riconsolidato, modificato o eliminato (Goldstein, 2016); in questi termini, è appropriata la definizione di recupero come *modificatore di memoria* (Bjork, 1975; Bjork & Bjork, 1992). Il recupero dell’informazione dalla memoria andrebbe infatti a modificare il sistema, costituendo esso stesso un potente evento di apprendimento (Bjork & Bjork, 1992). Nei paragrafi successivi saranno presentati alcuni studi che hanno indagato questo fenomeno e in particolare quelli che se ne sono occupati utilizzando materiale visivo, di particolare rilevanza per il lavoro qui presentato. Nell’ultimo paragrafo saranno riportate alcune delle teorie proposte in letteratura per spiegare gli effetti legati al testing.

2.1. Testing Effect: prospettiva storica e paradigmi

Come fatto notare da McDermott (2021), l'idea che praticare il recupero dell'informazione studiata possa facilitarne il successivo recupero è antecedente allo studio sperimentale della memoria. Uno dei primi studi a mettere in evidenza l'importanza di questo processo per il consolidamento mnestico è stato condotto da Gates (1917). Egli fornì a diversi gruppi di bambini di diversa età 9 minuti per l'apprendimento di due tipi di materiale verbale, costituiti da sillabe senza senso e biografie, andando a manipolare la percentuale di tempo (da 0 all' 80%) all'interno di questi 9 minuti spesa nel processo di *recitazione*. Quest'ultima può essere definita come l'alternanza tra la lettura del materiale, lo svolgere un test rispetto ad esso e l'ottenimento di un feedback sottoforma di rilettura del materiale nel momento in cui il suo recupero dalla memoria fallisce (McDermott, 2021). Nell'esaminare i risultati di un test di richiamo libero dopo i 9 minuti di studio, rispetto alle sillabe senza senso, emerse una chiara relazione tra il tempo speso nel processo di recitazione in fase di codifica e la performance finale di memoria. Maggiore la percentuale di tempo spesa in questo processo, migliore la performance di richiamo, in tutte le età testate. Diversamente, nel caso dell'altro materiale verbale, le biografie, questo pattern fu rilevato solo nei bambini più grandi (Gates, 1917).

Se Gates prende in esame un processo, quello di recitazione, che implica per definizione un feedback sottoforma di riesposizione al materiale di studio, Spitzer (1939) ha eseguito un esperimento in cui tale feedback non veniva fornito ai partecipanti. Egli presentò a dei bambini di classe sesta degli articoli di circa 600 parole da studiare in un periodo di 8 minuti. La manipolazione in questo studio riguardava il numero di test a scelta multipla svolti, privi di feedback, rispetto al materiale di studio e il periodo in cui questi venivano somministrati rispetto al test finale. Dai risultati fu chiaro come lo svolgere un test fosse in grado di attenuare la curva di dimenticanza rispetto al materiale studiato, soprattutto quando il test veniva svolto poco tempo dopo la fase di studio.

L'interesse recente verso il testing si deve a Roediger e Karpicke, i quali hanno pubblicato due importanti lavori sul tema. Il primo consiste in uno studio empirico (Roediger & Karpicke, 2006a), seguito da una review sull'argomento (Roediger & Karpicke, 2006b). Nel loro studio (2006a), essi indagarono il fenomeno del testing in condizioni didatticamente rilevanti, utilizzando testi in prosa come materiale di studio e un test di richiamo libero senza feedback. Inoltre, essi aggiunsero un'importante condizione di controllo costituita dalla riesposizione alla totalità del materiale di studio. Nell'esperimento 1 furono presentati agli studenti dei brevi passaggi di prosa relativi ad argomenti scientifici. Rispetto ad essi, i soggetti eseguivano o un test di richiamo libero senza feedback oppure ristudiavano il materiale prima di essere sottoposti ad un test finale di richiamo libero dopo 5 minuti, 2 giorni oppure una settimana. Nella condizione in cui il test finale venne somministrato dopo 5 minuti dalla fase di studio, i soggetti che avevano ristudiato il materiale ricordarono di più rispetto a coloro che erano stati sottoposti al test. Tuttavia, rispetto a coloro che svolsero il test finale dopo 2 giorni o una settimana, fu rilevato il pattern opposto. Il secondo esperimento svolto era finalizzato all'indagine degli effetti di test ripetuti. Erano pertanto previste tre condizioni: nella prima fase dell'esperimento, in una prima condizione i soggetti studiavano 4 volte i passaggi di prosa senza venire sottoposti ad alcun test, in una seconda studiavano 3 volte il materiale e venivano sottoposti ad un singolo test, mentre in una terza condizione ai soggetti veniva fatto studiare solo una volta il materiale prima di sottoporli a 3 test. Successivamente, in una seconda fase, era previsto un test finale che poteva essere somministrato dopo 5 minuti oppure una settimana. In questo esperimento era inoltre prevista la compilazione di un questionario a seguito della prima fase dell'esperimento, in cui i soggetti dovevano indicare tra le altre cose quanto bene pensavano che avrebbero ricordato una settimana dopo ciò che avevano studiato. In merito ai risultati, quando il test finale veniva fatto svolgere dopo 5 minuti, i soggetti che avevano studiato il materiale 4 volte ricordarono di più rispetto a coloro che lo avevano studiato 3 volte e svolto un test, che a loro volta mostrarono una performance di memoria migliore rispetto a coloro che avevano studiato i passaggi una singola volta ma svolto tre test. Tuttavia, dopo una

settimana fu rilevato un pattern esattamente opposto. Entrambi gli esperimenti pertanto mostrarono come lo svolgere un test sul materiale di studio fosse in grado di favorire il mantenimento a lungo termine. Al contrario, quando il test finale veniva somministrato poco tempo dopo era il ristudio a fornire un vantaggio. Allo stesso tempo l'aggiunta della condizione di ristudio ha consentito di escludere come il testing effect non dipenda dalla riesposizione al materiale durante il test, dal momento che i soggetti sottoposti al ristudio venivano riesposti al 100% del materiale ma, nonostante ciò, mostrarono una performance peggiore nel mantenimento a lungo termine. Oltre a ciò, gli autori rilevarono anche un altro importante dato: sebbene i soggetti che avevano studiato il materiale quattro volte mostrarono la peggior performance nel test finale dopo una settimana, nel questionario svolto nel secondo esperimento essi riportarono una migliore fiducia nei confronti del loro ricordo dopo una settimana rispetto agli altri due gruppi, che stimarono al contrario, senza differenze tra di loro, che avrebbero ricordato peggio quanto studiato. Pertanto, come sostenuto da Karpicke e colleghi (2009), sebbene lo svolgere un test su quanto studiato rappresenti un efficace strumento per il mantenimento a lungo termine, gli studenti non mostrano questa consapevolezza metacognitiva dell'effetto del test nel monitoraggio del proprio apprendimento ma al contrario manifestano una *illusione di competenza* a seguito di una lettura ripetuta del materiale. In un sondaggio svolto dagli autori sulle principali tecniche di studio adottate da un campione di studenti universitari è stato rilevato come la maggior parte di essi preferisca rileggere i propri appunti o libri di testo rispetto all'esecuzione di un test su quanto studiato. Tale rilettura ripetuta, sebbene non si riveli una strategia efficace per il mantenimento rispetto all'esecuzione di un test, innescherebbe una illusione di competenza incrementando la fluenza e la facilità con cui gli studenti elaborano il testo. Essi, pertanto, baserebbero le valutazioni del loro apprendimento su questi fattori, sebbene essi non siano predittivi del mantenimento futuro. Il fenomeno del testing effect si è dimostrato un fenomeno robusto (vedi Rowland, 2014 per una metanalisi), ed è stato replicato sia in laboratorio che in contesti didattici (vedi Roediger & Karpicke, 2006B per una review). In una review sull'argomento, McDermott riporta studi che mostrano come questo effetto sia stato replicato con

soggetti di varie età, dalla scuola elementare fino agli adulti, facendo ricorso a vari tipi di materiale verbale ma anche utilizzando materiale non verbale. Poiché importanti rispetto allo studio proposto in questa tesi, nel prossimo paragrafo verranno presentati alcuni studi che hanno portato evidenze del fenomeno anche rispetto a materiale di studio visivo.

2.2. Testing Effect e materiale di studio visivo

Molta della ricerca sul testing effect ha indagato questo fenomeno in riferimento a materiali verbali di diverso tipo (Dunlosky et al., 2013). In questo filone rientrano anche gli studi sopra descritti. Tuttavia, vari autori si sono occupati dell'indagine degli effetti di un test anche sulla memoria per stimoli visivi. Tra questi, Wheeler e Roediger (1992) hanno rilevato come lo svolgere almeno un test su una lista di immagini studiate andasse a favorirne il mantenimento a lungo termine dopo una settimana rispetto al non svolgerne alcuno. In vari studi sul fenomeno del testing sono stati utilizzati come stimoli volti e nomi presentati in maniera accoppiata (Carpenter & DeLosh, 2005; Landauer & Bjork, 1978; Morris & Fritz, 2002; Tse et al., 2010). Questi studi hanno messo in evidenza come lo svolgere un test anche su questo tipo di stimoli possa promuoverne il mantenimento rilevato in un test finale somministrato dopo 5 minuti (Carpenter e DeLosh, 2005), 30 minuti dopo (Landauer & Bjork, 1978) ma anche a distanza di 2 settimane (Morris & Fritz, 2002). Sebbene questi studi sottolineino la consistenza del fenomeno del testing rispetto a questo tipo di stimoli, lo studio svolto da Tse e colleghi (2010) ha rilevato come soggetti più anziani esposti a questi accoppiamenti tra volti e nomi, nel doverli ricordare in un test finale svolto dopo un'ora e mezzo beneficiassero meglio di uno studio ripetuto rispetto allo svolgimento di una serie di test privi di feedback, a differenza dei soggetti di mezza età in cui fu riscontrato il pattern opposto. Tuttavia, lo stesso beneficio del test ripetuto sul materiale di studio fu riscontrato anche sui soggetti più anziani in un secondo esperimento, in cui i test che venivano fatti svolgere prevedevano un feedback.

Altri studi hanno rilevato il vantaggio mnestico a cui il test può condurre anche rispetto all'apprendimento di materiale visuo-spaziale (Carpenter & Pashler, 2007; Rohrer et al., 2010). Carpenter e Pashler hanno presentato ai loro soggetti due mappe, su ciascuna delle quali erano riportati vari elementi come strade, fiumi ed edifici, e per le quali non era richiesta una risposta verbale esplicita. Una di queste mappe veniva studiata una seconda volta, mentre l'altra veniva sottoposta ad un test in cui essa veniva ripresentata assieme ad uno degli elementi che conteneva rimosso, con l'istruzione data ai partecipanti di individuare l'elemento mancante e di formarne un'immagine mentale. Tutti gli elementi della mappa venivano testati in questa maniera. La seconda fase dell'esperimento prevedeva un test finale, in cui veniva fornito un foglio bianco e indicato ai soggetti di disegnare le due mappe al meglio che potevano. A ricalcare la consistenza del testing effect, anche in questo caso fu rilevata una performance di memoria migliore per le mappe testate rispetto a quelle ristudiate una seconda volta. Risultati simili furono anche ottenuti da Rohrer e collaboratori (2010), sempre utilizzando mappe come stimoli visivi. Anche questi autori hanno riscontrato un vantaggio per il materiale testato rispetto a quello sottoposto a ristudio sia in un test finale identico a quello intermedio che in uno diverso.

Rispetto allo studio svolto da Carpenter e Pashler (2007), Kang (2010) ha fatto notare come la posizione degli elementi su una mappa possa essere registrata verbalmente, con questa elaborazione verbale che potrebbe andare a mediare gli effetti del test. Per questo motivo egli ha realizzato uno studio in cui gli stimoli utilizzati erano costituiti da materiale visuo-spaziale difficile da verbalizzare, ovvero caratteri cinesi accoppiati ai loro equivalenti in lingua inglese. In un primo esperimento, al gruppo nella condizione di ristudio questi accoppiamenti tra il carattere cinese e quello inglese venivano fatti studiare tre volte, mentre il gruppo sottoposto alla condizione di test li studiava una singola volta prima di essere sottoposto a due test di richiamo guidato con feedback, in cui si presentava la parola inglese richiedendo ai soggetti di visualizzare mentalmente il carattere cinese appropriato. A seguito della fase di apprendimento, entrambi i gruppi svolgevano un test di richiamo guidato dopo 10 minuti. Anche in questo esperimento, i soggetti sottoposti al test

ottennero una performance di memoria migliore rispetto al gruppo assegnato alla condizione di ristudio. Lo stesso risultato fu ottenuto anche in un secondo esperimento in cui il test finale di richiamo guidato veniva fatto svolgere dopo un giorno dalla fase di apprendimento e in un terzo esperimento in cui ciascuno gruppo ristudiava una metà degli stimoli e svolgeva un test rispetto all'altra metà. In maniera simile allo studio precedente, il test si è rivelato un efficace strumento di memoria anche per altri tipi di simboli, come quelli prodotti dalla tribù degli Asante del Ghana (Coppens et al., 2011).

2.3. Cornici teoriche proposte per la spiegazione del fenomeno

Come sostenuto da Karpicke (2017), la questione fondamentale che le teorie del retrieval-based learning dovrebbero affrontare è rivolta a come mai un recupero iniziale favorisca la probabilità di un recupero successivo dell'informazione rispetto a condizioni di controllo in cui questo processo non viene intrapreso. Verranno di seguito presentate alcune delle teorie proposte per spiegare gli effetti sulla memoria legati al test: l'elaborazione appropriata al trasferimento (Morris et al., 1977), la difficoltà desiderabile (Bjork, 1975; 1994) e il modello della biforcazione (Kornell et al., 2011).

2.3.1. Elaborazione appropriata al trasferimento

La teoria dell'elaborazione appropriata al trasferimento è stata proposta da Morris e colleghi (1977). In una serie di esperimenti questi autori misero in dubbio l'appropriatezza dei livelli di elaborazione (Craik & Lockhart, 1972; Craik & Tulving, 1975) mostrando come anche una codifica più superficiale come quella basata sulle caratteristiche fonologiche delle parole potesse condurre ad una performance di memoria migliore rispetto ad una codifica più profonda e basata sul significato quando il compito finale di riconoscimento fatto svolgere ai soggetti richiedeva di considerare il

suono delle parole presentate. All'interno di questa cornice teorica, la corrispondenza tra il tipo di elaborazione durante la codifica e il tipo di elaborazione durante il recupero condurrà ad una migliore performance di memoria (Goldstein, 2016). In riferimento al testing effect, lo svolgere un test sul materiale di studio consentirebbe agli studenti di mettere in atto il processo di recupero durante l'apprendimento e quindi di esercitare le stesse abilità necessarie per migliorare il successivo recupero dell'informazione (Roediger & Karpicke, 2006B). Questo processo può essere paragonato al percorso di uno studente che intende imparare a suonare il violino: in questo processo, l'esercitarsi a suonare lo strumento è fondamentale; diversamente, ascoltare qualcun altro che lo suona o leggere di come suonarlo non può andare a sostituire questo esercizio. In maniera analoga, dal momento che in una valutazione finale gli studenti dovranno praticare il recupero e utilizzare la propria conoscenza ha senso che essi pratichino il recupero e che la utilizzino durante le fasi iniziali dell'apprendimento (Karpicke, 2017). Se questa teoria offre da un lato una spiegazione dell'importanza del recupero per l'apprendimento, essa non propone un meccanismo che spieghi come e perché praticare il recupero favorisca l'apprendimento (Karpicke, 2017) e allo stesso tempo conduce ad una specifica previsione: la performance in un test finale dovrebbe essere migliore quando questo test è nello stesso formato di quello intermedio (Roediger & Karpicke, 2006B). Tuttavia, la semplice corrispondenza tra il formato dei due test non è sempre la condizione più desiderabile (McDermott, 2021): diversi studi mostrano come un test intermedio di richiamo possa favorire il mantenimento in misura maggiore rispetto ad un test intermedio di riconoscimento, a risposta multipla, indipendentemente dal formato del test finale (Karpicke, 2017; Roediger & Karpicke, 2006B). Sulla stessa linea, uno studio già sopraccitato in cui ciò è riscontrabile è quello di Rohrer e collaboratori (2010). Nel loro studio in cui sono state utilizzate delle mappe come materiale, sottoposte o a ristudio o ad un test intermedio, il test finale poteva avere o lo stesso formato di quello intermedio, consistente nel riportare nella posizione corretta sulla mappa il nome di ciascuna regione presentata, o un diverso formato in cui i nomi delle regioni non venivano ripresentati. Sebbene lo svolgimento del test intermedio favorisse la performance in entrambi i test

finali rispetto al ristudio, l'effetto rilevato fu addirittura maggiore per il test finale di diverso formato. Lo stesso dato fu riscontrato anche in un secondo esperimento in cui il test finale poteva avere nuovamente lo stesso formato di quello intermedio, che consisteva nel riportare sulla mappa vuota i nomi delle città, oppure richiedere di identificare la città che si sarebbe dovuta attraversare nel guidare lungo il percorso tra altre due località.

2.3.2. Difficoltà desiderabile e sforzo di recupero

Una spiegazione alternativa dell'importanza del recupero per il mantenimento dell'informazione si basa sull'idea generale per cui praticare il recupero costituisca uno sforzo per chi apprende, con questo sforzo che condurrebbe ad un potenziamento dell'informazione recuperata incrementando la probabilità di accedere ad essa e di utilizzarla nuovamente in futuro (Karpicke, 2017). Secondo Bjork (1994), l'introduzione di difficoltà durante l'apprendimento può fornire un vantaggio per chi apprende, andando a favorire la performance. In un suo lavoro, (Bjork,1975) egli prende in esame il fenomeno del recupero all'interno della cornice teorica dei livelli di elaborazione proposta da Craik e Lockhart (1972). Così come una codifica profonda ha come risultante una traccia di memoria rafforzata e duratura, allo stesso modo il beneficio di un recupero iniziale sul mantenimento varierà in maniera crescente in funzione della profondità o difficoltà di tale recupero iniziale. Questo secondo Bjork avverrebbe dal momento che un recupero più complesso e difficoltoso andrebbe a riattivare e a rafforzare la codifica di un elemento in memoria, incrementando allo stesso tempo il numero di vie di recupero per accedere a quella memoria. In questi termini, praticare il recupero sul materiale di studio costituirà una "difficoltà desiderabile" (Bjork, 1994), dal momento che tale pratica richiederà un maggior sforzo rispetto all'utilizzo di altre strategie di studio, come la lettura ripetuta del materiale, andando però allo stesso tempo a favorire un maggiore beneficio sul suo mantenimento a lungo termine (Karpicke, 2017). Allo stesso tempo però, se il test che viene fatto

svolgere è così difficile da far fallire spesso il recupero, il maggiore sforzo in questo processo andrà a discapito del successivo mantenimento (McDermott, 2021). Come rilevato da McDermott (2021), questa cornice teorica è una delle più popolari nel dibattito sul beneficio della pratica del recupero, ma allo stesso tempo presenta dei limiti. Essa, infatti, non specifica i meccanismi che dovrebbero produrre il beneficio mnestico del recupero sul successivo mantenimento, così come non è sempre chiaro cosa costituisca un recupero con sforzo (Karpicke et al., 2014). Inoltre, come rilevato da Karpicke (2017), proporre che le informazioni recuperate in memoria vengano rafforzate è essenzialmente una riproposizione del fenomeno della pratica del recupero, piuttosto che una sua spiegazione.

2.3.3. Il modello della biforcazione

Una diversa chiave di lettura è offerta da Kornell e colleghi (2011) e dal loro modello della biforcazione. Secondo questo modello, lo svolgere un test intermedio sul materiale di studio avrà come risultante una distribuzione biforcata degli elementi studiati. Il fenomeno può essere spiegato utilizzando lo stesso esempio proposto dagli autori, prendendo come riferimento due diversi set di elementi di studio che si distribuiscono normalmente: il set A, che verrà sottoposto ad un test intermedio, e il set B, che verrà sottoposto ad un ristudio. Degli elementi del set A sottoposti al test, alcuni verranno correttamente recuperati, per esempio il 57%, e questo recupero li andrà a potenziare ampiamente in termini di forza di memoria, a differenza degli elementi non recuperati, che non andranno incontro ad alcun potenziamento. Pertanto, la distribuzione del set A andrà incontro ad una biforcazione. Gli elementi del set B al contrario, venendo sottoposti al ristudio, verranno potenziati in maniera uniforme in termini di forza di memoria e si distribuiranno normalmente. Tuttavia, questo potenziamento per gli elementi ristudiati sarà inferiore rispetto a quello che si è verificato negli elementi correttamente recuperati nel set A. Con il passare del

tempo, la forza di memoria diminuirà in maniera simile per tutti gli elementi a causa del processo di dimenticanza; dopo un breve intervallo di tempo, il vantaggio uniforme conferito agli elementi del set B dal ristudio farà sì che una percentuale maggiore di questi elementi, ad esempio il 70%, superi ancora la soglia per il richiamo, rispetto a quanto avviene per gli elementi testati del set A, dove solo una porzione della distribuzione biforcata, il 57% in questo esempio, è andata incontro al potenziamento. Tuttavia, dopo un intervallo temporale più lungo e gli ulteriori effetti dell'oblio, il numero di elementi ristudiati che supera la soglia di richiamo diminuirà, scendendo ad esempio al 40%, mentre l'ampio vantaggio fornito dal recupero farà sì che in termini di forza di memoria tutto il 57% degli elementi testati e correttamente recuperati si troverà ancora al di sopra della soglia di richiamo. Oltre ad aver ricevuto evidenze empiriche da una serie di esperimenti svolti da Kornell e colleghi (2011), come rilevato da McDermott (2021), questo modello, da un lato è in linea con il dato per cui lo svolgere un test intermedio sia in grado di attenuare la curva di dimenticanza, dall'altro è coerente con il vantaggio a seguito del ristudio dopo un breve intervallo temporale dalla codifica rispetto a quanto osservato con il test, a differenza di quanto avviene dopo un intervallo temporale maggiore.

3. LO STUDIO ATTUALE

La memoria visiva a lungo termine è un sistema ad alta capacità (Standing, 1973), in grado di conservare un'elevata quantità di rappresentazioni dettagliate sia di immagini di singoli oggetti (Brady et al., 2008) che di scene complesse (Konkle et al., 2010). Di fronte a questo dato, ci si può chiedere se questo sistema possa essere ulteriormente potenziato. Diversi studi hanno preso in esame questo quesito eseguendo vari tipi di manipolazioni: dal variare la profondità di elaborazione durante la codifica (Baddeley & Hitch, 2017), alla manipolazione delle risorse attentive disponibili durante la codifica e della consapevolezza rispetto al compito di memoria (Evans & Baddeley, 2018), fino alla manipolazione della valenza emozionale delle immagini presentate (Dolcos et al., 2005).

La domanda relativa al potenziamento della memoria visiva a lungo termine è stata affrontata anche all'interno del nostro laboratorio di Parma, utilizzando come strumento di consolidamento mnestico lo svolgimento di due tipi di compito intermedio sulle immagini, presentate integre o degradate, codificate in precedenza: un compito di memoria vecchio/nuovo, che richiedeva un recupero episodico attivo della traccia di memoria per essere svolto, diversamente dall'altro compito previsto, quello di categorizzazione, che coinvolge la sola memoria semantica.

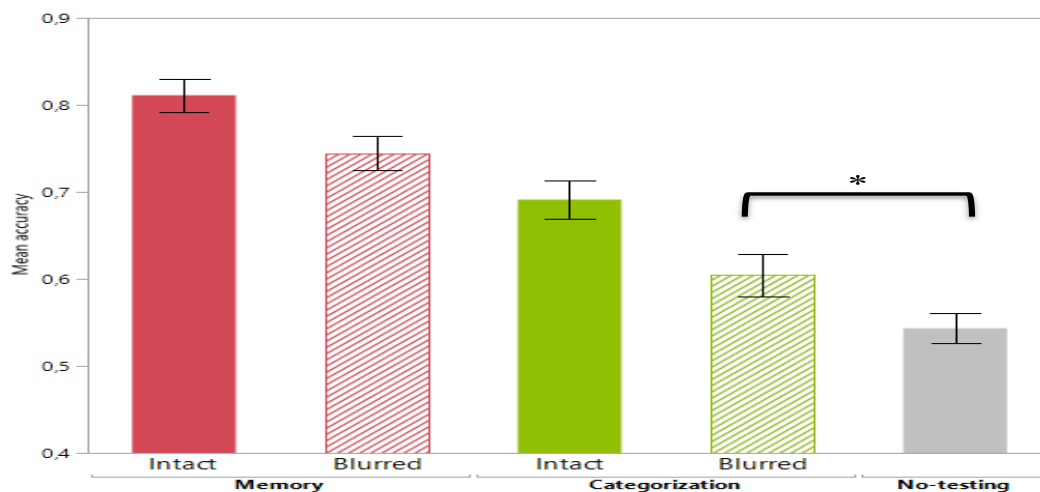


Figura 1: risultati dell'esperimento preliminare; accuratezza nel test finale in funzione del compito.

I risultati di questo esperimento preliminare hanno mostrato come lo svolgere un compito intermedio sulle immagini, sia integre che degradate, conferisse loro un vantaggio mnestico in un test finale somministrato una settimana dopo diversamente da quanto accadeva per le immagini non sottoposte ad alcun compito e codificate una singola volta in fase iniziale. Tuttavia, le immagini testate nel compito di memoria furono ricordate significativamente meglio di quelle sottoposte al compito di categorizzazione (Figura 1). Un risultato di questo tipo mette in evidenza un chiaro testing effect anche rispetto a stimoli come le scene naturali utilizzate. Recuperare la traccia di memoria nel compito vecchio/nuovo conduceva ad una migliore performance nel test finale, diversamente da quanto accadeva per le immagini non testate in cui questo recupero non avveniva. Anche la riesposizione allo stimolo assicurata dal compito di categorizzazione conferiva un modesto vantaggio nel test finale rispetto alle immagini non testate, tuttavia inferiore rispetto a quello legato al compito di memoria e al recupero episodico della traccia.

Tuttavia, è possibile offrire una diversa spiegazione di questi risultati inquadrandoli all'interno della cornice teorica della profondità di elaborazione (Craik & Lockhart, 1972; Craik & Tulving, 1975). È probabile che il tipo di compito intermedio che viene fatto svolgere al partecipante e la profondità di elaborazione che questo impone contribuiscano ad una ricodifica del materiale di studio durante questa fase. In questa prospettiva, il vantaggio legato al compito di memoria potrebbe essere ascritto al maggior grado di profondità associato ad esso, in termini di coinvolgimento semantico e di attenzione rivolta allo stimolo. Diversi studi negli anni hanno mostrato come il riconoscimento di una categoria target e l'estrazione del significato generale di un'immagine sia un processo altamente efficiente anche in condizioni molto degradate, sia con tempi molto bassi di esposizione dello stimolo presentato da solo sia quando questo è inserito all'interno di una presentazione visiva seriale rapida (vedi Potter, 2012). È logico supporre che questo processo di per sé già altamente efficiente non possa che diventarlo ulteriormente di fronte a un tempo di esposizione della singola immagine esteso a 3 secondi come nello studio preliminare sopracitato. Pertanto, sebbene il compito di categorizzazione implichi un ricorso alla memoria

semantica, di fronte all'istruzione di indicare se l'immagine presentata appartenga a categorie generali come "animali" o "mezzi di trasporto", è molto probabile che i soggetti portino a termine questo compito in una maniera relativamente superficiale e priva di difficoltà, facendo affidamento anche alla sola impressione dell'immagine per emettere la loro risposta, senza considerare appieno il significato dello stimolo (che l'immagine presentata sia un cane, un gatto o un leone, non fa alcuna differenza nell'emettere la risposta, dato che appartengono tutti alla stessa categoria). Questa ipotesi è avvalorata anche dai risultati della fase di test intermedio, che hanno mostrato come la performance di memoria nel compito di categorizzazione fosse vicina al 100% (Figura 2). Diversamente, di fronte all'istruzione di indicare se l'immagine è stata già vista oppure se si tratta di un'immagine nuova, oltre a recuperare l'immagine precedentemente codificata dalla memoria episodica i soggetti dovranno confrontare lo stimolo attualmente presentato con la propria rappresentazione mnestica. Per indicare se l'immagine attuale corrisponde a quella del cane che hanno già visto, ad esempio, piuttosto che indicare semplicemente se l'immagine presentata appartenga alla categoria generale "animale", i partecipanti dovranno considerare in maniera più fine e attenta il significato dello stimolo attualmente presentato (un cane è diverso da un leone o da un gatto) e allocare una maggiore quantità di risorse attentive per appurare che lo stimolo visto sia esattamente quello già codificato in precedenza, (ad esempio per appurare se l'immagine del Golden Retriever che vedo sullo schermo è lo stesso che ho già incontrato precedentemente o se si tratta di un'immagine simile di un cane della stessa razza). Dunque, sebbene il recupero episodico nel test di memoria possa avere un suo ruolo, i risultati potrebbero essere anche descritti da una profondità della ricodifica che avviene nella fase intermedia. Il mio lavoro di tesi si è incentrato sulla realizzazione di un esperimento, anch'esso basato su una versione modificata del classico

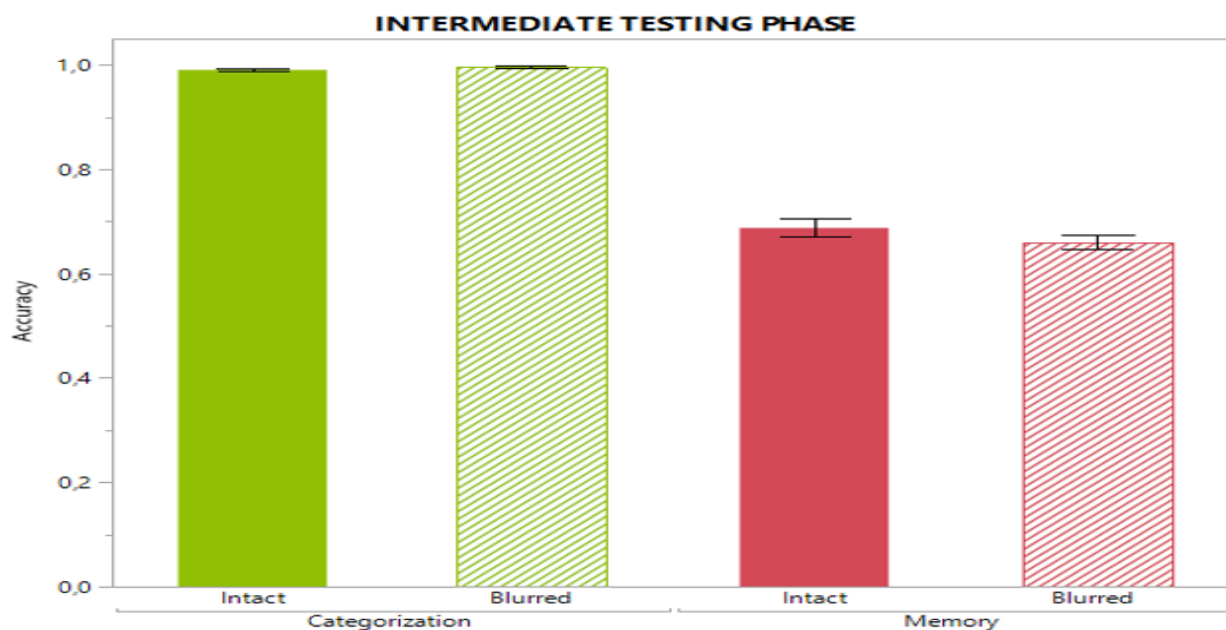


Figura 2: risultati della fase intermedia; accuratezza in funzione del compito e della qualità percettiva

paradigma del testing effect, in cui si intende testare l'ipotesi della profondità della codifica. Pertanto, era prevista una fase di codifica in cui venivano mostrate delle GIF animate appartenenti alle stesse due categorie semantiche dello studio precedente, mezzi di trasporto e animali, seguita da una seconda fase in cui alcuni di questi stimoli venivano sottoposti a due tipi di compito appaiati in termini di profondità di elaborazione, un rating di familiarità e un rating di piacevolezza, prima di essere sottoposte a un test finale una settimana dopo. Il ricorso al compito di piacevolezza è giustificato dalle evidenze riportate nel primo capitolo di questa tesi su come questo strumento rappresenti un efficace mezzo per il mantenimento a lungo termine tanto di materiale verbale quanto di materiale visivo come le scene naturali utilizzate nello studio attuale, promuovendo un'elaborazione semantica, in termini di significato, dello stimolo. Il rating di familiarità è stato scelto per avere un compito di memoria che richieda il recupero episodico attivo delle immagini diverso dal compito vecchio/nuovo utilizzato nell'esperimento preliminare. Allo stesso tempo, alcune delle scene viste durante la fase di codifica non venivano sottoposte ad alcun compito ma mostrate direttamente solo nel test finale. In riferimento a questo paradigma è possibile fare due

diverse previsioni: da un lato, se il vantaggio osservato nel test finale è legato al recupero episodico della traccia di memoria si dovrebbe osservare un vantaggio nel test finale per le immagini sottoposte alla valutazione di familiarità rispetto a quelle sottoposte al compito di piacevolezza e a quelle non sottoposte ad alcun compito, in cui questo recupero attivo della traccia non avviene; diversamente, se è la profondità della ricodifica nella fase intermedia a conferire il vantaggio mnestico allora si dovrebbe osservare un uguale vantaggio per le immagini sottoposte a due compiti appaiati in questi termini, in grado di promuovere lo stesso grado di analisi semantica e di mobilitare un'uguale quota di risorse attentive sullo stimolo, rispetto a quelle non sottoposte ad alcun compito in fase intermedia.

3.1. Materiali e metodi

3.1.1. Partecipanti

Per lo studio attuale sono stati reclutati 30 partecipanti dopo aver firmato un consenso informato. Il gruppo di partecipanti era composto da 19 donne e 11 uomini di età compresa tra i 18 e i 27 anni ($M=23.2$; $DS=2$). Prima dell'esperimento a ciascun partecipante è stato chiesto di compilare un questionario finalizzato a raccogliere alcune informazioni anagrafiche, relative alla presenza di patologie neurologiche, psicologiche o psichiatriche, allergie, problemi di vista, assunzione di farmaci, ore di sonno nella notte precedente, età, livello di istruzione e occupazione, genere e mano dominante. Tutti i soggetti hanno risposto a questo questionario. Dal momento che l'esperimento è stato eseguito interamente online, i partecipanti hanno ricevuto tramite e-mail istruzioni dettagliate circa le condizioni ottimali per svolgere entrambe le sessioni dell'esperimento, come la distanza da mantenere rispetto allo schermo, la necessità di svolgerle in una stanza in cui non erano presenti rumori o altro tipo di distrazioni.

3.1.2. Materiali e disegno

Gli stimoli utilizzati erano costituiti da 600 immagini neutre selezionate dal Web. 300 di queste appartenevano alla categoria “animali” (cani, gatti, conigli, uccelli, mucche, pecore, cavalli, animali selvatici come orsi, volpi e leoni) mentre le altre 300 appartenevano alla categoria “mezzi di trasporto” (automobili, biciclette, moto, camion, autobus). Per entrambe le categorie semantiche, erano presenti sia ambienti interni che esterni; inoltre, in alcune immagini era presente un solo elemento (per esempio un singolo animale o un singolo mezzo di trasporto) mentre in altre ne erano presenti due (nella stessa immagine, due mezzi di trasporto o due animali) (Figura 3).

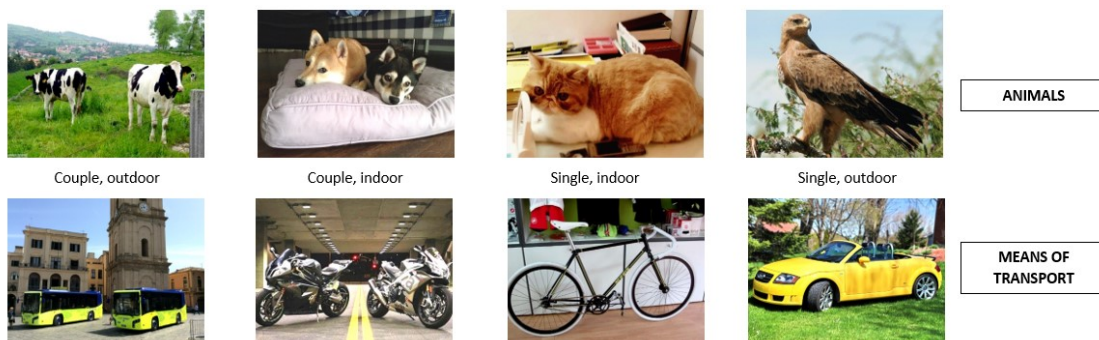


Figura 3: esempi degli stimoli utilizzati in questo esperimento

Dalle 600 immagini totali sono stati creati due set, ciascuno costituito da 300 immagini: 150 animali e 150 mezzi di trasporto. Entrambi i set sono stati bilanciati sia rispetto alla categoria semantica (animali vs mezzi di trasporto) ma anche rispetto ad altre tre condizioni: immagini in ambienti interni ed esterni; elementi presentati singolarmente o a coppie; la specie di animale e il tipo di mezzo di trasporto in modo tale che nei due set ci fossero, ad esempio, un numero simile di cani, gatti o auto. Tutte le immagini sono state rese omogenee rispetto ai seguenti parametri visivi

utilizzando il programma IrfanView: l'orientamento orizzontale, escludendo tutte le immagini con un orientamento verticale; la stessa dimensione (800x600). Delle 300 immagini che componevano ciascun set, 180 (90 animali e 90 mezzi di trasporto) sono state presentate come immagini dinamiche durante la prima fase definita come "Encoding" o codifica. 120 di queste sono state presentate nuovamente, durante la seconda fase definita "Fase intermedia", 10 minuti dopo la fase di encoding, assieme ad altre 120 immagini nuove (le rimanenti dalle 300 immagini del set). In questa seconda fase le immagini sono state divise equamente tra i due compiti (il rating di familiarità e il rating di valenza) e sono state presentate tutte percettivamente sfocate. Una settimana dopo, durante una fase di "Test finale", sono state presentate 360 immagini, tutte percettivamente integre, in un compito di scelta forzata a due alternative: 180 di queste erano immagini che i partecipanti avevano già visto precedentemente, durante la sola fase di encoding (60) o durante sia la fase di encoding che nella fase intermedia (120 delle 180 immagini). Le restanti 180 immagini erano immagini completamente nuove (Figura 4).

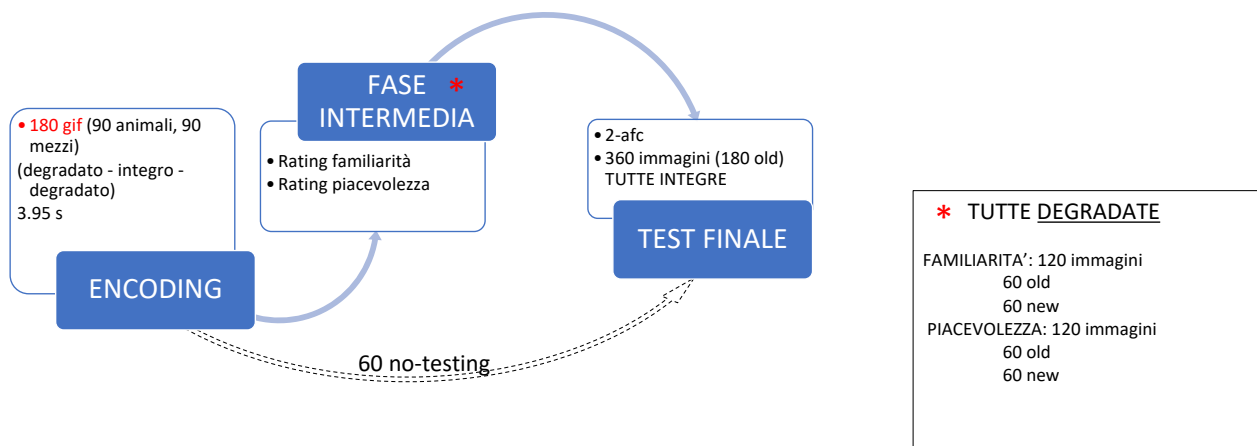


Figura 4: rappresentazione schematica del paradigma sperimentale

A partire dai due set, sono state costruite 4 sequenze (2 sequenze per ogni set), variando l'ordine di presentazione degli stimoli e la condizione a cui questi venivano assegnati (compito di familiarità, di piacevolezza o nessun compito). Dal set 1 sono state costruite le sequenze A e B; dal set 2 le sequenze C e D. L'ordine e la condizione assegnata agli stimoli in ciascuna sequenza erano sempre gli stessi e sono stati stabiliti in precedenza prima dell'inizio dell'esperimento. Questa pseudo-randomizzazione è stata scelta per non creare troppa variabilità tra i partecipanti. Durante questa fase preliminare è stata seguita un'importante regola per la creazione delle sequenze: nelle liste di immagini non vi erano più di tre stimoli consecutivi assegnati alla stessa condizione (categoria semantica, vecchio/nuovo, sinistra/destra nel compito finale).

3.1.3. Procedura

Encoding: dopo aver firmato un consenso informato (in cui veniva spiegato lo scopo generale dell'esperimento), i partecipanti hanno ricevuto una e-mail contenente sia le istruzioni per le condizioni ambientali ottimali da creare per svolgere l'esperimento (dal momento che questo è stato eseguito interamente online e non in laboratorio) sia i dati da inserire prima di iniziare l'esperimento. Prima di tutto, ai partecipanti è stato richiesto di posizionarsi ad una distanza di 50 cm dallo schermo del computer, in una stanza silenziosa, senza la presenza di possibili distrazioni, essendo sicuri di non venire disturbati durante lo svolgimento dell'esperimento; di rilevante importanza, i soggetti dovevano disporre di una connessione Internet stabile. Dopo aver cliccato il link fornito nella e-mail, i soggetti venivano condotti ad una pagina online in cui era presente una tabella in cui dovevano inserire un codice numerico specifico (in modo da salvaguardare l'identità dei partecipanti) e, alla voce "gruppo", una lettera dalla A alla D che indicava una delle 4 sequenze create per l'esperimento. A questo punto ai partecipanti veniva chiesto di attendere fino a quando la barra di download delle risorse non fosse diventata completamente grigia. Alla comparsa della frase

“tutte le risorse scaricate”, i partecipanti potevano cliccare “OK” per procedere. Successivamente, ai partecipanti era richiesto di compilare un questionario in cui dovevano indicare la presenza di patologie neurologiche, psicologiche o psichiatriche, allergie, problemi di vista, assunzione di farmaci, ore di sonno nella notte precedente, età, livello di istruzione e occupazione, genere e mano dominante. Completata questa parte preliminare, l’esperimento aveva inizio. A ciascun partecipante veniva data l’istruzione relativa al fatto che una serie di immagini sarebbero state presentate al centro dello schermo e che le avrebbero dovute semplicemente esplorare per tutta la durata dell’esposizione. Di seguito viene descritta la sequenza precisa: la comparsa di una croce di fissazione per 150 ms; una immagine dinamica (GIF) presentata per 4 secondi, che passava gradualmente dall’essere sfocata all’essere integra e poi nuovamente dall’essere integra all’essere sfocata (Figura 5); una croce di fissazione verde che i soggetti dovevano cliccare per procedere.

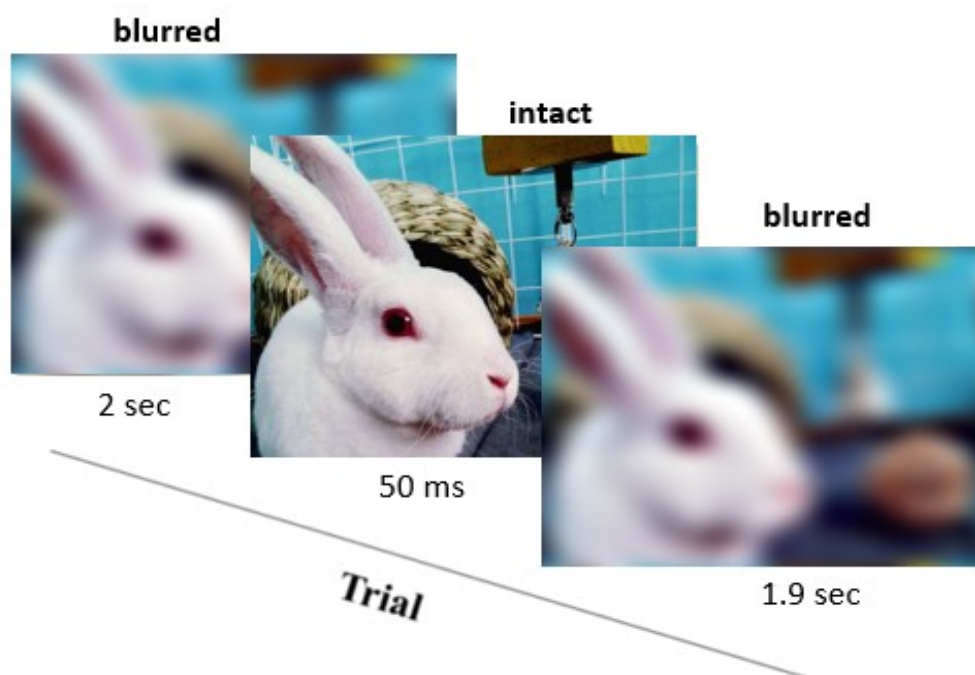


Figura 5: esempio di uno stimolo dinamico (GIF) utilizzato durante la fase di encoding: la stessa immagine passava gradualmente dallo sfocato all’intatto, e nuovamente dall’intatto allo sfocato.

Sono state utilizzate delle immagini dinamiche (GIF) durante questa fase di codifica dal momento che tutti gli stimoli utilizzati erano neutri, per favorire un maggiore coinvolgimento dei soggetti nella codifica. Questa prima fase è stata suddivisa in due blocchi: il blocco 1 consisteva di 45 immagini di animali e 45 di mezzi di trasporto; al termine del primo blocco vi era una pausa che durava due minuti (ma dopo un minuto, i soggetti potevano decidere di proseguire con il successivo blocco); il blocco 2 consisteva delle ultime 90 immagini. Per evitare possibili effetti di priorità o recenza, all'inizio e alla fine di ciascun blocco sono state inserite 4 immagini filler. Non sono state presentate più di 3 immagini consecutive appartenenti alla stessa categoria semantica. I quattro gruppi differivano rispetto all'ordine di presentazione dei due blocchi, in maniera da evitare possibili effetti legati all'ordine di presentazione: mentre nel gruppo A veniva presentato prima il blocco 1 e poi il blocco 2, nel gruppo B veniva presentato prima il blocco 2 e poi il blocco 1. Lo stesso veniva fatto dai gruppi C e D. Le 180 immagini dinamiche sono state costruite nel formato *.gif* utilizzando uno script di MATLAB e successivamente riconvertite nel formato *.mp4*, utilizzando CloudConvert, dal momento che la piattaforma Pavlovia, dove l'esperimento è stato caricato, non supportava il formato *.gif*. Ogni stimolo è stato presentato al centro dello schermo per un totale di 3.95 secondi: ciascuno stimolo dinamico era composto da 40 frame, a partire da un livello di degradazione di 4.0 per il frame 1, fino a un livello di degradazione di 9.0 per il frame 21 (che corrispondeva all'immagine intatta), per poi ritornare ad un livello di degradazione di 4.0 nel frame 40. Il cambiamento nella degradazione dell'immagine era graduale. Ciascun frame durava 100 ms, tranne per il frame 21 (che corrispondeva all'immagine intatta) che invece durava 50 ms. Di queste 180 immagini, 90 erano animali e 90 mezzi di trasporto; 120 immagini (60 animali e 60 mezzi di trasporto) sono state presentate nuovamente durante la fase intermedia, mentre le rimanenti 60 immagini (30 animali e 30 mezzi di trasporto) sono state presentate solo durante la fase di test finale, una settimana dopo. Alla fine della fase di encoding, i soggetti avevano a disposizione una pausa di 10 minuti. Durante la pausa, era loro consentito di allontanarsi temporaneamente dal computer e di fare altro. Quando la pausa stava per terminare (con un minuto

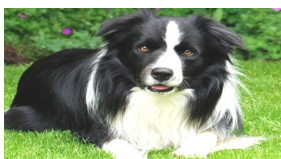
rimanente), un messaggio di avvertimento scritto in rosso compariva sullo schermo in modo tale da consentire ai soggetti di ritornare in tempo in posizione.

Fase intermedia: durante questa seconda fase ciascun partecipante doveva svolgere due compiti differenti: un rating di familiarità e un rating di piacevolezza: i due compiti venivano svolti separatamente dopo un intervallo di 5 minuti (di cui solo 3 minuti erano obbligatori), per evitare l'affaticamento dei soggetti. Per evitare un possibile effetto legato al compito che i soggetti svolgevano per primo, i soggetti assegnati al gruppo A o C svolgevano per primo il rating di familiarità, mentre i soggetti assegnati al gruppo B o D svolgevano prima quello di piacevolezza. All'inizio di ciascun compito venivano presentate 5 immagini filler per far prendere confidenza ai soggetti rispetto al compito. In entrambi i compiti, ai partecipanti veniva data l'istruzione di guardare le immagini presentate al centro dello schermo: inizialmente, compariva una croce di fissazione per 1 secondo; successivamente, lo stimolo, per una durata di 2 secondi, percettivamente degradato; infine, la domanda "quanto è sicuro di aver visto l'immagine nella fase precedente dell'esperimento?" per il rating di familiarità e la domanda "quanto è piacevole l'immagine?" per il rating di piacevolezza. Per entrambi i rating, era prevista una scala di risposta che andava da 1 (poco) a 4 (moltissimo). È stato scelto il range di valori da 1 a 4 per evitare che i soggetti potessero scegliere spesso un valore centrale se questo fosse stato presente e quindi per incoraggiare un giudizio più accurato nei due compiti. Le immagini, inoltre, in questa fase intermedia sono state presentate tutte percettivamente sfocate, con un valore di degradazione pari a 5.0. Si è deciso di presentarle sfocate dal momento che nell'esperimento precedente sopraccitato non è stata rilevata una interazione tra il tipo di compito svolto e il grado di degradazione dell'immagine. Pertanto, piuttosto che fornire entrambi i tipi di cue si è deciso di utilizzare solo immagini sfocate per rendere il compito più difficile ed osservare meglio i suoi effetti modulatori sulla memoria. I partecipanti dovevano rispondere utilizzando i tasti da 1 a 4 della tastiera e non era presente un tempo limite per emettere la risposta. Non sono stati presentati più di tre stimoli nuovi o vecchi consecutivi e non più

di tre di fila appartenenti alla stessa categoria semantica. Inoltre, nelle 4 sequenze, è stata ruotata anche la condizione di appartenenza delle immagini (compito di familiarità, piacevolezza, nessun compito). In ciascun compito sono stati presentati 120 stimoli (per un totale di 240 immagini): 60 immagini (30 animali e 30 mezzi di trasporto) erano state viste durante la fase di encoding, le altre 60 erano immagini nuove. Prima della fine di questa prima sessione sperimentale, i partecipanti dovevano rispondere ad alcune domande: indicare se c'erano stati problemi tecnici durante lo svolgimento dell'esperimento e, nel caso, quali; indicare in una scala di risposta da 1 a 5 il loro livello di stanchezza e quanto avevano trovato noioso l'esperimento.

Test finale: per questa fase finale, i partecipanti hanno ricevuto una seconda e-mail, contenente il link per accedere alla seconda sessione dell'esperimento, costituita da un unico blocco, e le stesse istruzioni presenti nella prima mail circa le condizioni ambientali ottimali da creare durante l'esperimento e i dati da inserire nella tabella che sarebbe apparsa sullo schermo prima di iniziare l'esperimento. Ai partecipanti veniva data l'istruzione che una serie di coppie di immagini sarebbero state presentate, rispetto alle quali avrebbero dovuto svolgere un compito di scelta forzata a due alternative (*two-alternative forced choice, 2AFC*): ai soggetti veniva richiesto di giudicare quale delle due immagini avevano già visto durante la prima sessione dell'esperimento. I partecipanti potevano rispondere premendo uno di due tasti sulla tastiera del computer: la lettera Z se l'immagine vecchia era presentata a sinistra o la lettera M se l'immagine vecchia era presentata a destra. Questa risposta veniva emessa durante la presentazione delle due immagini, incitando i partecipanti ad enfatizzare più l'accuratezza che la velocità nel fornire i loro giudizi. Pertanto, non erano presenti limiti di tempo nell'emettere la risposta. Si è deciso di utilizzare questo compito rispetto ad un compito vecchio/nuovo in accordo con quanto sostenuto da Brady e collaboratori (2023), secondo cui il primo compito sarebbe da preferire al secondo. In un compito vecchio/nuovo, in cui viene presentato un solo stimolo e il soggetto deve emettere un tipo di risposta se lo stimolo è già stato visto o un altro tipo se si tratta di uno stimolo nuovo, la risposta dei partecipanti sarà

influenzata dal fatto che uno stimolo vecchio o nuovo vada a generare un segnale di memoria al di sopra o al di sotto di un criterio di decisione adottato dai partecipanti. In un 2AFC, presentando uno stimolo vecchio casuale accoppiato ad uno stimolo nuovo casuale, i soggetti sceglieranno quello che evoca il segnale di memoria più forte. Pertanto, questo compito cattura l'intera distribuzione latente di forza di memoria dal momento che gli stimoli vecchio e nuovo in ciascun trial sono campioni casuali presi da un qualsiasi punto della distribuzione di memoria. Immediatamente dopo la risposta dei partecipanti, compariva un'altra schermata in cui i soggetti dovevano svolgere un rating di fiducia, dove veniva loro chiesto di specificare la sicurezza della loro precedente risposta tramite una scala di valutazione a 6 punti (1: poco; 6: moltissimo) (Figura 6). Durante questa fase, tutte le immagini presentate erano percettivamente intatte, per un totale di 360 immagini, 180 delle quali erano immagini vecchie mentre le altre 180 immagini erano completamente nuove. In questa fase finale, gli accoppiamenti di immagini, composti da una immagine nuova e una vecchia, sono stati realizzati in precedenza, randomizzando il loro ordine di presentazione, in modo tale che ci fosse la stessa sequenza stabilita per i soggetti assegnati ad ogni specifico gruppo. Per i gruppi A e B, in cui le immagini durante la fase di Encoding e in quella intermedia erano state prese dal set 1, le immagini nuove in questa fase finale sono state prese dal set 2; viceversa per i gruppi C e D.



2AFC

Z: l'immagine vecchia è a sinistra
M: l'immagine vecchia è a destra



Giudizio sulla sicurezza della
risposta nel 2AFC.

Figura 6: un esempio di trial nel test finale

Sono stati creati quattro tipi di accoppiamenti (Figura 7): 45 accoppiamenti “vecchia immagine di animale - nuova immagine di animale”; 45 accoppiamenti “vecchia immagine di mezzo – nuova immagine di mezzo”; 45 accoppiamenti “vecchia immagine di animale – nuova immagine di mezzo”; 45 accoppiamenti “vecchia immagine di mezzo – nuova immagine di animale”. La posizione di presentazione delle immagini vecchie, a destra o a sinistra, è stata bilanciata nei quattro tipi di accoppiamento. Sono state inoltre mantenute le seguenti regole: non più di tre immagini vecchie o nuove consecutive presentate nella stessa posizione e non più di tre accoppiamenti dello stesso tipo presentati consecutivamente. I quattro accoppiamenti sono stati anche bilanciati rispetto alla condizione di appartenenza degli stimoli nella fase intermedia: compito di familiarità, compito di piacevolezza, nessun compito. Queste tre condizioni sono state equamente distribuite nei quattro tipi di accoppiamenti ed equamente bilanciate rispetto alla posizione di presentazione a destra o a sinistra. Sebbene gli accoppiamenti fossero sempre di questi quattro tipi, essi venivano rieseguiti in ogni sequenza (A B C D). Pertanto, ognuna delle quattro sequenze aveva il proprio univoco 2AFC. Prima della fine dell’esperimento, i partecipanti dovevano rispondere ad alcune domande: indicare se ci fossero stati dei problemi tecnici durante l’esperimento e in caso indicare quali; indicare da 1 a 5 il loro livello di stanchezza e quanto avessero trovato noioso l’esperimento.



Figura 7: esempi dei quattro tipi di accoppiamenti utilizzati nel test finale

3.1.4. Analisi statistica dei dati comportamentali

In questo esperimento sono stati raccolti solo dati comportamentali che sono stati successivamente analizzati utilizzando i software JMP e SPSS. Rispetto alla fase intermedia sono stati indagati i punteggi di rating ed è stata eseguita una ANOVA 2x2 utilizzando come fattori la condizione delle immagini e il tipo di compito svolto. Nella fase di test finale, è stata dapprima considerata l'accuratezza nel 2AFC in funzione del tipo di compito svolto in fase intermedia eseguendo una ANOVA a tre livelli con il tipo di compito come fattore. Successivamente l'accuratezza nel 2AFC finale è stata considerata rispetto ai punteggi di rating forniti nei due compiti intermedi utilizzando i punteggi e il tipo di compito come fattori in un'ANOVA 4x2. In un'analisi successiva sono stati considerati i punteggi di rating di fiducia rispetto al tipo di compito svolto in fase intermedia e rispetto all'accuratezza nel test finale tramite un'ANOVA 3x2 con il tipo di compito e l'accuratezza come fattori. Infine, sono stati considerati gli effetti legati alla categoria semantica di appartenenza dell'immagine. I punteggi di rating in fase intermedia sono stati analizzati attraverso un'ANOVA 2x2 con la categoria semantica e il tipo di compito come fattori mentre successivamente è stata analizzata l'accuratezza nel 2AFC in fase finale eseguendo una ANOVA 3x2 con il tipo di compito svolto in fase intermedia e la categoria come fattori.

4. Risultati

4.1. Fase intermedia

La Fase intermedia è stata svolta 10 minuti dopo la fase di Encoding. Sono state presentate 240 immagini: 120 nel rating di familiarità, 120 nel rating di piacevolezza. In ciascun compito, la metà delle immagini era stata già incontrata nella fase di Encoding, l'altra metà era nuova. Sono stati

calcolati i punteggi medi di rating nei due compiti per le immagini vecchie e nuove (Figura 8). Una ANOVA 2x2 utilizzando come fattore la condizione (vecchie VS nuove) delle immagini e il compito (familiarità VS piacevolezza) in fase intermedia ha messo in evidenza un effetto principale della condizione [$F(1, 29) = 61.953, p < .001, \eta_p^2 = .681$] e un'interazione significativa [$F(1, 29) = 69.759, p < .001, \eta_p^2 = .706$]. Nel compito di familiarità, i soggetti hanno valutato come più familiari le immagini vecchie precedentemente incontrate nella fase di Encoding rispetto a quelle nuove [$F(1, 29) = 76.106, p < .001, \eta_p^2 = .724$]. Diversamente, nel compito di piacevolezza, non sono state rilevate differenze nei punteggi di rating di piacevolezza per le immagini vecchie viste in Encoding e quelle nuove [$F(1, 29) = 2.118, p = .156, \eta_p^2 = .068$].

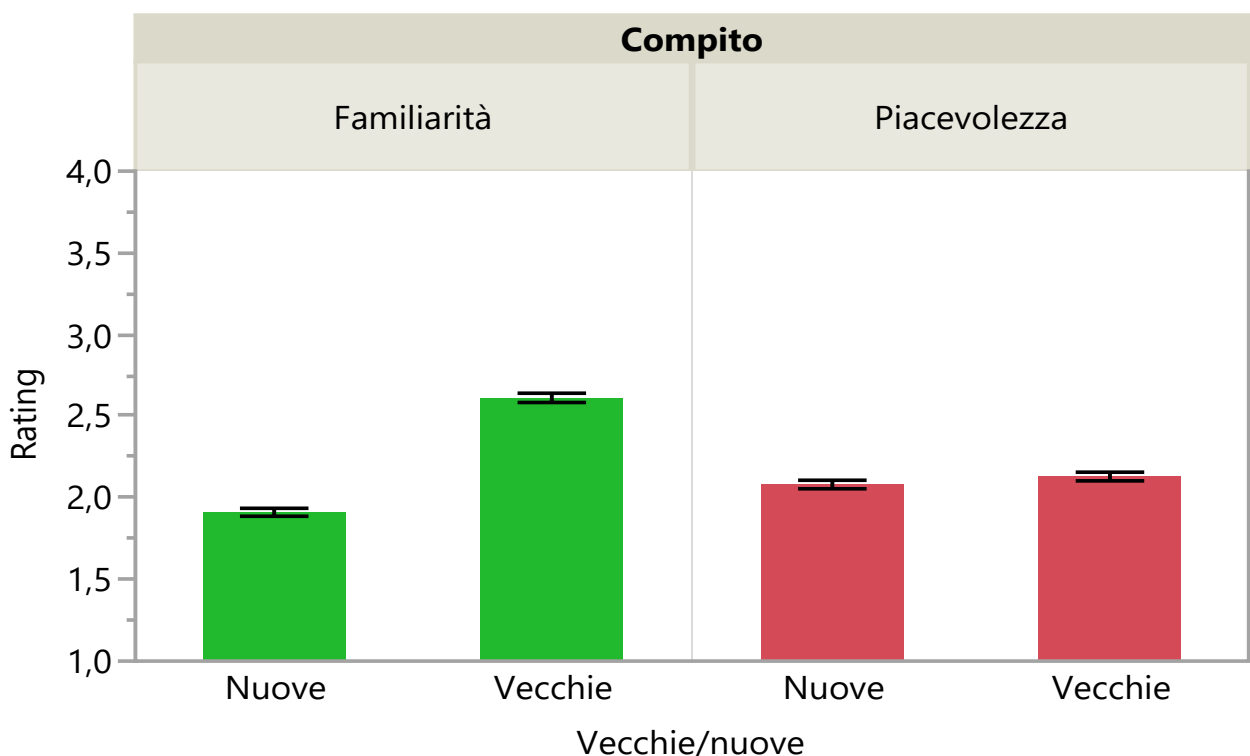


Figura 8: punteggi medi di rating nei due compiti per le immagini vecchie e nuove.

4.2. Test finale

4.2.1. Accuratezza nel 2AFC rispetto al compito svolto in fase intermedia

È stata calcolata l'accuratezza media nel 2AFC svolto una settimana dopo nella fase di Test finale in funzione del compito svolto sull'immagine nella fase intermedia (Compito di familiarità, piacevolezza, nessun compito) (Figura 9). Un'ANOVA a tre livelli utilizzando come fattore il compito svolto in fase intermedia (Familiarità VS Piacevolezza VS Nessun Compito) ha rilevato un effetto principale del compito [$F(2, 28) = 31.746, p < .001, \eta_p^2 = .694$]. In maniera critica, non è stata rilevata nessuna differenza significativa tra l'accuratezza nel 2AFC per le immagini sottoposte al rating di familiarità nella fase intermedia e quelle sottoposte al rating di piacevolezza [$F(1, 29) = 1.115, p = .300, \eta_p^2 = .037$]. Al contrario, per le immagini non sottoposte ad alcun compito in fase intermedia, l'accuratezza nel 2AFC è risultata significativamente inferiore rispetto a quella per le immagini sottoposte sia al rating di familiarità [$F(1, 29) = 57.275, p < .001, \eta_p^2 = .664$] che al rating di piacevolezza in fase intermedia [$F(1, 29) = 48.590, p < .001, \eta_p^2 = .626$].

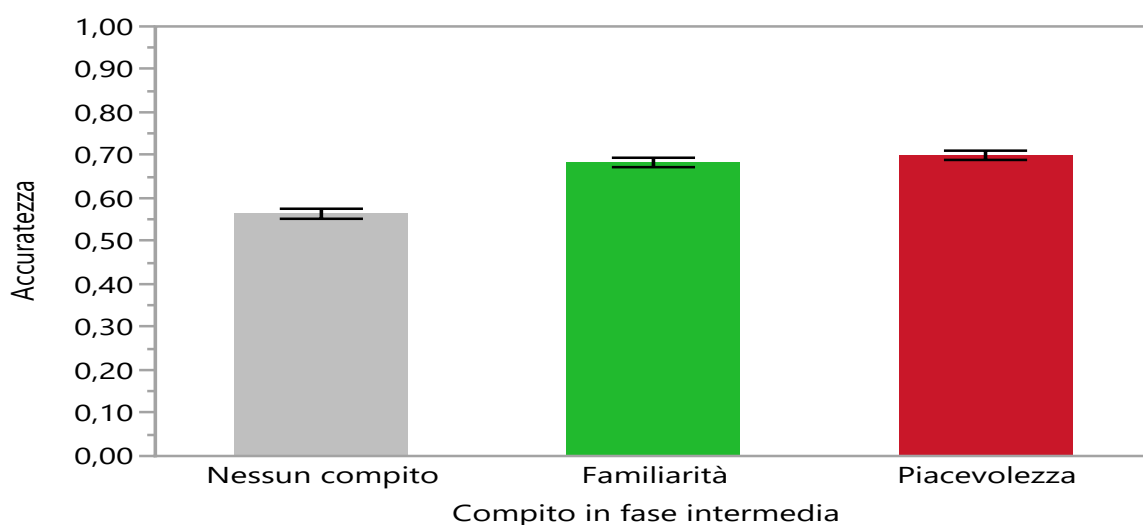


Figura 9: accuratezza nel 2AFC finale in funzione della condizione in fase intermedia

Successivamente, è stata misurata l'accuratezza nel 2AFC finale svolto una settimana dopo rispetto ai vari punteggi di rating in fase intermedia forniti nei due compiti, quello di piacevolezza e quello di familiarità (Figura 10). È stata dunque eseguita un'ANOVA 4x2 utilizzando come fattore i punteggi di rating in fase intermedia (1 VS 2 VS 3 VS 4) e il tipo di compito in fase intermedia (Familiarità VS Piacevolezza). È stato rilevato un effetto principale dei punteggi di rating [$F(3, 23) = 11.492, p < .001, \eta_p^2 = .600$] mentre non è stato rilevato un effetto principale del compito né un'interazione. In particolare, è stato rilevato un trend lineare rispetto ai soli punteggi di rating [$F(1, 25) = 24.431, p < .001, \eta_p^2 = .494$] indipendentemente dal tipo di compito svolto in fase intermedia.

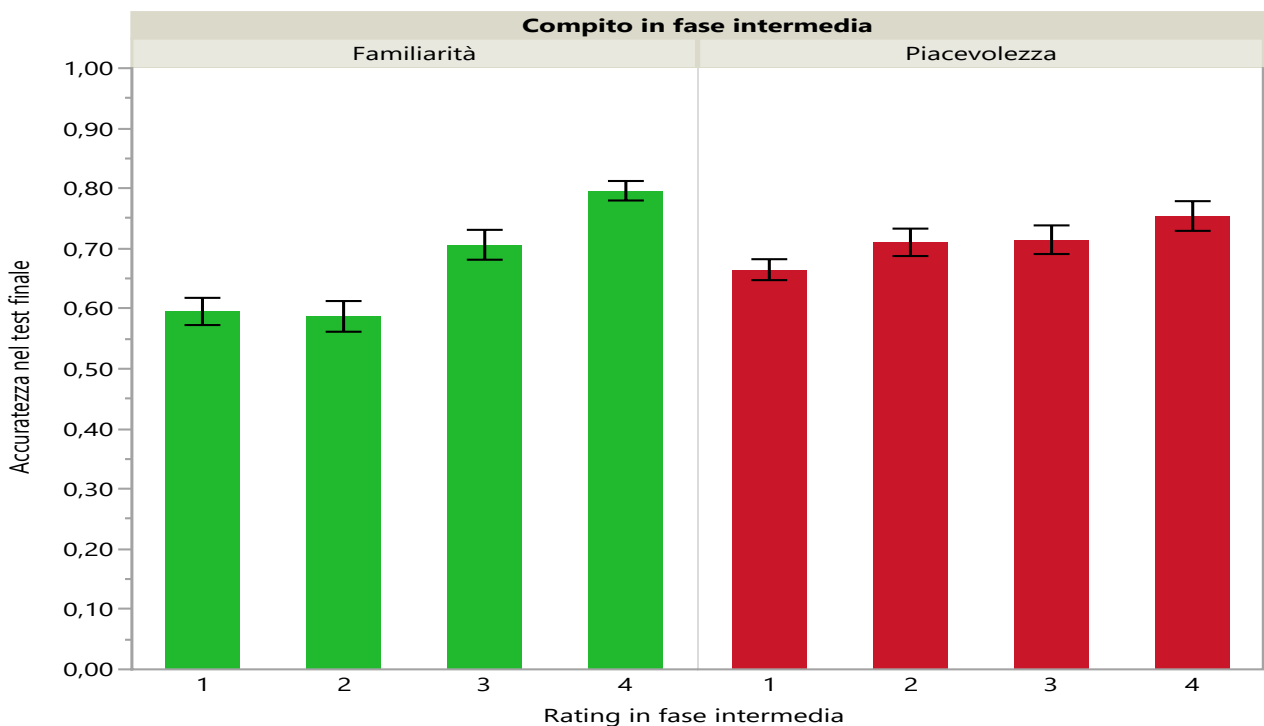


Figura 10: Accuratezza nel test finale in funzione dei punteggi di rating nei due compiti in fase intermedia.

4.2.2. Rating di fiducia nella fase di test finale

Nella fase di Test finale, i soggetti, dopo aver emesso la propria risposta nel 2AFC dovevano svolgere un rating di fiducia, indicando con un numero da 1 (Poco) a 6 (Moltissimo) quanto fossero sicuri della propria risposta. La figura 11 mostra la distribuzione dei punteggi di rating per le immagini sottoposte alle tre condizioni in fase intermedia, riconosciute e non riconosciute nel test finale. Tendenzialmente, i soggetti si sono dimostrati meno sicuri per quelle immagini non sottoposte ad alcun compito; sia quando le ricordano che quando non le ricordano, tendono a prevalere i punteggi di rating di fiducia più bassi. Diversamente, negli altri due compiti, tendono a prevalere i punteggi più bassi nel caso delle immagini non ricordate, mentre per quelle ricordate tendono a prevalere i punteggi estremi di rating di fiducia.

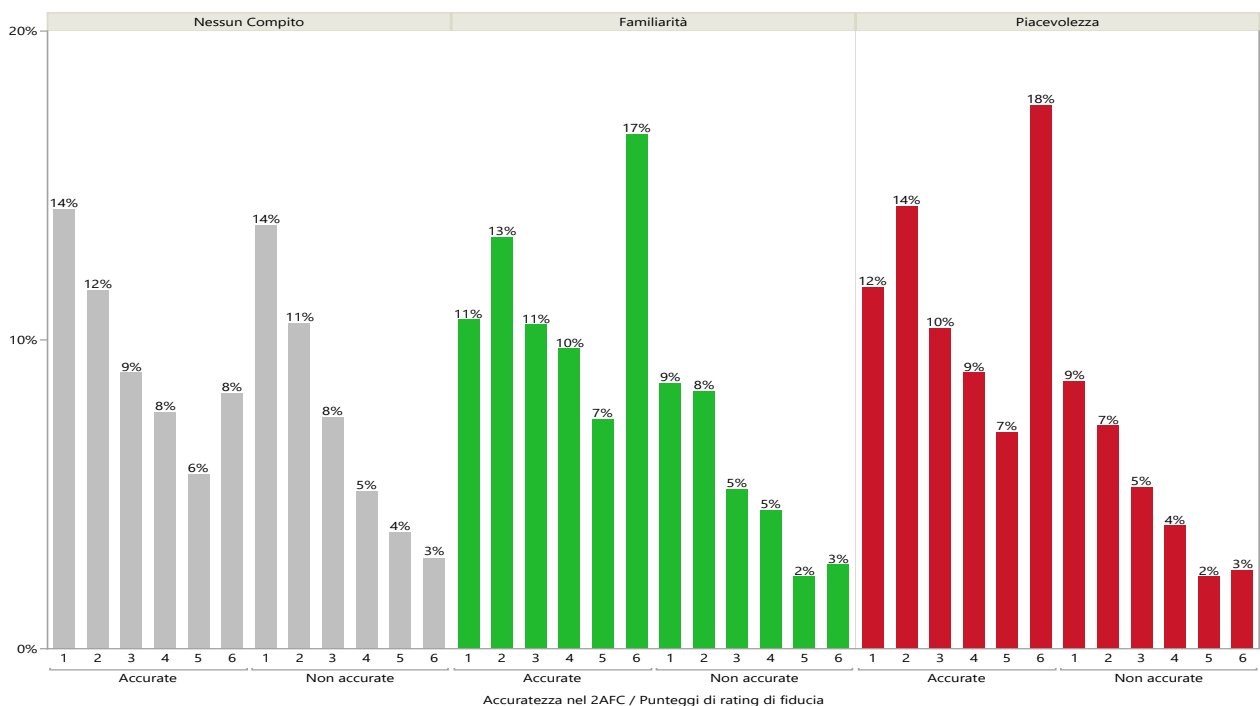


Figura 11: Distribuzione dei punteggi del rating di fiducia in percentuale, rispetto al compito intermedio e all'accuratezza nel test finale.

In maniera coerente, i soggetti si sono dimostrati più sicuri nel caso delle risposte accurate rispetto alla valutazione fornita nel caso delle risposte non accurate [$F(1, 29) = 68.925, p < .001, \eta_p^2 = .704$]. Sono stati calcolati i punteggi del rating di fiducia nella fase di test finale in funzione del compito svolto in fase intermedia, per le immagini correttamente riconosciute e per quelle non accurate (Figura 12). È stata quindi eseguita un'ANOVA 3x2 utilizzando come fattori il tipo di compito in fase intermedia (Nessun compito VS Piacevolezza VS Familiarità) e l'accuratezza nel test finale di scelta forzata a due alternative (Accurate vs Non accurate). È stato rilevato un effetto principale del compito [$F(2, 28) = 13.907, p < .001, \eta_p^2 = .498$], un effetto principale dell'accuratezza [$F(1, 29) = 67.357, p < .001, \eta_p^2 = .699$] e un'interazione anch'essa significativa [$F(2, 28) = 11.649, p < .001, \eta_p^2 = .454$]. Non è stata rilevata alcuna differenza significativa nei punteggi del rating di fiducia per le immagini non riconosciute sottoposte ai tre tipi di compiti in fase intermedia, mentre nel caso delle immagini accurate, il non svolgere alcun compito in fase intermedia ha condotto ad una sicurezza nelle proprie risposte espressa attraverso i punteggi di rating inferiore rispetto a quella per le immagini sottoposte al compito di familiarità [$F(1, 29) = 46.791, p < .001, \eta_p^2 = .617$] e a quello di piacevolezza [$F(1, 29) = 63.831, p < .001, \eta_p^2 = .688$]. In maniera analoga a quanto rilevato per l'accuratezza, anche nel caso dei punteggi del rating di fiducia per le immagini accurate non è stata rilevata nessuna differenza significativa tra quelle sottoposte al compito di piacevolezza e al compito di familiarità in fase intermedia [$F(1, 29) = .477, p = .495, \eta_p^2 = .016$].

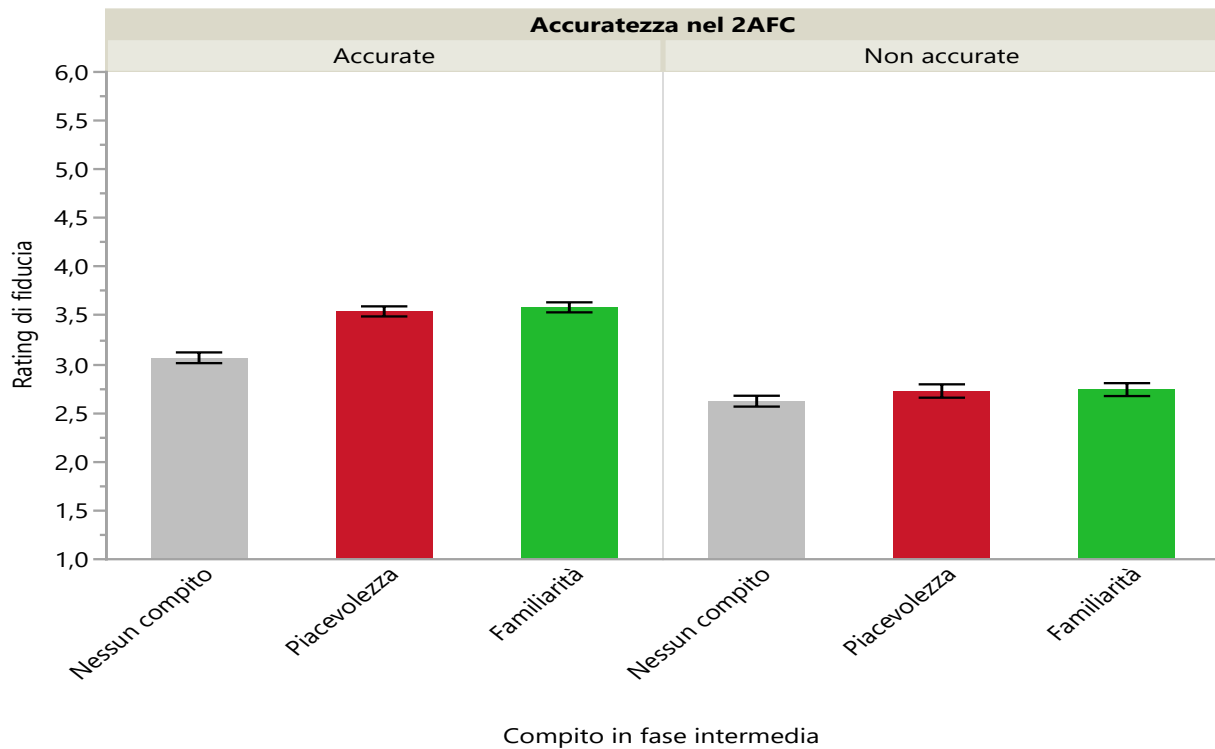


Figura 12: Rating di fiducia in funzione del tipo di compito in fase intermedia, per le immagini accurate e non accurate.

4.3. Effetti legati alla categoria dell'immagine

In questo esperimento, ai soggetti venivano presentate immagini appartenenti a due categorie semantiche: animali e mezzi di trasporto. Innanzitutto, sono stati calcolati i punteggi di rating nella fase intermedia nei due compiti in funzione della categoria di appartenenza dell'immagine (Figura 13). Le immagini considerate in questa analisi erano tutte immagini vecchie già incontrate nella fase di Encoding. È stata quindi eseguita un'ANOVA 2x2 inserendo come fattore il tipo di categoria semantica (Animali VS Mezzi) e il tipo di compito in fase intermedia (Familiarità VS Piacevolezza). È stato rilevato un effetto principale della categoria [$F(1, 29) = 89.306, p < .001, \eta^2$

= .755], un effetto principale del compito [$F(1, 29) = 30.880, p < .001, \eta^2 = .516$] e un'interazione significativa [$F(1, 29) = 40.381, p < .001, \eta^2 = .582$]. Nel compito di familiarità, i soggetti hanno valutato le immagini di animali come più familiari rispetto alle immagini di mezzi di trasporto [$F(1, 29) = 20.053, p < .001, \eta^2 = .409$], mentre nel compito di piacevolezza, le immagini di animali sono state valutate come più piacevoli rispetto a quelle di mezzi di trasporto [$F(1, 29) = 87.217, p < .001, \eta^2 = .750$]. Successivamente, è stata calcolata l'accuratezza nel 2AFC finale svolto una settimana dopo in funzione del tipo di categoria di appartenenza dell'immagine e del tipo di compito in cui è stata presentata l'immagine durante la fase intermedia (Figura 14). Pertanto, è stata eseguita un'ANOVA 3x2 utilizzando come fattori il tipo di compito in fase intermedia (Piacevolezza VS Familiarità VS Nessun compito) e il tipo di categoria (Animali VS Mezzi). È stato rilevato solo un effetto principale del compito [$F(2, 28) = 31.746, p < .001, \eta^2 = .694$]. Le immagini di animali non sottoposte ad alcun compito in fase intermedia sono state ricordate peggio nel compito finale rispetto a quelle sottoposte al compito di familiarità [$F(1, 29) = 22.366, p < .001, \eta^2 = .435$] e a quelle sottoposte al compito di piacevolezza [$F(1, 29) = 20.694, p < .001, \eta^2 = .416$]. Lo stesso pattern è stato rilevato anche per le immagini di mezzi di trasporto, con quelle non sottoposte ad alcun compito ricordate peggio rispetto a quelle sottoposte al compito di familiarità [$F(1, 29) = 45.407, p < .001, \eta^2 = .610$] e a quelle sottoposte al compito di piacevolezza [$F(1, 29) = 34.268, p < .001, \eta^2 = .542$]. In maniera consistente, non sono state rilevate differenze nell'accuratezza né tra le immagini di animali sottoposte al compito di piacevolezza e quelle sottoposte al compito di familiarità, né per quelle dei mezzi sottoposte all'uno o all'altro compito. Infine, indipendentemente dal tipo di compito a cui sono state sottoposte in fase intermedia, non sono state rilevate differenze significative nell'accuratezza per le immagini di animali e mezzi.

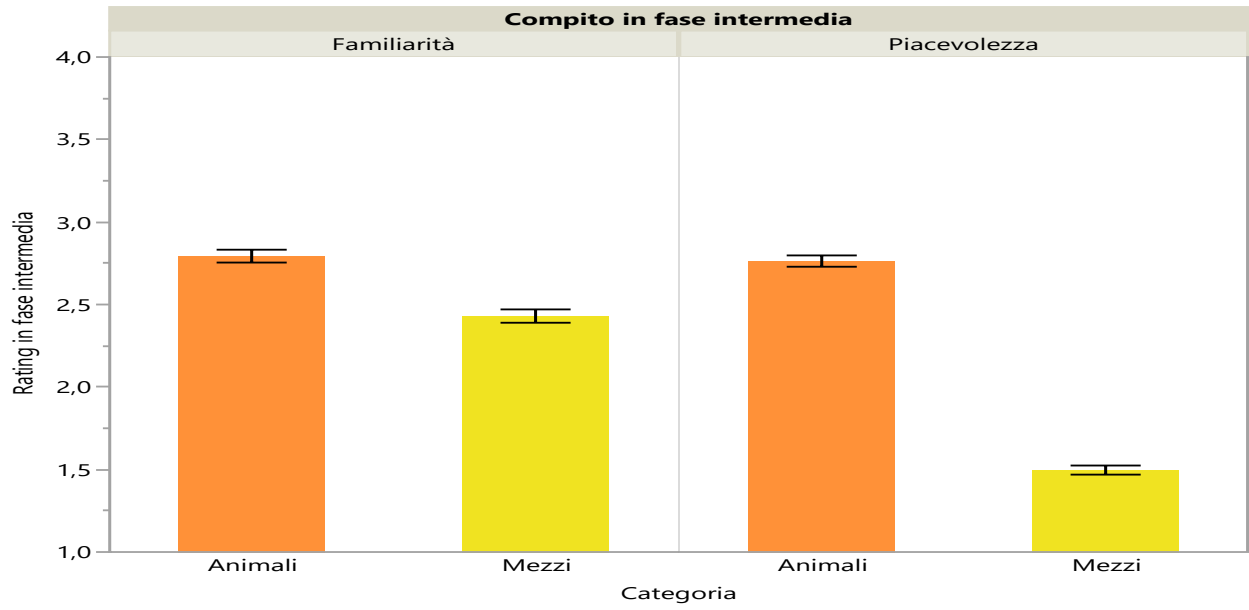


Figura 13: punteggi di rating nei due compiti in fase intermedia in funzione della categoria semantica

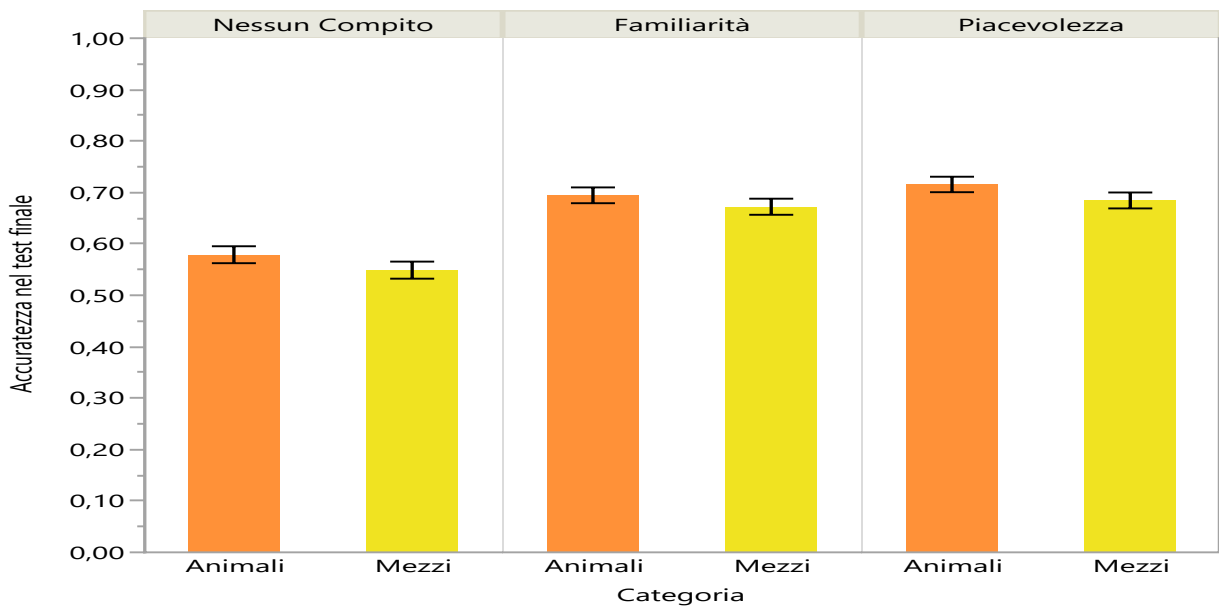


Figura 14: Accuratezza nel 2AFC finale in funzione della categoria di appartenenza dell'immagine e del compito cui è stata sottoposta in fase intermedia.

5. Discussione

In questo studio l'indagine si è incentrata sulla possibilità che la memoria a lungo termine visiva per scene naturali, un sistema già di per sé ad alta capacità (Standing, 1973), potesse essere ulteriormente modulata. In particolare, la manipolazione eseguita riguardava il tipo di compito svolto nella fase intermedia di una versione modificata del paradigma tipicamente utilizzato negli studi sul testing effect (Roediger & Karpicke, 2006a; 2006b), consistente di tre fasi. Nella prima fase, quella di Encoding, sono state presentate delle immagini dinamiche (GIF) appartenenti a due categorie semantiche neutre (animali e mezzi di trasporto), per la durata di 4 secondi, nei quali l'immagine passava da uno stato di degradazione a uno di integrità percettiva per poi ritornare nuovamente ad uno stato di degradazione. La scelta di utilizzare delle gif in questa fase è motivata dalla volontà di favorire un maggior coinvolgimento dei soggetti di fronte alla presentazione di immagini tutte emozionalmente neutre. Successivamente, era prevista una fase intermedia in cui una parte delle immagini già viste in fase di Encoding veniva nuovamente presentata e sottoposta a due tipi di compito, mentre le rimanenti venivano ripresentate direttamente nell'ultima fase, quella di test finale, una settimana dopo. In questa fase intermedia è stato dunque manipolato il tipo di compito svolto sulle immagini ripresentate: un rating di familiarità, in cui i soggetti dovevano giudicare quanto trovassero familiari le immagini presentate e un rating di piacevolezza, in cui i soggetti dovevano giudicare quanto trovassero piacevole l'immagine presentata. Il primo è un compito che per essere svolto richiede un recupero episodico attivo della traccia di memoria; il secondo è il compito di elezione utilizzato negli studi sulla profondità di elaborazione per garantire una codifica profonda in termini di analisi semantica dello stimolo. Si è scelto di utilizzare questi due rating in modo tale da avere due compiti appaiati in termini di profondità di elaborazione (Craik & Lockhart, 1972; Craik & Tulving, 1975) e quindi di analisi semantica e di risorse attentive rivolte allo stimolo ma che allo stesso tempo differiscono in virtù dei processi coinvolti, con il recupero episodico richiesto dal solo rating di familiarità diversamente dal rating di piacevolezza dove questo

non è necessario per svolgere il compito. L'obiettivo di questo studio era quindi quello di indagare i processi che entrano in gioco nel fornire un eventuale vantaggio mnestico in un test finale somministrato dopo una settimana. Nell'ultima fase, pertanto, eseguita dopo una settimana, i soggetti svolgevano un compito di memoria nel formato di una scelta forzata a due alternative (*two-alternative forced choice, 2AFC*), considerato da Brady e colleghi (2023) come quello migliore e da preferire nell'indagine della performance di memoria.

In riferimento a quest'ultima, nel 2AFC finale per le immagini sottoposte ad un compito, indipendentemente da quale di questi si trattasse, rispetto a quelle non sottoposte ad alcun compito ma viste in fase di Encoding e riviste solo nel test finale una settimana dopo, è stato nuovamente rilevato un testing effect per le scene naturali, in maniera coerente con i risultati dell'esperimento preliminare sopraccitati. Le immagini sottoposte ad un compito, indipendentemente dal fatto che si trattasse del compito di familiarità o di piacevolezza, sono state ricordate meglio rispetto alle immagini non sottoposte ad alcun compito ma mostrate solo nella fase di Encoding e poi direttamente in quella di Test finale (Figura 9). Se questo dato è coerente con la letteratura che mostra come lo svolgere un test sul materiale di studio possa migliorare la performance di memoria in un test di memoria finale, il confronto tra la performance di memoria nel 2AFC svolto una settimana dopo per le immagini sottoposte al compito di familiarità e quelle sottoposte al compito di piacevolezza consente di gettare luce sui processi coinvolti nel fornire questo vantaggio, e di mettere in dubbio l'idea tradizionale che nel testing effect questo sia sempre legato al recupero del materiale dalla memoria (Karpicke, 2017). Infatti, confrontando queste due condizioni, è emerso come non ci fosse nessuna differenza significativa nella performance di memoria nel test finale per le immagini sottoposte a questi due compiti in fase intermedia (Figura 9), un dato coerente con l'interpretazione che non sia tanto il recupero del materiale di studio dalla memoria a conferire il vantaggio mnestico (dal momento che questo recupero non è richiesto nel compito di piacevolezza), quanto la profondità di elaborazione (Craik & Lockhart, 1972; Craik & Tulving, 1975) in termini di analisi semantica e di risorse attentive rivolte allo stimolo nel momento in cui questo viene

“ricodificato” all’interno della fase intermedia nei due compiti.

Se questo è stato rilevato in merito all’accuratezza nel 2AFC, anche i risultati del rating di fiducia che veniva svolto dopo l’emissione della risposta conducono ad una conclusione analoga. Nel momento in cui i soggetti ricordano l’immagine vecchia (Figura 12, parte sinistra del grafico), le immagini sottoposte ai due compiti in fase intermedia sono associate ad una maggiore sicurezza nella risposta emessa nel test di memoria rispetto a quelle non testate; tuttavia, nuovamente, svolgere il compito di familiarità o di piacevolezza non ha condotto ad una differenza nella sicurezza della propria risposta tra i due compiti. I risultati delle analisi sui punteggi di rating intermedio nei due compiti (Figura 8) e sull’accuratezza nel 2AFC in funzione dei diversi punteggi di rating intermedio (Figura 10) mostrano anche come la memoria per l’immagine vada a modulare effettivamente i punteggi di familiarità; le immagini vecchie sono valutate come più familiari rispetto a quelle nuove. Al contrario, le immagini vecchie non sono state valutate come più piacevoli di quelle nuove. Allo stesso tempo, rispetto all’accuratezza nel 2AFC finale, è stato rilevato solamente un trend lineare rispetto ai punteggi di rating, mentre questo non è emerso per il tipo di compito così come non è emersa un’interazione significativa tra il compito e i punteggi di rating. Pertanto, è solo un miglior punteggio fornito all’immagine nel compito intermedio, senza nessuna specificità del tipo di compito in cui questo è emesso, a modulare la memoria a lungo termine per l’immagine. Che il vantaggio mnestico sia attribuibile alla profondità della ricodifica in fase intermedia nei due compiti è suggerito anche dalle analisi sulle due categorie semantiche di immagini mostrate ai soggetti.

In termini di contenuto semantico delle immagini, i soggetti hanno valutato come più familiari le immagini di animali già incontrate rispetto a quelle dei mezzi di trasporto, anch’esse già incontrate in precedenza; allo stesso tempo, gli animali sono stati valutati come più piacevoli rispetto ai mezzi trasporto (Figura 13). A fronte di questa modulazione nei due compiti in fase intermedia, come mostrato in Figura 14, le analisi hanno mostrato come non ci fosse alcuna differenza nell’accuratezza nel 2AFC finale una settimana dopo per gli animali e i mezzi sottoposti

sia al compito di familiarità che a quello di piacevolezza. Pertanto, sebbene gli animali venissero recuperati meglio nella fase intermedia rispetto ai mezzi e giudicati come più piacevoli, questo non si rifletteva in una migliore performance per questa categoria semantica nel test finale svolto una settimana dopo. Sebbene i risultati ottenuti siano consistenti e vadano tutti nella direzione secondo cui sia il fatto di ingaggiare lo stimolo in un compito che richieda un certo grado di analisi semantica e cognitiva affinché questo venga poi ricordato meglio, e non il recupero mnestico dello stesso, si potrebbe anche ipotizzare che tale vantaggio sia semplicemente legato ad una riesposizione allo stimolo in fase intermedia. Ovvero, in entrambi i compiti, il soggetto rivede lo stimolo diversamente da quanto accade per le immagini non sottoposte ad alcun compito, e questo conduce ad un vantaggio mnestico una settimana dopo nel test finale. Tuttavia, questa interpretazione non è sostenuta dai risultati dell'esperimento precedente, in cui si osserva che lo svolgere un compito vecchio/nuovo nella fase intermedia ha conferito un vantaggio mnestico maggiore rispetto a quanto ottenuto dallo svolgere un compito di categorizzazione, a parità di riesposizione. Allo stesso modo, i risultati ottenuti in questo esperimento non sono interpretabili nei termini di sforzo nel recupero (Bjork, 1975), dal momento che il compito di piacevolezza, in virtù di quelle che sono le istruzioni fornite ai partecipanti, verosimilmente non richiede questo processo. In relazione a quest'ultimo punto, non è da escludere che anche il compito di piacevolezza implichi un recupero del materiale di studio nel momento in cui viene fornita questa valutazione. Il paradigma utilizzato nel presente studio, tuttavia, non consente di esplorare a pieno questa possibilità. È rilevante tuttavia osservare che il rating di piacevolezza non differisce per immagini vecchie e nuove in fase intermedia, suggerendo quindi che i due processi, familiarità e piacevolezza, siano in qualche modo indipendenti. Studi successivi potrebbero indagare se anche il compito di piacevolezza implichi un recupero del materiale di studio facendo affidamento a sonde diverse dalla risposta comportamentale e a misure implicite di memoria, ad esempio utilizzando strumenti come l'elettroencefalogramma o l'eye-tracker. Vi sono indicazioni in letteratura di come sia possibile ricavare una misura indiretta di memoria estraendo i potenziali evocati dal tracciato

EEG (*event-related potentials*, *ERP*, vedi Rugg & Curran, 2007, per una review), sebbene un precedente studio abbia fallito nell'evidenziare questi indici di memoria di riconoscimento in un contesto di memoria implicita di immagini neutre (Ferrari et al., 2013). In riferimento ai movimenti oculari, vi sono indicazioni in letteratura di una modulazione del comportamento oculare in funzione della memoria per le immagini (Damiano & Walther, 2019).

In conclusione, i risultati di questo esperimento hanno messo in evidenza come lo svolgere un compito intermedio sul materiale di studio costituito da scene naturali sia in grado di condurre ad un miglioramento della performance di memoria in un test finale svolto dopo una settimana. Tuttavia, nel caso di questi stimoli, ricchi di dettagli visivi, la natura specifica del processo implicato in ciascun compito non sembra essere il fattore cruciale nel conferire il vantaggio mnemonico, diversamente da un'elaborazione profonda in termini di analisi semantica e risorse attentive rivolte allo stimolo assicurata da entrambi i compiti utilizzati.

BIBLIOGRAFIA

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation*, 2, 89-195.

Baddeley, A. D., & Hitch, G. J. (2017). Is the Levels of Processing effect language-limited? *Journal of Memory and Language*, 92, 1-13.

Bjork, R.A. (1975). Retrieval as a memory modifier: an interpretation of negative recency and related phenomena. In Solso, R.L. (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 123–144). Erlbaum, Hillsdale, NJ.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). The MIT Press.

Bjork, R.A., & Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In Healy, A.F., Kosslyn, S.M., & Shiffrin, R.M. (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes*, vol. 2 (pp. 35–67). Erlbaum, Hillsdale, NJ.

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, 103(4), 751-757.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38), 14325–14329.

- Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, *30*, 421–449.
- Carpenter, S. K., & DeLosh, E. L. (2005). Application of the Testing and Spacing Effects to Name Learning. *Applied Cognitive Psychology*, *19*, 619-636.
- Carpenter, S. K., & Pashler H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, *14*(3), 474-478.
- Coppens, L. C., Verkoeijen, P. P. J. L., & Rikers, R. M. J. P. (2011). Learning Adinkra symbols: The effect of testing. *Journal of Cognitive Psychology*, *23*(3), 351-357.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of Processing and the Retention of Words in Episodic Memory. *Journal of Experimental Psychology: General*, *104*(3), 268-294.
- Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding and retrieval. *Cognition*, *184*, 119-129.
- Dolcos, F., LaBar, K. S., & Cabeza, R. (2005). Remembering one year later: role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(7), 2626–2631.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58.

- Evans, K. K., & Baddeley, A. (2018). Intention, attention and long-term memory for visual scenes: It all depends on the scenes. *Cognition*, *180*, 24–37.
- Ferrari, V., Bradley, M. M., Codispoti, M., Karlsson, M., & Lang, P. J. (2013) Repetition and brain potentials when recognizing natural scenes: task and emotion differences. *Social Cognitive and Affective Neuroscience*, *8*(8), 847-854.
- Gates, A.I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *40*.
- Gibson, J. J. (1977). The theory of affordances. In R. Shaw, & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ: Erlbaum.
- Goldstein, E. B. 2016. *Psicologia cognitiva: Connettere mente, ricerca, ed esperienza comune*. Piccin Nuova Libreria S.p.A., Padova.
- Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, *82*(3), 472-481.
- Hyde, T. S., & Jenkins, J. J. (1973). Recall for Words as a Function of Semantic, Graphic, and Syntactic Orienting Tasks. *Journal of Verbal Learning and Verbal Behavior*, *12*, 471-480.
- Intraub, H., & Nicklos, S. (1985). Levels of Processing and Picture Memory: The Physical Superiority Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(2), 284-298.
- Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition*, *38*(8), 1009-1017.
- Karpicke, J.D. (2017). Retrieval-based learning: A decade of progress. In Wixted, J. T. (Ed.), *Cognitive Psychology of Memory, Learning and Memory: A Comprehensive Reference* (Volume 2,

pp. 487–514). Academic Press: Cambridge, MA, USA.

Karpicke, J. D., Butler, A. C., & Roediger, H. L. III (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, *17*(4), 471-479.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: the role of categories in visual long-term memory. *Psychological science*, *21*(11), 1551–1556.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*, 85-97.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In Gruneberg, M. M., Morris, P. E., & Sykes, R. N. (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

McDermott, K. B. (2021). Practicing Retrieval Facilitates Learning. *Annual Review of Psychology*, *72*, 609-633

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of Processing Versus Transfer Appropriate Processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519-533.

Morris, P. E., & Fritz, C. O. (2002). The improved name game: better use of expanding retrieval practice. *Memory*, *10*(4), 259–266.

Osiurak, F., Rossetti, Y., & Badets, A. (2017). What is an affordance? 40 years later. *Neuroscience and Biobehavioral Reviews*, *77*, 403-417.

Packman, J. L., & Battig, W. F. (1978). Effects of different kinds of semantic processing on memory for words. *Memory & Cognition*, *6*(5), 502-508.

- Patterson, K. E., & Baddeley, A.D. (1977). When face Recognition Fails. *Journal of Experimental Psychology: Human Learning and Memory*, 3(4), 406-417.
- Postman, L., & Kruesi, E. (1977). The Influence of Orienting Tasks on the Encoding and Recall of Words. *Journal of Verbal Learning and Verbal Behavior*, 16, 353-369.
- Potter, M. C. (2012). Recognition and memory for briefly presented scenes. *Frontiers in Psychology*, 3(32), 1-9.
- Roediger, H. L. III, & Karpicke, J. D. (2006a). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 17(3), 249-255.
- Roediger, H. L. III, & Karpicke, J. D. (2006b). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, 1(3), 181-210.
- Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests Enhance the Transfer of Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 233-239.
- Rowland, C. A. (2014). The Effect of Testing Versus Restudy on Retention: a Meta-Analytic Review of the Testing Effect. *Psychological Bulletin*, 140(6), 1432-1463.
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, 11(6), 251-257.
- Schulman, A. I. (1974). Memory for words recently classified. *Memory and Cognition*, 2(1A), 47-52.
- Spitzer, H. F. (1939). Studies in retention. *The Journal of educational psychology*, 30(9), 641-656.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 25, 207-222.

Tse, C. S., Balota, D. A., & Roediger, H. L. III (2010). The Benefits and Costs of Repeated Testing on the Learning of Face-Name Pairs in Healthy Older Adults. *Psychology and Aging*, 25(4), 833-845.

Tulving, E. (2000). Concepts of memory. In Tulving, E., Craik, F. I. M. (Eds.), *The Oxford Handbook of Memory* (pp. 33–43). New York, NY: Oxford University Press, Inc.

Warrington, E. K., & Ackroyd, C. (1975). The effect of orienting tasks on recognition memory. *Memory & Cognition*, 3(2), 140-142.

Walsh, D. A., & Jenkins, J. J. (1973). Effects of Orienting Tasks on Free Recall in Incidental Learning: “Difficulty,” “Effort,” and “Process” Explanations. *Journal of Verbal Learning and Verbal Behavior*, 12, 481-488.

Wheeler, M. A., & Roediger, H. L. III (1992). Disparate effects of repeated testing: Reconciling Ballard’s (1913) and Bartlett’s Results. *Psychological Science*, 3(4), 240-245.