

Introduzione

2. Sensazione

2.1. I sensi come strumenti di misura

2.2. Legge di Weber

2.2.1. Misurare una soglia differenziale

PES

JND

2.2.2. La JND è proporzionale allo stimolo di riferimento

2.2.3. Generalità e limiti della legge di Weber

2.3. Legge di Fechner

2.3.1. La funzione logaritmica

2.3.2. Un'unità di misura fechneriana: il decibel

2.4. Legge di Stevens

2.4.1. *Magnitude estimation*

2.4.2. *Magnitude production*

2.4.3. *Cross-modal matching*

2.4.4. La funzione potenza

2.5. Psicofisica esterna e interna

2.5.1. Valore adattivo della sensibilità ai rapporti

2.5.2. Il caso dell'acuità

2.6. Sensibilità alla frequenza

2.6.1. Sensibilità acustica alla frequenza temporale

2.6.2. Sensibilità visiva alla luce

2.6.3. Sensibilità visiva al frequenza spaziale

2.7. Adattamento sensoriale e codifica neurale

2.8. Epilogo

3. Processi percettivi di base

3.1. Indeterminazione ottica e percezione visiva

- 3.1.1. Indeterminazione geometrica e fotometrica
- 3.1.2. La catena psicofisica
- 3.1.3. Energia luminosa e informazione ottica
- 3.1.4. Codificazione e organizzazione

3.2. Unificazione/segregazione

3.3. Articolazione figura/sfondo

3.4. Completamento amodale e organizzazione figurale

3.5. Struttura dello spazio visivo

- 3.5.1. Assi cardinali del campo visivo
- 3.5.2. Informazioni sulla profondità

Informazione ottica monoculare

Informazione ottica binoculare

Informazione extra-ottica monoculare

Informazione extra-ottica binoculare

3.6. Epilogo

4. Riconoscimento

4.1. Un concetto eterogeneo

- 4.1.1. "Vedere come"
- 4.1.2. Ritrovare il già noto
- 4.1.3. Illusioni di riconoscimento

4.2. Modelli di riconoscimento

- 4.3.1. Confronto di sagome (*template matching*)
- 4.3.2. Reti di caratteristiche (*feature nets*)
- 4.3.3. Descrizioni strutturali (*structural descriptions*)
- 4.3.4. Quale modello per il riconoscimento umano?

4.3. Effetto inversione

4.4. Percezione categoriale

4.5. Epilogo

Introduzione ai capp. 2, 3, 4

I capitoli su sensazione, percezione e riconoscimento riguardano l'*esperienza diretta*, un oggetto d'indagine non del tutto ovvio. Quando ci si chiede "cosa fa la mente?", vien da pensare in modo abbastanza naturale ai casi in cui "mente" compare nel linguaggio quotidiano, in espressioni come "venire in mente", "tenere a mente" oppure "calcolo mentale". Capacità come ricordare o eseguire operazioni aritmetiche senza supporti esterni appaiono come chiare manifestazioni del buon funzionamento di uno speciale organo, il sistema mente/cervello.

In genere è meno chiaro in che senso l'esperienza diretta degli osservatori – il mondo come loro appare – sia un prodotto mentale. Intendiamoci sui termini. Quando parliamo di "esperienza" pensiamo di solito a ciò che abbiamo imparato, magari casualmente, nel corso della vita; c'è chi ne ha tanta e chi poca. Nel linguaggio scientifico, invece, esperienza diretta è tutto ciò che l'osservatore vive come un dato, nel momento considerato: le cose viste e toccate, con il loro colore e la loro durezza; i vari oggetti del mondo, incluso quell'oggetto particolare che è il corpo dell'osservatore. Perché, quindi, i problemi scientifici relativi all'esperienza diretta sono spesso difficili da individuare? Molto dipende dal fatto che essa si presenta per quello che non è; cioè, dal fatto che il mondo percepito è *fenomenicamente oggettivo* [Riquadro A.1]. In chiave evolucionistica, l'oggettività fenomenica dell'esperienza diretta viene a volte chiamata *illusione di realtà*, allo scopo di evidenziare che si tratta di una simulazione *geneticamente soggettiva* (cioè prodotta dall'organismo) del mondo fisico, tendenzialmente adeguata alle necessità poste dall'interazione con l'ambiente e dalla comunicazione con i conspecifici, ma non del tutto veridica.

La contrapposizione critica tra mondo fisico e mondo percepito (inteso come sinonimo di "direttamente esperito") è la strategia con cui cercheremo di superare il *realismo ingenuo*, la convinzione spontanea ma inadeguata che di mondo ce ne sia uno solo, quello in cui le cose sono sempre come appaiono. Tale contrapposizione è una componente fondamentale del *realismo critico*, rappresentato emblematicamente dalla rivoluzione galileiana [Riquadro A.2]. Vedremo che rispetto alle altre scienze, che hanno per oggetto la realtà indipendente dall'osservatore, la psicologia è emersa relativamente tardi, come conseguenza delle difficoltà concettuali inerenti l'applicazione del metodo scientifico allo studio dell'osservatore [Riquadro A.3].

I prossimi tre capitoli seguono una tripartizione classica, non esente da difetti, del flusso di elaborazione dell'informazione contenuta negli stimoli sensoriali. Per

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

esempio, nel caso della visione il percorso dal mondo fisico all'osservatore è rappresentabile come una serie di condizioni causali: una *catena psicofisica* composta da stimolo distale (indipendente dall'osservatore), stimolo prossimale (l'immagine catturabile in un punto di vista), sensazione (la registrazione, tipicamente parziale, degli stimoli prossimali), percezione (prodotto dell'organizzazione delle sensazioni in unità normalmente corrispondenti agli stimoli distali), riconoscimento (risultante dal confronto tra le percezioni e l'informazione depositata in memoria). In un modello rigidamente sequenziale, ciascuno stadio di elaborazione sarebbe condizione necessaria e sufficiente per il successivo. Benché i confini tra sensazione, percezione e riconoscimento siano controversi, allo scopo di illustrare almeno i fatti e i problemi più rilevanti abbiamo seguito questa tripartizione che, all'interno della modalità sensoriale più studiata, ricalca la consueta distinzione tra visione di basso, intermedio, alto livello (*low-, mid-, high-level vision*).

Il bello della diretta – dell'esperienza diretta – è che i dati osservati non hanno etichetta d'origine. Quando incontriamo un amico, il colore della sua nuova maglietta (dipendente quasi soltanto dalla codificazione neurale della luce che colpisce l'occhio, un processo di basso livello), la sua andatura (dipendente dall'organizzazione percettiva delle traiettorie tracciate dai vari punti del suo corpo, un processo di livello intermedio), il senso di calda familiarità che il suo volto suscita in noi (dipendente dalla nostra storia individuale, incorporata in processi ad alto livello) appaiono tutte come proprietà fenomenicamente immediate. E non è facile capire se l'esperienza diretta risulti soltanto da processi *bottom-up* (stimolo prossimale \Rightarrow sensazione \Rightarrow percezione \Rightarrow riconoscimento) o anche da processi *top-down* (in cui gli schemi che guidano il riconoscimento co-determinano l'elaborazione degli stimoli).

In sostanza, facendo proprio l'aforisma “Non vediamo le cose come sono, ma come siamo”, coniato dalla scrittrice americana Anaïs Nin (1903-1977), l'indagine scientifica si propone di identificare i processi interni al sistema mente/cervello che portano le cose ad apparire così come appaiono [Riquadro A.4].

Riquadro A.1

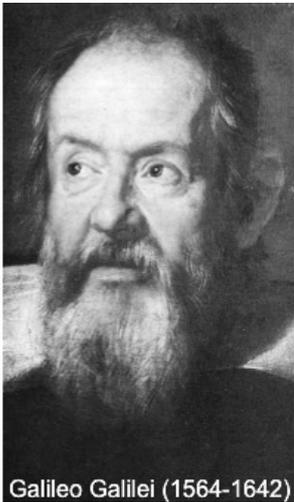


Wolfgang Köhler (1887-1967)

“Ma come posso dire che una sedia, per esempio, è un’esperienza oggettiva, se debbo ammettere che essa dipende da certi processi interni al mio organismo? Su questa base la sedia non diventa soggettiva? Sì e no. In questo preciso momento abbiamo cambiato il significato dei termini soggettivo e oggettivo” (W. Köhler, *Gestalt Psychology*, 1929, p. 24, New York, Liveright).

Il passo è tratto dalle pagine in cui lo psicologo tedesco Wolfgang Köhler introduce la fondamentale distinzione tra *fenomenicamente oggettivo* e *geneticamente soggettivo*: l’esperienza diretta (vedere una sedia davanti a noi e sentirla sotto le nostre mani) si presenta come oggettiva, nel senso che non viene vissuta come prodotta dall’atto osservativo, pur essendo generata da processi interni all’osservatore.

Riquadro A.2

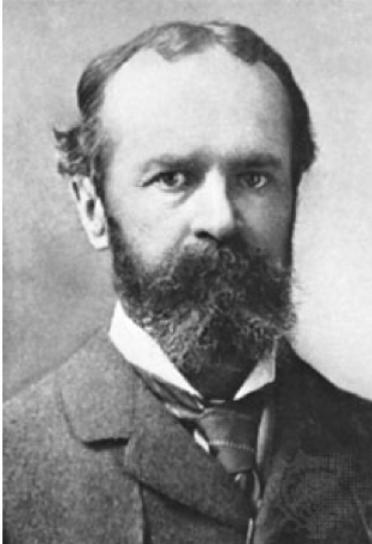


Galileo Galilei (1564-1642)

“Io vo movendo una mano ora sopra una statua di marmo, ora sopra un uomo vivo. Quanto all’azione che vien dalla mano, rispetto ad essa mano è la medesima sopra l’uno e l’altro soggetto, ch’è di quei primi accidenti, cioè moto e toccamento, né per altri nomi vien da noi chiamata [...] Or quella titillazione è tutta di noi, e non della penna, e rimosso il corpo animato e sensitivo, ella non è più altro che un puro nome. Ora, di simile e non maggiore essistenza credo io che possano esser molte qualità che vengono attribuite a i corpi naturali, come sapori, odori, colori ed altre.” (Galileo Galilei, *Il Saggiatore*, 1623).

Il ragionamento è semplice e geniale. Nessuno dubita che il solletico dipenda interamente dalla sensibilità di chi lo prova, pur essendo provocato da un’azione meccanica esterna. Allo stesso modo possiamo ipotizzare (e molta della scienza successiva a Galileo ha cercato di confermarlo per le varie modalità sensoriali) che molte altre proprietà, ingenuamente vissute come oggettive, siano piuttosto gli effetti che particolari stimoli esterni producono nel corpo dell’osservatore. È notevole che Galileo paragoni al solletico non solo sapori e odori – vissuti spesso come caratteristiche della sensibilità (“la bocca amara”) oltre che come attributi degli oggetti esterni (“lo zucchero è dolce”) – ma anche il colore, che osservatori pur non tanto ingenui irresistibilmente trattano come una proprietà degli oggetti materiali.

Riquadro A.3



William James (1842-1910)

“In tutto il libro ho mantenuto il punto di vista delle scienze naturali [...] Ogni scienza naturale assume alcuni dati in modo acritico [...] La psicologia, la scienza delle menti individuali e dei loro limiti, assume come propri dati (1) i *pensieri e sentimenti* e (2) un *mondo fisico*, distribuito nello spazio-tempo, con il quale tali menti coesistono e che (3) *esse conoscono*.” (W. James, *The Principles of Psychology*, 1890).

Difficile trovare un passo più denso di idee chiave. Nella prefazione al suo celebre trattato, lo psicologo e filosofo americano William James definisce in modo insuperato la psicologia come scienza delle menti naturali (limitate e potenzialmente diverse da individuo a individuo), esplicita i dati che questa ambiziosa disciplina deve prendere per buoni e qualifica la relazione che lega i contenuti mentali (tra i quali le percezioni) al mondo fisico con il termine “conoscenza”.

Le parole di James riecheggiano nell'incipit di *Gestalt Psychology* (Köhler, 1929): “Per la psicologia proprio come per tutte le altre scienze sembra esserci un unico punto di partenza: il mondo come me lo trovo davanti, ingenuamente e acriticamente”.

Riquadro A.4



Kurt Koffka (1886-1941)

Why do things look as they do? “Perché le cose appaiono come appaiono?” Questa domanda cruciale, che tutti considerano come la più importante nella storia della scienza della percezione, venne posta dallo psicologo tedesco Kurt Koffka nei *Principles of Gestalt Psychology* (1935).

Koffka esclude come insufficienti due possibili risposte: le cose appaiono come appaiono perché (1) sono quello che sono, nel senso che corrispondono agli stimoli distali (la risposta del realismo ingenuo); (2) corrispondono agli stimoli prossimali (che essendo a contatto con l'osservatore rispettano un requisito della causazione fisica). La vera risposta, per Koffka, è che le cose appaiono così grazie all'organizzazione che il sistema mente/cervello impone agli stimoli prossimali, concepiti come vincoli esterni di un processo comunque regolato da tendenze interne. L'esperienza diretta risulterebbe quindi dall'equilibrio tra forze esterne e interne.

Si tenga presente che il “perché” di Koffka riguarda solo il primo dei quattro livelli in cui l'etologo olandese Niko Tinbergen (1907-1988) propose di articolare la spiegazione scientifica del comportamento degli organismi: causale, adattivo, ontogenetico, filogenetico. Anche per l'esperienza diretta (accessibile all'individuo ma preclusa agli altri osservatori) possiamo chiederci: (a) quali processi/meccanismi ne sono la causa prossima; (b) qual è la sua funzione; (c) come si sviluppa in un dato organismo; (d) qual è la sua storia evolutiva, in una specie e tra le specie.

2. Sensazione

2.1. I sensi come strumenti di misura

In condizioni semplificate, quando lo stimolo varia lungo una sola dimensione, i sistemi sensoriali funzionano in pratica come strumenti di misura delle grandezze fisiche fondamentali (peso, lunghezza, durata). Certo, essi sono qualcosa di più e di diverso. Per esempio il sistema visivo, con cui si può stimare “a occhio” quanto lungo è un tavolo, non si è evoluto per far questo e viene facilmente superato da una stecca di legno graduata.

Benché il modello dell'organo di senso come strumento di misura sia parziale, l'approccio scientifico centrato sull'isolamento delle variabili e sulla spiegazione dei sistemi complessi in base alle proprietà dei loro costituenti elementari ha giocato un ruolo importante anche in psicologia, in particolare attraverso la psicofisica, disciplina nata all'inizio dell'Ottocento e ancora in grado di fornire una solida impalcatura metodologica per lo studio della mente. La psicofisica ha affrontato problemi cruciali come la scelta dell'unità di misura della sensazione e il rapporto tra misure fisiche e misure psicologiche.

2.2. Legge di Weber

La prima pietra dell'edificio della psicofisica venne posta da Ernst H. Weber [Riquadro 2.1], il quale trovò una risposta elegante a una domanda molto ben posta: quali sono le somiglianze e le differenze tra le varie modalità sensoriali? Weber utilizzò la *discriminazione di intensità*, un compito adatto a tutte le modalità – considerato che sensazioni di qualità anche molto diversa (pesantezza, chiarezza, altezza) variano comunque per intensità – e che fornisce una buona definizione operativa di *sensibilità*: un osservatore è molto sensibile se discrimina intensità molto vicine.

Nella tradizione aristotelica i sensi periferici (visione, udito, tatto, gusto, olfatto) confluiscono nel *senso comune*, una modalità sovraordinata che mette in relazione le informazioni veicolate dai primi. La capacità di discriminare l'intensità di due sensazioni – avvertendo una più forte dell'altra – potrebbe dipendere dal funzionamento di ciascun senso periferico o da una modalità centrale sovraordinata.

Ipotizziamo dunque che i sistemi sensoriali siano strumenti di misura delle varie grandezze fisiche: la visione per le grandezze fotometriche (intensità, composizione spettrale) e geometrico-cinematiche (lunghezza, densità, velocità); l'udito per le variazioni di ampiezza e frequenza degli stimoli acustici; il tatto attivo, che coinvolge

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 pelle, postura e muscoli, per il peso sollevato; il tatto passivo per la pressione esercitata sulla pelle; il gusto e l'olfatto per la concentrazione di determinate sostanze. In ciascun caso possiamo determinare una soglia differenziale, cioè la minima differenza tra due stimoli sufficiente per la loro discriminazione. In un compito in cui l'osservatore esegue molte prove in cui deve confrontare uno stimolo di valore fisso (standard) con uno stimolo variabile (confronto), la soglia differenziale è l'incremento che innalza la percentuale di risposte "confronto > standard" dal 50 all'75%, cioè dal livello del caso (assenza di discriminazione) a un valore convenzionalmente adottato come indicativo di una discriminazione sufficientemente attendibile.

2.2.1. Misurare una soglia differenziale

Ecco come possiamo procedere per misurare la soglia differenziale per il peso sollevato. Poiché il volume e l'aspetto materiale possono influenzare le aspettative dell'osservatore e quindi la percezione del peso, utilizzeremo oggetti esternamente uguali: per esempio, riempiamo d'acqua delle bottigliette di plastica opaca (per evitare che il livello di riempimento influenzi il giudizio) e le peseremo con una bilancia. Faremo in modo che la bottiglietta standard pesi 100 g e che le 8 bottigliette di confronto pesino rispettivamente 93, 95, 97, 99, 101, 103, 105, 107 g. Non serve che uno stimolo di confronto abbia peso uguale allo standard; anzi, se tutti sono diversi dallo standard lo sperimentatore può in tutta onestà garantire all'osservatore che in ogni confronto le due bottigliette hanno peso diverso, anche se gli sembrano di peso uguale. In ogni prova il partecipante solleva in sequenza lo standard e uno stimolo di confronto (o viceversa), e indica quello che gli sembra più pesante.

Questa procedura viene chiamata *metodo degli stimoli costanti* in quanto lo sperimentatore, invece di variare in modo continuo lo stimolo di confronto come nel *metodo dell'aggiustamento* [descritto nel Riquadro 2.2], sceglie da 5 a 9 stimoli di confronto, per poi presentarli uno alla volta insieme allo standard, chiedendo all'osservatore di indicare quello di intensità maggiore.

La **Figura 2.1** mostra la distribuzione delle proporzioni di risposte "confronto > standard", nell'ipotesi che all'osservatore sia stata presentata una sequenza casuale di 160 prove, prodotto di 10 ripetizioni \times 2 ordini di sollevamento (standard \Rightarrow confronto; confronto \Rightarrow standard) \times 8 stimoli di confronto. Nel grafico non vanno rappresentate le proporzioni di risposte corrette ma – cosa poco intuitiva – le proporzioni di risposte di un tipo, tra le due alternative "confronto > standard" e "confronto < standard". Convenzionalmente si sceglie la prima perché gli stimoli di confronto in ascissa sono in

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 ordine crescente da sinistra a destra, e quindi la curva interpolante sarà monotona crescente. La distribuzione di 8 punti tende a disporsi lungo una tipica curva a S, chiamata ogiva o *curva psicometrica*, essendo caratteristica di molte misurazioni psicologiche. La curva è interpolabile a occhio oppure utilizzando una procedura più elaborata [Riquadro 2.3], ma l'importante è l'obiettivo: la misurazione del *punto di eguaglianza soggettivo* (PES) e della JND.

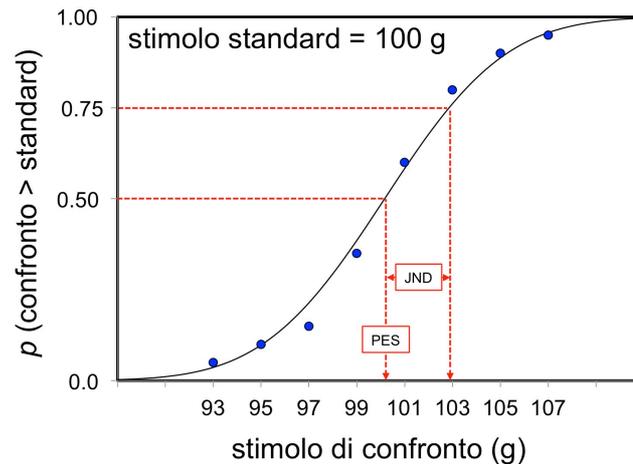


Figura 2.1. La *curva psicometrica* interpola le 8 proporzioni di risposta “confronto > standard” ottenute con il *metodo degli stimoli costanti*. Il PES è il valore in ascissa corrispondente all'intersezione tra la curva psicometrica e il livello $p(\text{confronto} > \text{standard}) = 0,50$. La JND è l'incremento di intensità dello stimolo di confronto necessario a elevare $p(\text{confronto} > \text{standard})$ da 0,50 a 0,75.

PES. Seguite la linea tratteggiata orizzontale per il livello $p(\text{confronto} > \text{standard}) = 0,50$ fino a toccare la curva psicometrica della Figura 2.1 e scendete in verticale fino all'ascissa individuando il PES, pari a 100,16 g. Questo è il valore dello stimolo di confronto che rende equiprobabili le risposte “confronto > standard” e “standard > confronto”. La differenza [PES - standard] è il cosiddetto *errore costante* (qui pari a 0,16 g), che dovrebbe essere molto piccolo se l'ordine di presentazione è bilanciato.

JND. L'altro valore significativo è riferito al livello convenzionale $p(\text{confronto} > \text{standard}) = 0,75$. Confrontare i valori assoluti delle JND provenienti da ricerche differenti richiede cautela, in quanto sono utilizzati anche altri livelli (per esempio, lo 0,84 derivato dalla curva normale, di cui la psicometrica è la cumulativa), tutti sensati purché superiori al caso, cioè a 0,50. Nel nostro esempio, la soglia differenziale è l'incremento di peso, rispetto al PES, necessario a portare la probabilità di risposta “confronto > standard” da 0,50 a 0,75, che per la curva psicometrica della Figura 2.1, relativa a uno standard di 100 g, corrisponde a una JND di 2,69 g.

2.2.2. La JND è proporzionale all'intensità di riferimento

Si potrebbe ipotizzare che il sistema sensoriale di discriminazione del peso funzioni come una buona bilancia elettronica, in grado di registrare con uguale accuratezza sia la differenza tra 100 e 110 g sia la differenza tra 500 e 510 g. Weber trovò qualcosa di molto diverso. Ripetendo la misurazione della soglia differenziale con uno standard di 500 g, Weber scoprì che l'incremento di circa 3 g, che nell'80% dei casi consente di dire correttamente che una bottiglietta di 103 g pesa più di quella di 100 g, non è assolutamente sufficiente a far discriminare una bottiglia di 503 g da una di 500 g. Lo stesso livello di discriminazione da 500 g si ottiene quando lo stimolo di confronto pesa 515 g.

Weber trasse la conclusione che, con buona approssimazione, l'incremento ΔI – lungi dall'essere costante – cresce al crescere dell'intensità I dello standard, mantenendo con essa una relazione di semplice proporzionalità, tale per cui a rimanere costante è il rapporto $\Delta I / I$, chiamato *frazione di Weber*. Questa relazione vale in tutte le modalità sensoriali, anche se il valore della frazione cambia.

Tabella 2.1. Valori tipici di k (la costante di Weber) e c (il coefficiente della funzione di Fechner definita nella sezione 2.3.1) per varie modalità sensoriali (vista, udito, ecc.) e per proprietà intramodali differenti (intensità e altezza del suono). Siccome c è una funzione inversa di k , la sensibilità è elevata se k è piccolo e c grande, bassa se k è grande e c piccolo).

modalità/proprietà	k	$c=1/\log(1-k)$
intensità di un suono	0,003	768,7
chiarezza di una luce	0,02	116,3
peso sollevato	0,03	77,9
odore	0,07	34,0
intensità della pressione sulla pelle	0,14	17,6
altezza di un suono	0,15	16,5
sapore (concentrazione salina)	0,20	12,6

2.2.3. Generalità e limiti della legge di Weber

La **Tabella 2.1** evidenzia come la capacità discriminativa, misurata in base alla costante di Weber, vari molto a seconda della modalità/proprietà considerata. La sensibilità è definita come l'inverso della soglia (sensibilità = $1/\text{soglia}$). Quindi, se il valore di k è piccolo, la sensibilità è elevata (un piccolo incremento produce una nuova sensazione); se il valore di k è grande, la sensibilità è scarsa (il passaggio a una nuova sensazione richiede un incremento maggiore). I valori dipendono anche dall'osservatore e dalle condizioni di misura, ma veicolano un'idea sostanzialmente corretta: nonostante l'intensità sia un attributo comune a tutte le modalità sensoriali, la sensibilità riflette caratteristiche specifiche della modalità o modulo intrasensoriale (per esempio, chiarezza e lunghezza visive), piuttosto che il funzionamento di un apparato sovramodale ispirato al senso comune aristotelico.

In verità, la legge di Weber è un'idealizzazione. Per esempio, la sensibilità umana alla luce copre una gamma amplissima di intensità, ai cui estremi il valore della frazione di Weber è più elevato (sensibilità minore). La Figura 2.2 mostra l'andamento a U del valore della frazione di Weber al crescere della luminanza (la quantità di luce che raggiunge l'occhio da una zona unitaria del campo visivo, misurata in cd/m^2 , candele al metro quadrato). Le luminanze in ascissa, estese negli ambiti *mesopico* (visione al crepuscolo) e *fotopico* (visione diurna), coprono 5 unità logaritmiche, entro le quali la frazione $\Delta I / I$ vale circa 0,02 (valore indicativo riportato nella **Tabella 2.1**) solo per le luminanze centrali. In ambito *scotopico* (visione notturna, per intensità inferiori a $0,01 \text{ cd}/\text{m}^2$) la frazione di Weber cresce ancora.

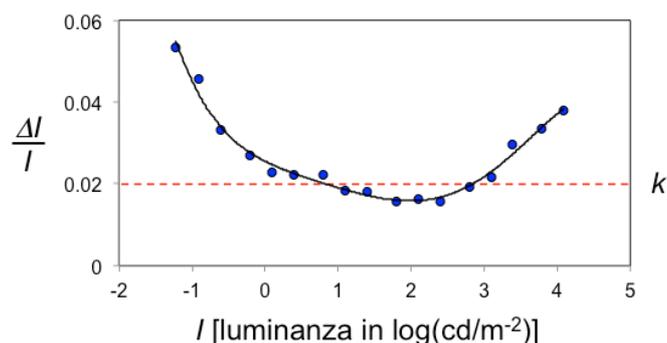


Figura 2.2. Nel caso della sensibilità alla luce, se si rappresenta il valore la frazione di Weber (in ordinata) in funzione dell'intensità I dello stimolo di riferimento (in ascissa), si ottiene un andamento a U. La legge di Weber, rappresentata dalla linea tratteggiata orizzontale, è rispettata nell'intervallo centrale della gamma di intensità. I 6 punti a sinistra riguardano la visione *mesopica* (da circa $0,01$ a $3 \text{ cd}/\text{m}^2$); gli altri la visione *fotopica* (occhio adattato alla luce).

2.3. Legge di Fechner

L'espressione *legge di Weber* fu introdotta dal fisico, filosofo e psicologo tedesco Gustav Theodor Fechner (1801-1887), che riconobbe nella proporzionalità delle JND un fondamentale invariante sensoriale, utile a rispondere a una domanda complementare a quella di Weber. Come cresce la sensazione al crescere dello stimolo? Cioè, dati tre stimoli fisici ben distinguibili ed equispaziati sul continuum dell'intensità (per esempio, pari a 200, 300, 400), la sensazione prodotta dallo stimolo 300 starà esattamente a metà tra quella prodotta dallo stimolo 200 e quella prodotta dallo stimolo 400?

La domanda riguarda uno specifico aspetto del *parallelismo psicofisico*, posizione teorica secondo la quale il mondo fisico e il mondo fenomenico sono diversi, ma paragonabili a due universi in cui ogni stimolo corrisponde a un'unica sensazione, e viceversa. Assumendo che i due mondi siano paralleli, Fechner si chiese se la loro metrica è uguale in tutto, o se il mondo fenomenico – sotto il profilo dei rapporti tra intensità – non sia piuttosto paragonabile a una regolare deformazione del mondo fisico.

2.3.1. La funzione logaritmica

La soluzione al parallelismo psicofisico proposta da Fechner è una legge ricavata matematicamente combinando alcuni assunti con l'ipotesi che la legge di Weber sia ferrea, cioè valga su tutto il continuum delle intensità (diversamente da quanto mostrato nella [Figura 2.2](#)). La formulazione più semplice della legge psicofisica di Fechner è:

$$S = c \log(I);$$

dove la variabile indipendente è I , intensità dello stimolo espressa come multiplo della soglia assoluta (quindi come numero puro, e non come grandezza fisica), la variabile dipendente è S , intensità della sensazione, mentre c è una costante – caratteristica della modalità sensoriale – pari a $[1/\log(1-k)]$, dove k è la frazione di Weber.

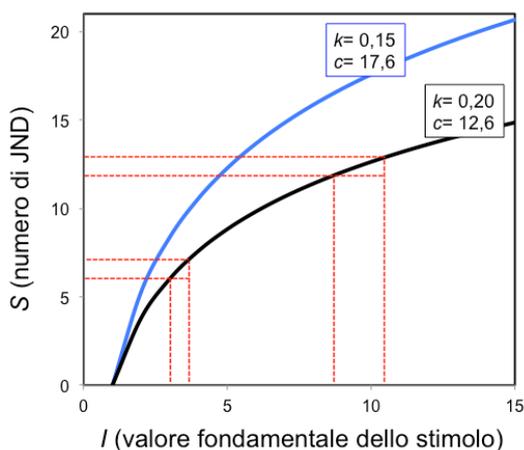


Figura 2.3. Rappresentazione grafica della legge psicofisica $S = c \log(I)$, per due diversi valori di c (e quindi di k , di cui c è una funzione inversa). Per la curva con $c=12,6$ le linee tratteggiate evidenziano che intervalli uguali sulla scala delle sensazioni S corrispondono a intervalli proporzionali sulla scala delle intensità fisiche I .

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

La **Figura 2.3** contiene la rappresentazione della legge logaritmica di Fechner per due valori di c . Per Fechner, una modalità è molto sensibile se due intensità-stimolo sono divise da molte JND. Il valore di c è rivelatore: poiché $\log(10) = 1$, tra la soglia assoluta e uno stimolo 10 volte più intenso ci sono circa 116 JND per la chiarezza della luce, 78 per il peso, solo 13 per il sapore di sale (Tabella 2.1).

Il primo assunto di Fechner è intuitivo: esiste una soglia assoluta, sotto la quale lo stimolo fisico esiste ma è così debole da risultare privo di corrispettivo psicologico nella sensazione. L'esistenza di una soglia assoluta implica che il parallelismo psicofisico si attui tra due semirette con origini sfasate (Figura 2.4).

Il secondo assunto – molto criticato ma cruciale per l'eleganza del modello – è che tutte le JND siano soggettivamente uguali: in una modalità con $k = 0,20$ il vissuto di differenza tra 10 e 12 sarebbe eguale al vissuto di differenza tra 70 e 84, tra 200 e 240, e così via (aspetto evidenziato nella **Figura 2.4** dalla corrispondenza tra passi eguali sul continuum delle sensazioni e passi proporzionalmente crescenti sul continuum degli stimoli fisici).

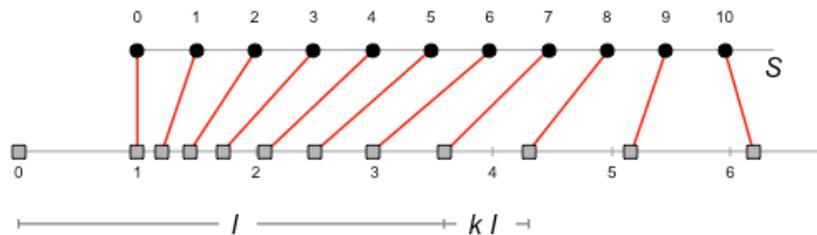


Figura 2.4. Il *parallelismo psicofisico* postulato da Fechner è caratterizzato da due proprietà: (i) l'origine del continuum psicologico S è spostata verso destra, rispetto all'origine del continuum fisico I ; (ii) intervalli eguali sul continuum psicologico S corrispondono a intervalli diseguali ma proporzionalmente crescenti sul continuum fisico I . Come evidenziato dal regolo in basso, la rappresentazione si riferisce al caso in cui $k = 0,20$ ($c = 12,6$).

2.3.2. Un'unità di misura fechneriana: il decibel

Più che alla sua accuratezza nel rappresentare l'accrescimento soggettivo delle sensazioni sopra soglia, la legge di Fechner deve la propria importanza all'eleganza formale, al punto da essere incorporata nella definizione di unità di misura come il *decibel*. In ambito acustico, il dB_{SPL} (*decibel* per il *Sound Pressure Level*) non esprime una misura assoluta dell'intensità dello stimolo (l'analogo del peso in grammi), ma piuttosto una quantità approssimativamente equivalente al numero di JND di un osservatore ideale la cui soglia assoluta p_0 sia pari a $0,00002$ pascal (2×10^{-5} Pa, una piccolissima variazione di pressione, pari a quella provocata da una zanzara lontana 3 m,

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 in una stanza ben isolata). La definizione completa, conforme alla legge di Fechner, è $\text{dB}_{\text{SPL}} = 20 \log(p/p_0)$. Il valore della costante c (pari a 20) è stato scelto in modo da far corrispondere l'unità dB_{SPL} circa a una JND.

Pertanto, un suono quantificato in 20 dB_{SPL} (un debole fruscio) ha intensità fisica pari a 2×10^{-4} Pa (10 volte la soglia assoluta) e uno di 80 dB_{SPL} (rumore di una strada molto trafficata a 30 m) ha intensità fisica pari a 0,2 Pa (10.000 volte la soglia assoluta). La soglia del dolore viene situata intorno a 140 dB_{SPL} , pari a 2×10^2 Pa (10^7 volte la soglia assoluta). Accettando la logica di Fechner, possiamo concludere che i suoni udibili, da quello appena avvertibile a quello che provoca dolore (e al quale non dovete esporvi), coprono un'estensione pari a 140 JND, tutte – ricordiamolo ancora una volta – corrispondenti a intervalli fisici proporzionalmente crescenti.

2.4. Legge di Stevens

Nella *psicofisica indiretta* di Fechner la funzione psicofisica ad accrescimento logaritmico è un'inferenza basata sulle misure della soglia assoluta e di più soglie differenziali lungo il continuum di intensità dello stimolo. Un'alternativa radicale è la *psicofisica diretta* di S. S. Stevens (1906-1973), che propose di ottenere la funzione psicofisica utilizzando – senza mediazioni – i giudizi quantitativi espressi dagli osservatori. I tre metodi codificati da Stevens sono i seguenti.

2.4.1. *Magnitude estimation*

Nella *magnitude estimation* (stima di grandezza) l'osservatore attribuisce un numero a ciascuna intensità sopra soglia. Per facilitare il confronto tra le stime di osservatori diversi, lo sperimentatore può proporre a tutti di ancorarsi a uno stesso *modulo*, suggerendo per esempio di chiamare 10 una certa intensità di riferimento. Provate. Ritagliate 7 dischi di varia grandezza. Lasciate il disco piccolo sempre a disposizione dell'osservatore e stabilite che la sua area vale 10. Poi mostrate, uno dopo l'altro, gli altri dischi, chiedendo all'osservatore di esprimere con un numero la grandezza percepita. Ripetete la procedura alcune volte, in modo da poter calcolare un valore medio della stima numerica di ciascun disco.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

2.4.2. *Magnitude production*

La *magnitude production* (produzione di grandezza) è la procedura inversa rispetto alla precedente. Dato un numero assegnato dallo sperimentatore, l'osservatore deve aggiustare (o scegliere tra quelli dati) il valore dello stimolo che meglio lo rappresenta. Poniamo che vi venga mostrato un disco di grandezza 10 e che abbiate la possibilità di regolare la grandezza di un altro disco (per esempio proiettato sul muro). Se lo sperimentatore vi chiede di produrre 30, voi cercherete di generare un disco che appare tre volte più grande del disco proposto come modulo. Una variante della *magnitude production* è la bisezione, in cui l'osservatore deve generare lo stimolo che gli appare intermedio tra i due stimoli dati.

2.4.3. *Cross-modal matching*

Nel *cross-modal matching* (confronto intermodale) l'osservatore regola l'intensità di una sensazione in una modalità – per esempio, un suono più o meno forte – rendendola uguale all'intensità dello stimolo presentato dallo sperimentatore. La *magnitude estimation* può essere considerata una sorta di *cross-modal matching*, se si ipotizza che la rappresentazione interna dei numeri (cap. XX Zorzi sui numeri) costituisca una modalità autonoma, interfacciabile con le scale di intensità che caratterizzano le diverse modalità sensoriali.

2.4.4. La funzione potenza

I metodi della psicofisica diretta – inclusa la bisezione, utilizzata per costruire una scala dei grigi a passi soggettivamente eguali da due pionieri della psicofisica, Joseph Plateau (1801-1883) e Joseph Delboeuf (1831-1896) – producono dati interpolabili mediante una funzione di accrescimento della sensazione diversa da quella di Fechner, la funzione potenza rappresentata dalla seguente formula generica:

$$S = m I^n;$$

dove S rappresenta l'intensità della sensazione direttamente stimata (o prodotta) dall'osservatore e I l'intensità dello stimolo valutato, mentre m è un coefficiente che dipende dalle specifiche unità di misura utilizzate e n è l'esponente caratteristico della specifica modalità o proprietà considerata (Tabella 2.2).

Come i valori di k e c della psicofisica indiretta, anche il valore dell'esponente n della psicofisica diretta varia molto da modalità a modalità. Ma c'è un'importante differenza: mentre la funzione logaritmica di Fechner è sempre di tipo compressivo (posto che incrementi proporzionali del valore fisico della JND corrispondono a

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 intervalli soggettivi sempre uguali), la funzione potenza genera una famiglia che include curve sia compressive (per $n < 1$) sia espansive (per $n > 1$), oltre alla singolarità (per $n = 1$) rappresentata dalla retta $S = m I$.

Tabella 2.2. Valori tipici dell'esponente n della funzione potenza di Stevens, che esprime la velocità con cui cresce la sensazione al crescere dello stimolo.

modalità/proprietà	n
scossa elettrica sulle dita	3,50
esposizione al freddo su tutto il corpo	1,70
calore di un oggetto metallico a contatto sul braccio	1,60
durata di uno stimolo acustico (rumore bianco)	1,10
lunghezza di una linea	1,00
area di un disco	0,70
intensità acustica di una nota pura (3000 Hz)	0,67
chiarezza di una zona illuminata su sfondo buio	0,33

La **Figura 2.5** illustra i tre casi, in cui la sensazione cresce molto più lentamente dello stimolo (chiarezza), come lo stimolo (lunghezza di una linea) e più velocemente dello stimolo (calore sul braccio).

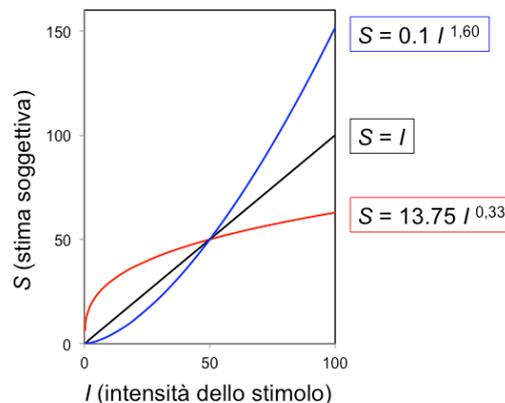


Figura 2.5. La legge psicofisica di Stevens è descritta dalla *funzione potenza* $S = m I^n$, con due parametri liberi: m , il cui valore dipende dalle unità scelte per lo stimolo e per la stima soggettiva, non ha significato teorico; n rende la curva negativamente ($n < 1$) o positivamente ($n > 1$) accelerata, o la fa degenerare in una retta ($n = 1$). Gli esponenti utilizzati sono quelli per la chiarezza di una zona illuminata ($n = 0,33$), per la lunghezza di una linea ($n = 1$) e per il calore sul braccio ($n = 1,60$). I valori di m sono scelti in modo da rendere salienti le differenze fra i tre andamenti.

2.5. Psicofisica esterna e interna

Le leggi di Weber, Fechner e Stevens si applicano tutte all'ambito che lo stesso Fechner aveva definito *psicofisica esterna*, riferita alla relazione tra sensazioni e stimoli collocati al di fuori dell'organismo. Fechner aveva tuttavia anticipato l'esigenza di sviluppare una *psicofisica interna*, riferita alla relazione tra le sensazioni e gli eventi

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 neurali – suscitati dagli stimoli – che vanno considerati come correlati, o condizioni ultime, delle sensazioni.

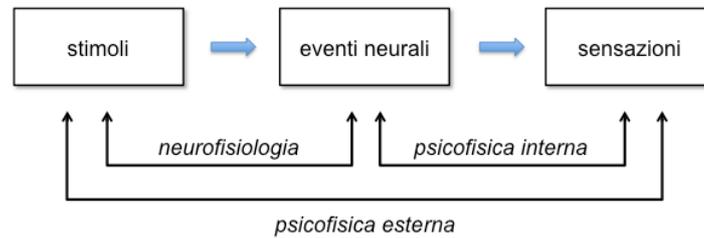


Figura 2.6. Fechner aveva ben chiaro che la *psicofisica esterna*, relativa al legame tra sensazioni e stimoli esterni all'organismo, va interpretata come composizione di due funzioni, quella indagata dalla *neurofisiologia* (che mappa gli stimoli esterni negli eventi neurali) e quella qualificabile come *psicofisica interna* (che mappa gli eventi neurali nelle sensazioni).

La **Figura 2.6** illustra lo schema che, a partire dalla seconda metà del Novecento, guida l'approccio correlazionale allo studio delle funzioni sensoriali, percettive e cognitive, e che ha l'obiettivo di trattare la psicofisica interna sostanzialmente come un'equivalenza o un isomorfismo (per usare il termine preferito da Köhler nell'ambito della teoria della Gestalt). Tale schema permette di formulare un'interpretazione della legge di Fechner. I dati neurofisiologici, infatti, indicano che la risposta dei neuroni, consistente nell'aumento di frequenza degli impulsi rispetto al livello base, è una funzione approssimativamente logaritmica dell'intensità fisica dello stimolo; proprio come l'accrescimento delle sensazioni rivelato dalle risposte comportamentali, motorie o verbali. La corrispondenza fra dati della neurofisiologia e dati della psicofisica esterna obbedisce al criterio dell'accumulazione di prove convergenti cui si ispira tutta la ricerca sui meccanismi sensoriali.

2.5.1. Valore adattivo della sensibilità ai rapporti

L'identificazione del correlato neurale di un'esperienza sensoriale risponde alla prima domanda di Tinbergen, quella relativa alla causa prossima (si veda il Riquadro A.4 e **il cap. xx**); ma non rende esplicito il significato adattivo della proporzionalità delle JND o dell'accrescimento logaritmico della sensazione. Per comprendere il valore adattivo della sensibilità ai rapporti (piuttosto che alle differenze) dobbiamo far riferimento alla distinzione tra *stimolo distale* e *stimolo prossimale*, essenziale per un senso remoto come la visione, in cui la comunicazione fra l'oggetto fisico e l'osservatore è mediata dalle proiezioni ottiche.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

Prendiamo *due oggetti distali* di altezza diversa, uno sgabello di 60 cm e una sedia di 40 cm. A 2 m dall'osservatore quello scarto di 20 cm corrisponde a una *differenza prossimale* pari a circa 1,8 mm sulla retina; ma a 4 m di distanza quello stesso scarto di 20 cm diventa 0,9 mm sulla retina, e così via. Se la sensibilità visiva fosse funzione delle differenze prossimali assolute, sgabello e sedia dovrebbero apparire di grandezza via via più simile quanto più l'osservatore si allontana (poiché la differenza prossimale tra le immagini dei due oggetti diviene sempre più piccola): un assurdo per un organismo mobile come il nostro, che raccoglie continuamente immagini provenienti da un mondo fatto in prevalenza di oggetti con una grandezza distale invariante, ma che – come conseguenza della variabilità della distanza di osservazione – deve elaborare *immagini retiniche* sempre diverse. Ma qualcosa rimane invariante anche sulla retina: sono i rapporti tra le grandezze retiniche, che riproducono fedelmente i rapporti tra le grandezze distali. Il fatto che la sedia sia alta $\frac{2}{3}$ dello sgabello si traduce nel fatto che sempre, indipendentemente dalla distanza di osservazione, l'immagine retinica della sedia sarà $\frac{2}{3}$ di quella dello sgabello, e la differenza tra le due immagini retiniche sarà sempre la metà della grandezza retinica della sedia.

L'invarianza dei rapporti è il possibile fondamento ecologico della legge di Weber, di evidente valore adattivo per un organismo che ha bisogno di muoversi tra gli oggetti, usando le informazioni contenute nelle immagini proiettate sulla retina. Per sviluppare un'interazione efficiente con l'ambiente, l'osservatore deve poter percepire le proprietà dello stimolo distale e non quelle dello stimolo prossimale, che varie circostanze (prima fra tutte, l'insieme dei movimenti esplorativi) rendono molto variabile anche quando lo stimolo distale è costante.

2.5.2. Il caso dell'acuità

Uno strumento che debba rilevare stimoli distribuiti nello spazio (e il nostro sistema visivo, a livello elementare, fa pure questo) è caratterizzato dalla sua *acuità*, cioè dalla capacità di risolvere i dettagli, discriminando un elemento da due elementi separati da un intervallo vuoto. Il nome di Weber è legato anche alla scoperta di una curiosa eterogeneità della pelle del corpo, organo del tatto passivo. Chiudete gli occhi e chiedete a un amico di far scorrere le punte di un compasso sul vostro volto, da un orecchio all'altro, passando intorno alla bocca (una punta sopra e una sotto le labbra). Quando l'apertura effettiva del compasso è costante, sentirete un movimento divergente da un orecchio verso la bocca e poi convergente verso l'altro orecchio. L'effetto, chiamato *illusione di Weber*, è spiegabile in base alla densità dei recettori tattili, più

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 numerosi vicino alla bocca che all'orecchio. L'illusione ha un corrispettivo statico nel diverso valore, in varie parti del corpo, del *minimum separabile* (la più piccola separazione tra due stimolazioni, prima che vengano confuse con una sola). Per esempio, chiudete gli occhi e fatevi toccare il polpastrello del dito indice con un compasso. Sentirete di essere toccati da due punte (e non avrete la sensazione di un'unica punta) anche quando esse distano solo 5 mm una dall'altra. Poi fatevi toccare sulla schiena. Sentirete due punte soltanto se queste distano almeno 50 mm.

Questo modo di misurare l'acuità nel tatto passivo ha un corrispettivo in vari metodi di misura dell'acuità visiva, in condizioni statiche e dinamiche. Nelle migliori condizioni statiche, un osservatore rileva correttamente la separazione tra due piccole luci, nel buio totale, quando queste distano un minuto di grado di angolo visivo (1/60 di grado). Questo limite – già notevole, visto che un grado di angolo visivo corrisponde sulla retina a circa 0,3 mm – diviene ancora più basso se l'acuità visiva viene misurata chiedendo all'osservatore di discriminare l'allineamento di due piccoli segmenti. Questo compito – chiamato *discriminazione di Vernier* (dal nome di un matematico e inventore francese del Seicento) – rivela una *super-acuità* nell'ordine dei 5 secondi di grado di angolo visivo, addirittura superiore a quella ipotizzabile in base alla stimolazione indipendente dei recettori retinici.

Come l'acuità tattile passiva, anche l'acuità visiva è fortemente influenzata dalla zona stimolata. Verso la periferia del campo visivo l'acuità peggiora, come conseguenza di una minore densità dei recettori retinici e delle cellule corticali da questi direttamente attivate. L'incremento di grandezza necessario a compensare la perdita di acuità visiva dovuta all'eccentricità viene chiamato *fattore di ingrandimento corticale* (*cortical magnification factor*; Figura 2.7) e rappresenta un buon esempio di convergenza tra prestazioni in un compito psicofisico (identificazione di lettere) e il substrato neurale della discriminabilità delle immagini.



Figura 2.7. Illustrazione del fattore di ingrandimento corticale. Fissando il centro della figura, ciascuna lettera avrà un'ampiezza corticale circa simile e quindi sarà ugualmente identificabile.

2.6. Sensibilità alla frequenza

In molti degli esempi fatti finora (peso, lunghezza, area), le sensazioni sono omologhe alle corrispondenti grandezze fisiche. Non è sempre così quando prendiamo in considerazione la frequenza di un evento periodico. Sia in campo acustico sia in campo visivo la frequenza (il numero di cicli nell'unità di tempo o spazio fisico) può corrispondere a una qualità sensoriale lontana dal concetto di numerosità relativa.

2.6.1. Sensibilità acustica alla frequenza temporale

Lo stimolo prossimale per il suono è la variazione periodica di pressione di un mezzo come l'aria, provocata da un evento meccanico (il battacchio che urta la campana, il diapason che vibra). Lo stimolo più semplice è l'onda sinusoidale, variabile per ampiezza (distanza verticale tra picco e valle dell'onda, percepita come intensità sonora sul continuum debole-forte) e frequenza (numero di cicli nell'unità di tempo, percepita come altezza sonora sul continuum grave-acuto).

Uno stimolo acustico a bassa frequenza viene percepito come “basso” e uno ad alta frequenza come “alto”. In sostanza, quindi, si potrebbe pensare che un adeguato modello psicofisico dell'altezza sonora sia una coppia di semirette verticali, una per lo stimolo fisico in Hz (Hertz, numero di cicli al secondo) e una per l'altezza percepita, dalla nota più grave al sibilo più acuto. Poiché vale la legge di Fechner, basterebbe ruotare di 90 gradi la **Figura 2.4** e modificarla tenendo presente che la costante di Weber per l'altezza sonora vale circa 0,15 (**Tabella 2.1**).

Tuttavia, il modello lineare trascura un importante aspetto qualitativo, cioè l'equivalenza ciclica dei suoni prodotti da stimoli con frequenza una doppia dell'altra, separati cioè dall'intervallo musicale chiamato *ottava*. Poiché il suono varia sia per intensità (l'altezza) sia per qualità (l'equivalenza tra note uguali ma di altezza diversa), il modello spaziale più adeguato è un'elica tridimensionale: all'aumentare della frequenza il suono sale lungo una scala a chiocciola, crescendo in altezza ma anche occupando una posizione via via diversa sul cerchio delle note musicali. La struttura elicoidale del suono – un universale percettivo condiviso da più specie e con precisi correlati neurali – esemplifica la codificazione sensoriale multidimensionale di stimoli unidimensionali.

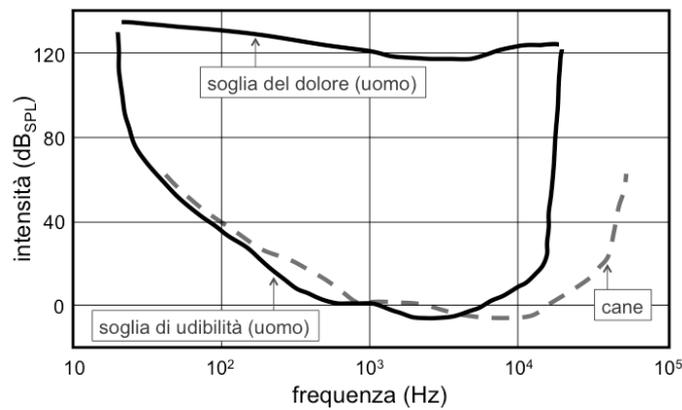


Figura 2.8. Le tre linee descrivono l'andamento della soglia di udibilità di stimoli sinusoidali, in funzione della frequenza in Hz (su ascissa logaritmica), per un osservatore umano normale e per il cane (curve a U, rispettivamente piena e tratteggiata), nonché l'andamento della soglia del dolore nell'osservatore umano, per intensità pari o superiori ai 120 dB_{SPL}. La regione compresa tra le due linee piene rappresenta l'insieme degli stimoli acustici che l'osservatore umano percepisce come suoni. Il cane sente gli ultrasuoni, cioè risponde a stimoli acustici non udibili dall'osservatore umano.

Prescindendo dalla qualità tonale, la sensibilità uditiva in funzione della frequenza è normalmente descritta da grafici come quello della Figura 2.8. Si noti che per l'osservatore umano la soglia in decibel assume un valore negativo per frequenze tra 1000 e 5000 Hz. Ciò deriva dall'aver definito i decibel come misura di intensità relativa alla soglia assoluta, che per il suono è convenzionalmente riferita a uno stimolo acustico di 1000 Hz (cfr. la sezione 2.3.2).

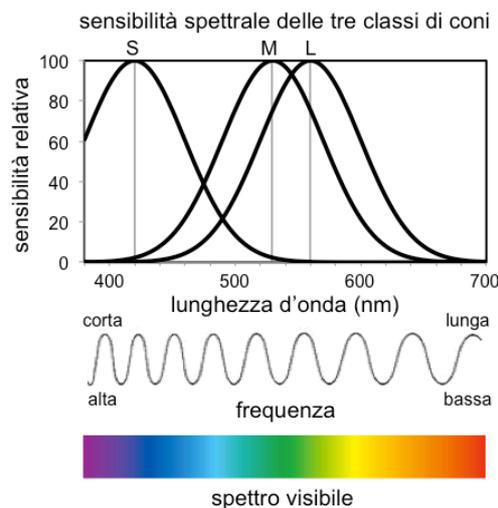


Figura 2.9. Il grafico in alto descrive la sensibilità spettrale delle tre classi di coni, che costituiscono il sistema di codificazione delle miscele di radiazioni. La sensibilità di ciascuna classe è qui approssimata da una curva normale, con picchi rispettivamente a 420 nm (coni *short*, S), 530 nm (coni *medium*, M) e 560 nm (coni *long*, L). In basso viene illustrata la relazione inversa tra lunghezza d'onda e frequenza e la sequenza di colori visibili nella porzione visibile dello spettro, in corrispondenza di ciascuna lunghezza d'onda.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

2.6.2. Sensibilità alla luce

Considerare il sistema visivo come un semplice strumento di misura è utile quando ci si riferisce alla quantità complessiva di luce che arriva in un punto della retina. Ben più complessa (quindi non facilmente riconducibile al modello dello strumento di misura) è la codificazione sensoriale della composizione spettrale della luce, che si traduce nell'esperienza del colore. La **Figura 2.9** illustra la ricchezza dell'informazione contenuta nelle differenze di composizione spettrale tra le varie regioni dell'immagine. Per catturare almeno parte di tale informazione il sistema visivo deve soppesare (per così dire) le componenti principali dello spettro e codificarne la prevalenza relativa.

Un buon modello della codificazione retinica della composizione spettrale è fornito dalla *teoria tricromatica* di Young-Helmholtz [**Riquadro 2.4**], che si basa sul *principio dell'univarianza* della risposta di ciascuna classe di fotorecettori, le cellule della retina in grado di trasdurre l'energia luminosa in segnale nervoso. I fotorecettori retinici sono di due tipi: bastoncelli (specializzati per la visione notturna) e coni (specializzati per la visione diurna). Mentre la sensibilità dei bastoncelli è descritta da un'unica funzione spettrale, nella retina dell'osservatore umano normale troviamo tre classi di coni, con picchi di sensibilità in zone diverse dello spettro di lunghezza d'onda. Le tre classi di coni – indicate con le iniziali S, M, L (*short, medium, long*) – codificano ogni miscela di luce mediante la combinazione di tre valori, ciascuno corrispondente alla quantità complessiva di attivazione di una classe.

Il *principio di univarianza* descrive un'importante caratteristica generale della codifica neurale. Anche quando lo stimolo è multidimensionale, la risposta del singolo neurone è unidimensionale: il neurone risponde di più o di meno, ma la quantità di attivazione (una variabile intensiva) dipende sia dall'intensità dello stimolo sia dalla sua prossimità al picco della funzione che descrive la sensibilità lungo lo spettro.

Come mostra la **Figura 2.10**, un osservatore con una sola classe di recettori (pur sensibili su quasi tutto lo spettro visibile) “non vede i colori”: al massimo può vedere una zona più chiara/scura di un'altra. Ecco perché. La sensibilità di una singola classe di recettori (nell'esempio, i coni M) varia in funzione della lunghezza d'onda, per cui la quantità di attivazione in M rimane uguale quando l'intensità della radiazione compensa la minore sensibilità. Un osservatore con una sola classe di recettori può annullare ogni differenza fra le due metà del dischetto, eguagliando la radiazione a , di 490 nm, e la radiazione b , di 600 nm. Siccome per i coni M la sensibilità a 490 nm è tre volte quella a 600 nm, per eguagliare le due radiazioni di diversa lunghezza d'onda (e far scomparire il contorno verticale che divide in due il dischetto) l'osservatore dovrà solo aumentare

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15
 l'intensità della radiazione a 600 nm, fino a renderla tre volte maggiore dell'intensità della radiazione a . Un osservatore tricromatico (con tre classi di coni attivi) e un osservatore dicromatico (con due classi di coni attivi) non potrebbero mai rendere uguali le radiazioni a e b modificando soltanto l'intensità di una delle due.

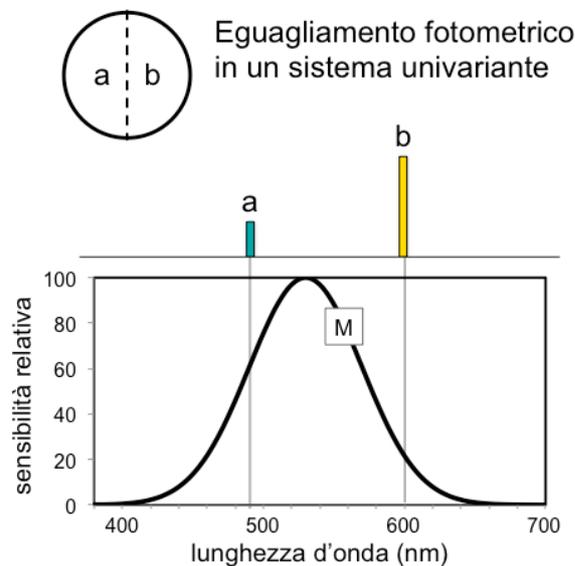


Figura 2.10. In un eguagliamento fotometrico l'osservatore deve rendere indistinguibili le due zone del disco in alto. L'osservatore univariante può eguagliare la radiazione a manipolando solo l'intensità della radiazione b . Il grafico in basso mostra che la sensibilità della classe di coni M a 490 nm è tripla rispetto a quella a 600 nm. L'osservatore univariante è facilmente ingannabile, essendo cieco alla differente composizione spettrale delle zone a e b del dischetto quando l'intensità in b è tre volte superiore a quella in a (rapporto illustrato dalle due barre dell'istogramma centrale).

Un osservatore con una sola classe di recettori viene chiamato *univariante*, perché può eguagliare l'attivazione provocata dalla radiazione in a variando l'intensità di una sola radiazione in b . Ogni osservatore umano è univariante alla luce delle stelle, dopo mezz'ora di adattamento al buio, perché in quel caso funzionano soltanto i bastoncelli, la cui curva di sensibilità (non illustrata nella [Figura 2.9](#)) ha un picco a circa a 500 nm. Se vi trovate all'aperto, i fiori potranno essere appena distinguibili dalle foglie, ma solo perché più scuri o più chiari; e in casa, prima di accendere la luce, vi potrà capitare di infilare una calza verde e una rossa prendendole per uguali. È questo il probabile fondamento del detto "Di notte tutti gatti sono bigi".

La [Figura 2.11](#) illustra un osservatore con due classi di coni (M e L) in cui la risposta a una singola radiazione a 560 nm in a viene eguagliata dalla risposta a un miscela ben calibrata, contenente una radiazione a 530 nm (di intensità pari circa ai 3/5 di quella in a) e una a 610 nm (di intensità pari circa ai 6/5 di quella in a). Questo è un

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15
osservatore bivalente, perché l'attivazione prodotta da una radiazione a risulta equivalente a quella prodotta da una miscela metamERICA b contenente almeno due radiazioni di intensità ben scelta.

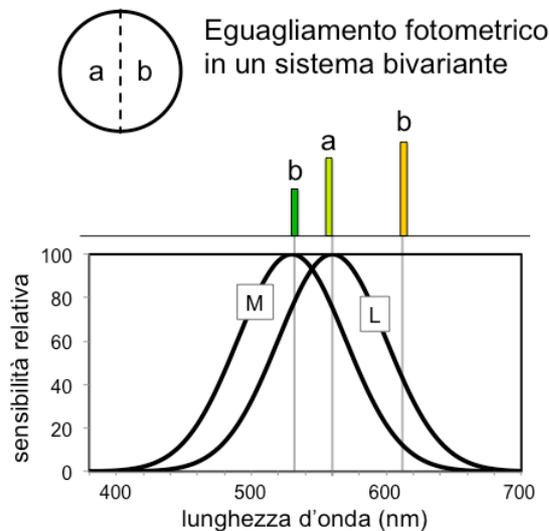


Figura 2.11. Un *osservatore bivalente* può eguagliare la risposta a una radiazione a manipolando le intensità di due radiazioni componenti la miscela b . L'eguagliamento fotometrico si ottiene quando, sia per i coni M sia per i coni L, le risposte alla radiazione a e alla miscela b sono identiche. Attivazioni identiche sia nella classe M sia nella classe L non sarebbero ottenibili utilizzando in b una sola radiazione.

L'osservatore umano normale con tre classi di coni (Figura 2.9) è trivariante, nel senso che l'attivazione prodotta da una radiazione a è eguagliabile miscelando in b almeno tre radiazioni di intensità ben scelta. Potete esercitarvi a individuare alcune soluzioni trivarianti generalizzando il procedimento usato nelle Figure 2.10 e 2.11.

Lo spazio di attivazione dei coni SML è un solido tridimensionale; ma se si prescinde dalla componente intensiva e si considerano solo i rapporti fra le tre attivazioni, allora la cromaticità è rappresentabile come posizione all'interno del cosiddetto triangolo di Maxwell (dal nome del suo ideatore, il grande fisico scozzese James C. Maxwell, 1831-1879), che costituisce il prototipo di altri spazi dei colori (Figura 2.12). Al centro del triangolo troviamo un colore acromatico (un grigio scuro se l'intensità complessiva dell'attivazione è debole, un bianco se l'intensità è molto elevata) in quanto le tre attivazioni SML sono perfettamente bilanciate (il punto acromatico è quindi equidistante da ciascuno dei tre vertici).

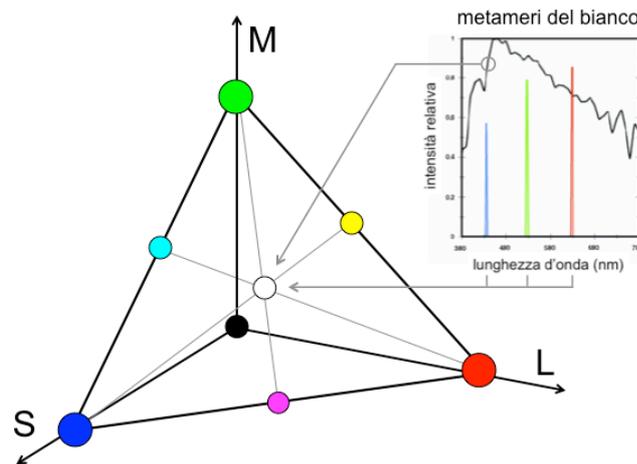


Figura 2.12. Lo spazio di attivazione dei coni è il triedro definito dalle tre semirette SML, all'origine del quale sta il nero (corrispondente all'assenza di attivazione in ciascun sottosistema). Sezionando il triedro con un piano si ottiene il *triangolo di Maxwell*: ciascun punto del triangolo corrisponde a un particolare rapporto fra le tre attivazioni, con il bianco equidistante dai tre vertici. Per l'osservatore descritto nella **Figura 2.9** i colori ai vertici sono virtuali, posto che è impossibile attivare una classe di coni senza attivarne contemporaneamente almeno un'altra. L'inserto a destra mostra due *metameri* della luce solare: il bianco prodotto da una sorgente artificiale standard (linea nera) e quello prodotto dalla miscela dei tre primari RGB (tre picchi che, se osservati uno alla volta, appaiono blu, verde, rosso). Per l'osservatore trivariante della **Figura 2.9** le due miscele, fisicamente eterogenee, mappano nello stesso punto dello spazio di attivazione dei coni e quindi appaiono dello stesso colore.

2.6.3. Sensibilità visiva alla frequenza spaziale

La possibilità di discriminare la composizione spettrale è indubbiamente utile, consentendoci di trovare facilmente frutta e fiori tra il fogliame anche quando questi mandano all'occhio la stessa intensità di luce. Ma in molti casi (come dimostra il cinema in bianco e nero) forme e movimenti possono essere veicolati efficacemente da immagini composte da zone diverse solo per intensità, a condizione che il contrasto sia sufficiente. In un mondo acromatico lo stimolo elementare è un reticolo sinusoidale, variabile per contrasto (ampiezza dell'onda che descrive il profilo di luminanza) e frequenza (numero di cicli nell'unità di spazio).

La sensibilità umana al contrasto in funzione delle frequenza spaziale è illustrata dall'andamento a U rovesciata osservabile a sinistra nella **Figura 2.13**. Il grafico a destra mostra che nell'adulto la sensibilità al contrasto più elevata (soglia più bassa) si ottiene in visione centrale per una frequenza spaziale di 3-4 cicli per grado di angolo visivo (3-4 alternanze bianco-nero in un cm, alla distanza di osservazione di 57 cm). Per l'adulto in visione periferica, a circa 10 gradi di eccentricità, la sensibilità diminuisce e trasla verso le frequenze più basse. Nel

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 bambino di tre mesi – le cui capacità visive sono misurabili sfruttando la preferenza visiva per gli stimoli più articolati (*preferential looking*) – la sensibilità è ancora minore e comunque traslata verso le frequenze più basse.

Diversamente da quanto accade per la sensibilità uditiva alla frequenza temporale degli stimoli acustici e alla sensibilità visiva per la frequenza spaziale delle radiazioni luminose, l'attributo visivo corrispondente alla frequenza spaziale è omologo alla corrispondente grandezza fisica. Reticoli a bassa frequenza appaiono radi; reticoli ad alta frequenza appaiono fitti.

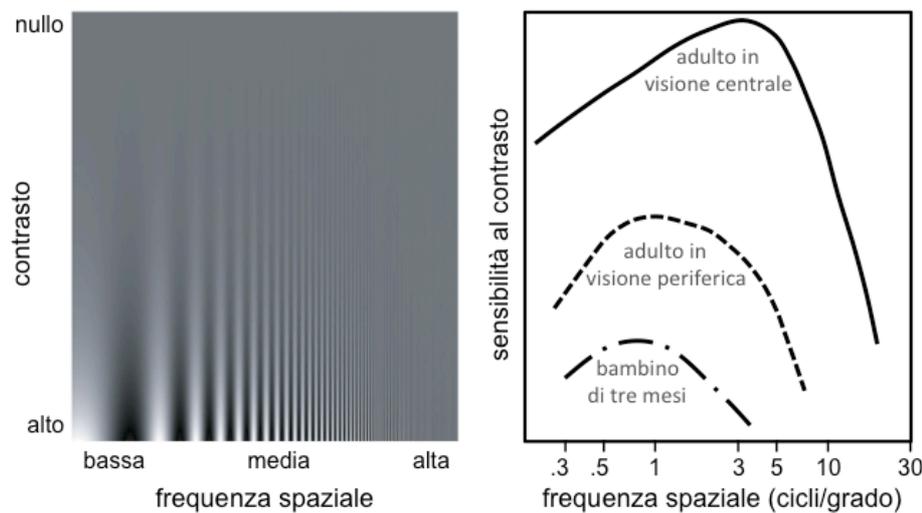


Figura 2.13. Nell'immagine a sinistra il contrasto varia da nullo (grigio omogeneo, in alto) ad alto (variazione sinusoidale da chiaro a scuro, in basso). Alla normale distanza di lettura la variazione chiaro-scuro appare discriminabile fino a livelli più bassi di contrasto (verso l'alto) quando la frequenza è intermedia. Come mostra il grafico a destra, la visibilità dei reticoli di diversa frequenza spaziale varia in funzione dell'osservatore e della zona del campo visivo.

2.7. Adattamento sensoriale e codifica neurale

L'*adattamento sensoriale* – cioè la perdita di sensibilità conseguente all'esposizione prolungata a una stimolazione costante (condizione innaturale ma facilmente riproducibile) – è un fenomeno cruciale per la scoperta del codice usato dal cervello per la registrazione delle proprietà elementari degli stimoli, al punto che è stato definito l'*elettrodo dello psicofisico*. Grazie ad esso, infatti, è possibile sondare i processi neurali utilizzando una procedura comportamentale.

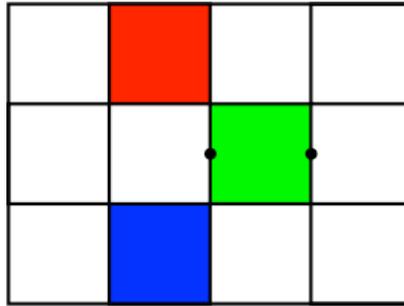


Figura 2.14. Immagine consecutiva negativa. In visione libera potete controllare che la scacchiera è bianca, salvo per i tre quadrati colorati. Fissate ora, contando fino a 10, il punto centrale e poi spostate lo sguardo sul punto di destra. In alcuni dei quadrati prima bianchi compariranno dei nuovi colori, dipendenti dall'adattamento selettivo sviluppato durante la breve fissazione del primo punto. Se continuate ad alternare la fissazione fra i due punti, solo alcuni quadrati resteranno acromatici.

L'adattamento può prodursi in ciascuno dei sottosistemi sensoriali deputati alla registrazione delle proprietà elementari (e anche relativamente complesse) degli stimoli, a riprova della generalità dei principi sottostanti la codifica neurale. Conseguenze semplici ma rilevanti dell'adattamento alla composizione spettrale della stimolazione sono le immagini consecutive negative (*negative afterimages*; **Figura 2.14**), un fenomeno chiaramente retinotopico, cioè legato alla topografia retinica. Per esempio, se una zona della retina viene affaticata dall'esposizione prolungata al verde (è il caso del quadrato a destra, mentre fissate il punto centrale) per un certo tempo la sua sensibilità a quelle radiazioni sarà indebolita, mentre resterà integra la capacità di rispondere alle altre radiazioni. Per questo motivo, quando l'occhio si sposta sul punto di destra, il quadrato alla sua destra non appare bianco ma lilla (il colore corrispondente a un bianco cui sia stato tolto il verde, cioè alla miscela rosso-blu delle componenti residue dello spettro). Altrettanto vale per il giallino del quadrato che sulla retina rimpiazza il blu e per il ciano che rimpiazza il rosso. Si noti che i colori cromatici osservabili nell'immagine consecutiva della **Figura 2.14** appaiono più luminosi di quelli dei quadrati inizialmente bianchi, che danno luogo (essendo l'immagine consecutiva una sorta di negativo fotografico) a un grigio più scuro.

La **Figura 2.15** illustra l'analogo effetto nel dominio della frequenza spaziale. L'effetto richiede di spostare lo sguardo in una piccola zona fra due reticoli di diversa frequenza spaziale, producendo uno specifico adattamento sensoriale, cioè la perdita selettiva di sensibilità per la frequenza spaziale del reticolo che cade in quella zona retinica. Spostare lo sguardo durante l'adattamento serve a prevenire il formarsi delle immagini consecutive negative, illustrate nella **Figura 2.14**. Tali immagini disturberebbero l'osservazione dell'effetto consecutivo a carico della frequenza spaziale,

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 che consiste nell'allontanamento apparente della frequenza dei reticoli osservati nella fase secondaria, rispetto alla frequenza di adattamento: il reticolo più fitto del reticolo di adattamento apparirà ancora più fitto; quello più rado del reticolo di adattamento apparirà ancora più rado.



Figura 2.15. Effetto consecutivo della frequenza spaziale. I reticoli a destra hanno la stessa frequenza spaziale, ma appariranno diversi dopo l'adattamento ai reticoli di sinistra. Per ottenere l'effetto dovete avere la pazienza di adattare il sistema visivo ai reticoli di sinistra (quello a bassa frequenza sopra e quello ad alta frequenza sotto) spostando lo sguardo sui quattro pallini centrali: contate lentamente fino a 60 spostando gli occhi da un pallino all'altro (se vi adattate più a lungo l'effetto sarà ancora migliore) e poi spostate lo sguardo sul pallino centrale di destra, fra i due reticoli obiettivamente uguali. Quello superiore apparirà un po' più fitto di quello inferiore.

Gli effetti di adattamento riguardano molte altre proprietà, visive e non. Un effetto particolarmente forte è l'*illusione della cascata*, facilmente osservabile soffermandosi a fissare una roccia immobile nel mezzo del precipitare dell'acqua (oppure cercando sul web una delle tante versioni della *waterfall illusion*). Se dopo mezzo minuto spostate lo sguardo sulle rocce immobili o sugli alberi, vedrete un paradossale movimento verso l'alto di tutta la parte del campo visivo prima occupata dalla cascata.

Anche l'illusione della cascata è spiegabile in base all'idea che la percezione del livello neutrale corrispondente all'immobilità (come il bianco nella Figura 2.14) derivi dal bilanciamento tra le risposte di varie classi di neuroni. Di norma le rocce e gli alberi appaiono immobili perché i deboli segnali inviati dai neuroni specializzati per i movimenti verso l'alto e verso il basso si equilibrano. L'adattamento alla cascata porta all'affaticamento selettivo dei neuroni specializzati per il movimento verso il basso. Quando portiamo lo sguardo su una scena immobile, questi neuroni rispondono di meno, dando luogo a uno squilibrio della risposta media verso l'alto.

Gli effetti consecutivi dell'adattamento hanno importanti somiglianze con gli *effetti di contrasto* osservabili in configurazioni statiche, in cui il contesto esterno produce una distorsione dello stimolo incluso. La **Figura 2.16** mostra due esempi classici di tali effetti: il contrasto simultaneo di bianchezza e l'illusione d'inclinazione

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 (*tilt illusion*). Entrambi gli effetti (come molti altri, a carico di altre proprietà elementari delle sensazioni) indicano che la registrazione sensoriale delle proprietà degli stimoli dipende dal contesto spaziale, oltre che temporale (come dimostrato dagli effetti di contrasto successivo, prima discussi). In generale, i sistemi sensoriali sembrano influenzati dal valore relativo della stimolazione; cioè dal rapporto fra lo stimolo e il suo contesto, adiacente nello spazio o nel tempo.

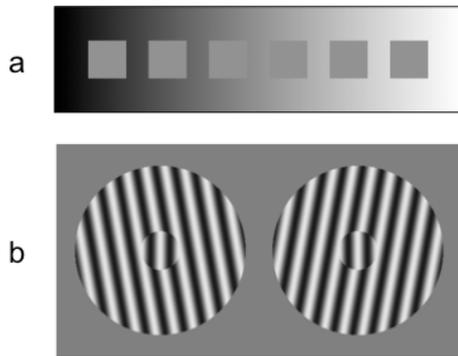


Figura 2.16. Due fenomeni di contrasto simultaneo. (a) I sei grigi sono fisicamente eguali, ma appaiono diversi, in funzione del contrasto con lo sfondo. (b) Entrambi i reticoli al centro sono verticali, ma quello a sinistra, incluso in un reticolo inclinato in senso antiorario, appare leggermente inclinato in senso orario; viceversa per quello di destra.

2.8. Epilogo

Nella scienza dell'esperienza diretta, lo studio della sensazione è la parte riguardante la registrazione delle grandezze fisiche fondamentali. La *registrazione sensoriale* avviene in modo efficiente ma non del tutto fedele, se per fedeltà si intende la riproduzione esatta delle proprietà metriche e dimensionali degli stimoli messi a disposizione dall'ambiente. In molti casi la registrazione sensoriale è assimilabile a una misurazione effettuata mediante uno strumento (come un metro, una bilancia o un termometro). Non accade così per la codificazione sensoriale della composizione spettrale, che comporta una massiccia riduzione di dimensionalità: si passa dalle numerosissime radiazioni componenti l'input ottico, ciascuna con la propria intensità, alle sole tre dimensioni dello spazio dei colori dell'osservatore umano.

La registrazione sensoriale non si adegua nemmeno all'ideale della copia "punto a punto": come dimostrano molte illusioni, la sensazione è influenzata, oltre che dal corrispondente stimolo locale, anche dalla stimolazione adiacente. È forse questo il motivo principale per abbandonare il modello della registrazione a favore di un modello alternativo, basato sui concetti di *interpretazione* e *rappresentazione selettiva*. È quanto faremo nel capitolo sui processi percettivi di base.

Riquadro 2.1. Costante di Weber.



Ernst H. Weber (1795-1878)

La discriminazione tra due intensità non dipende dalla loro differenza ma dal loro rapporto. Weber (1834) giunse a questa conclusione confrontando le soglie differenziali in varie modalità sensoriali. La formula più comune per quella che Fechner chiamò *legge di Weber* è:

$$\Delta I / I = k ;$$

dove ΔI è l'incremento necessario a produrre una differenza appena percepibile (soglia differenziale o *just noticeable difference*, JND) tra uno stimolo standard di intensità I e lo stimolo di confronto di intensità $(I + \Delta I)$, mentre k è la "costante di Weber".

Una formulazione più trasparente è:

$$\Delta I = k I .$$

Qui è esplicito che la JND *non* è costante, ma proporzionale all'intensità di riferimento. Per esprimere che la discriminazione tra due stimoli dipende dal loro rapporto (e non dalla loro differenza) osserviamo che $I_1 / I_2 = (1 + k)$, facilmente derivabile dalle precedenti: $(I + \Delta I) / I = (I + k I) / I = (1 + k)$.

La relazione individuata da Weber in ambito sensoriale ha un illustre antecedente in ambito economico, dove un secolo prima il matematico svizzero Johann Bernoulli (1667-1748) trovava ovvio ipotizzare che il valore soggettivo di un incremento di ricchezza nominalmente identico diminuisse al crescere della ricchezza già posseduta: "Non c'è dubbio che un guadagno di mille ducati sia più significativo per un povero che per un ricco, nonostante il guadagno sia uguale per entrambi". L'ipotesi di Bernoulli ha trovato fondamento empirico nella ricerca contemporanea sulle decisioni (cap. xx decisioni")

Riquadro 2.2. Metodi della psicofisica classica.



Gustav T. Fechner (1801-1887)

Negli *Elemente der Psychophysik* (1860) Fechner formulò la legge logaritmica e codificò tre procedure psicofisiche: metodo dell'aggiustamento, metodo dei limiti, metodo degli stimoli costanti. Ipotizziamo di voler trovare il quadrato che appare di area uguale a un disco standard.

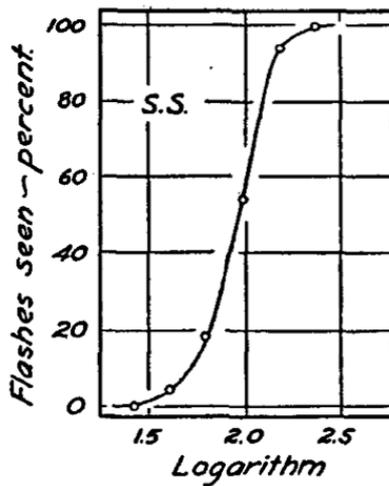
Nel *metodo dell'aggiustamento* lo sperimentatore modifica in modo continuo l'area del quadrato, fino a quando l'osservatore la giudica identica a quella del disco standard. La procedura (ripetibile a piacere) è un po' approssimativa, ma rapida e vantaggiosa per uno studio preliminare.

Nel *metodo dei limiti* lo sperimentatore predispone una ventina di quadrati di area diversa, a intervalli uguali, compresi tra un quadrato che appare chiaramente minore del disco a uno che appare chiaramente maggiore. Poi alterna una serie crescente e una decrescente. Nella prima serie, crescente, l'osservatore dice "più piccolo" per varie prove, fino a quando non dice "più grande" (se rispetta le istruzioni e non usa la risposta "uguale"); a quel punto lo sperimentatore prende nota del punto di transizione da una risposta all'altra e presenta la serie decrescente, che comincia con prove in cui la risposta "più grande" è prima facile e poi più difficile. Quando l'osservatore passa alla risposta "più piccolo" lo sperimentatore annota il nuovo punto di transizione e continua ad alternare serie crescenti e decrescenti, fino a quando dispone di una decina di punti di transizione, la cui media costituisce la miglior stima del *punto di eguaglianza soggettivo* (PES).

Il *metodo degli stimoli costanti* è descritto nel testo a proposito del peso sollevato. Per eguagliare il disco possiamo scegliere 4 quadrati più grandi e 4 più piccoli, in modo da avere 8 aree di confronto collocate a intervalli uguali tra un'area chiaramente più piccola del disco a una chiaramente più grande. Deciso il numero di ripetizioni di ogni coppia confronto-standard (poniamo 10), si presenta all'osservatore la sequenza di 80 prove in ordine casuale.

Nella versione originaria, il metodo dei limiti e il metodo degli stimoli costanti richiedono di sottoporre all'osservatore molte prove poco informative, in cui lo stimolo di confronto è lontano dal PES da stimare. Tale difetto viene eliminato in alcune varianti *adattive*, come il *metodo della scala* (*staircase method*), che, sulla base di una prima stima grossolana del PES, genera stimoli di confronto per i quali la discriminazione dallo standard è sempre più difficile.

Riquadro 2.3. Soglia assoluta.



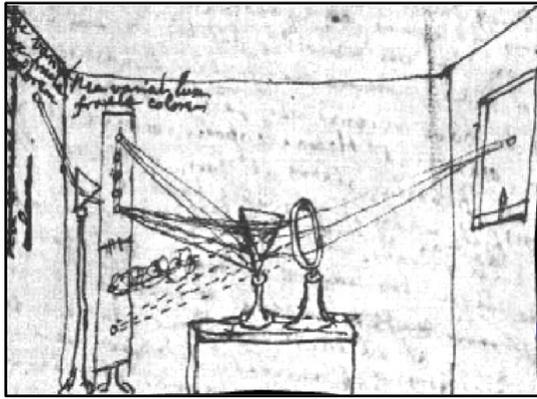
La *soglia assoluta* è definita come la minima intensità rilevabile, cioè discriminabile dalla condizione in cui lo stimolo è assente. In questo senso la soglia assoluta è un caso speciale di *soglia differenziale*, quello dell'incremento rispetto all'intensità zero.

In una delle pietre miliari della psicofisica della visione, Hecht, Shlaer e Pirenne (1942) misurarono il *numero minimo di fotoni* in grado di produrre una sensazione di luce, per un osservatore perfettamente adattato al buio e stimolato da un breve piccolo flash nella zona del campo visivo, a 20° di eccentricità, in cui la sensibilità retinica è massima. Il grafico descrive la prestazione di S.S., il secondo autore (tra

i 7 osservatori c'erano gli autori), espressa come percentuale di risposte "ho visto la luce" su 50 flash. Il numero di fotoni per flash varia da 24 a 221, su un'ascissa logaritmica. Per la soglia assoluta gli autori scelsero il 60% di risposte positive, che per S.S. corrispondeva a uno stimolo prossimale con circa 100 fotoni (2 unità logaritmiche) in un millisecondo di esposizione. Poiché ai fotorecettori retinici arriva solo un decimo dell'energia luminosa misurata in prossimità della pupilla e poiché il flash, pur piccolo, stimolava una regione popolata da circa 500 fotorecettori, Hecht *et al.* stimarono che in media un flash risulta visibile quando 10 fotorecettori assorbono un singolo fotone (la quantità elementare di luce, nell'ottica corpuscolare). Questa stima rivela che l'occhio umano può comportarsi – in condizioni favorevoli – come uno strumento meravigliosamente adattato al limite fisico dell'energia misurata.

Tuttavia, il compito utilizzato da Hecht *et al.* è criticabile, poiché non tiene sotto controllo la propensione (*bias*) a favore o contro la risposta positiva. Un osservatore con *bias* conservativo (poco propenso a dire sì) otterrebbe una soglia più elevata rispetto a uno con *bias* liberale (molto propenso a dire sì). Un compito che mette al riparo dal rischio di confondere la vera sensibilità (la soglia per l'osservatore *unbiased*) con il *bias* (fattore soggettivo indipendente dallo stimolo) è la scelta forzata tra due alternative (*two-alternative forced choice, 2AFC*), in cui si presentano due intervalli, spaziali o temporali, uno solo dei quali contiene lo stimolo, mentre nel compito *yes/no* l'intervallo è unico e l'osservatore sa che uno stimolo, pur debole, c'è sempre. Nel compito 2AFC, in cui l'osservatore deve scegliere l'intervallo che contiene lo stimolo, la curva psicometrica sale dal 50 al 100% delle risposte corrette, dato che per stimoli troppo deboli l'osservatore andrà a caso tra sinistra/destra o prima/dopo (e quindi non si otterranno percentuali di risposte corrette inferiori al 50%). Per la soglia assoluta stimata mediante il compito 2AFC viene tipicamente scelto il 75% di risposte corrette.

Riquadro 2.4. Scomposizione della luce solare.



La scienza moderna del colore inizia in una stanza buia del Trinity College di Cambridge. Un po' prima del 1670, il grande fisico e matematico inglese Isaac Newton (1642-1727) vi condusse alcuni esperimenti rivoluzionari in base ai quali concluse che il bianco (così appare una macchia di luce quando un sottile fascio di luce solare penetra dall'esterno) in realtà contiene tante radiazioni diversamente colorate. Newton giunse a questa conclusione utilizzando un prisma di vetro,

in grado di rivelare che le radiazioni contenute nella luce solare sono diversamente rifrangibili.

La sequenza degli esperimenti è geniale. Per dimostrare che non era il prisma a colorare la luce bianca (ma che esso semplicemente dispiegava i raggi paralleli che producono il bianco) Newton interpose un secondo prisma lungo la traiettoria di una radiazione elementare scomposta dal primo prisma, arrivando a dimostrare che ogni radiazione elementare non è ulteriormente rifrangibile. Ciò contraddice l'idea – basata sull'evidenza fenomenologica – che il bianco sia un colore puro, qualificandolo invece come miscela dei colori più diversi. Cosa altrettanto interessante, Newton dimostrò che per riottenere il bianco non sono indispensabili tutte le radiazioni ottenute dalla prima scomposizione con il prisma. Un buon bianco (magari meno intenso, ma comunque acromatico) è ottenibile miscelando un piccolo numero di radiazioni elementari: addirittura solo tre, purché ben spaziate sullo spettro.

I *metameri* (miscela fisicamente diverse che appaiono indistinguibili) sono la chiave per la comprensione del sistema di codificazione della luce. Come il bianco, anche i colori spettrali – prodotti cioè da radiazioni elementari – sono mimabili miscelando in maniera opportuna altre radiazioni che non li contengono. La corrispondenza tra composizione spettrale e colore percepito, infatti, non è iniettiva: mentre ogni composizione spettrale corrisponde a uno specifico colore, di regola ogni colore corrisponde a più composizioni spettrali. Ciò rende possibile, tra l'altro, la tecnologia del colore artificiale, in grado di generare un'ampia tavolozza di tinte mediante composizione additiva di poche radiazioni primarie che, in isolamento, appaiono di colore rosso, verde, blu (*red, green, blue*; RGB).

I fondamenti della codificazione sensoriale della composizione spettrale vennero posti, all'inizio dell'Ottocento, dal medico inglese Thomas Young (1773-1829), il quale comprese che Newton, suggerendo che ogni radiazione elementare attivi una specifica classe di recettori retinici, aveva suggerito un'idea incompatibile con l'elevata risoluzione spaziale della visione. Se a ognuna delle radiazioni elementari corrispondesse un recettore specializzato, in ogni piccola zona della retina dovrebbero essere stipati moltissimi recettori diversi. Nel rispetto dell'anatomia Young propose un sistema di codificazione, recepito poi da Hermann von Helmholtz, basato su tre sole classi di recettori, ciascuna sensibile su un'ampia zona dello spettro, ma con picchi di sensibilità distanziati quanto basta a produrre risposte differenti (Figura 2.9).

3. Processi percettivi di base

3.1. Indeterminazione ottica e percezione visiva

Se la sensazione è il terreno della *psicofisica locale*, centrata sulle grandezze fisiche fondamentali, allora la percezione è il terreno della *psicofisica globale*, centrata sul ruolo di ciò che sta intorno allo stimolo locale, definibile in termini generici come “contesto”. Ma perché la percezione è così influenzata dalle variabili contestuali, al punto da rendere impraticabile il modello della registrazione “punto a punto”, utilizzato nel **cap. 2**?

La risposta rinvia a un limite cruciale dell'informazione sensoriale, molto evidente nel caso della visione, la modalità su cui ora ci concentreremo. La distribuzione di luce che arriva a un punto di vista è in grado di dire molto sulla forma e costituzione materiale degli oggetti fisici, ma non tutto. I limiti della luce come veicolo di informazione sul mondo vengono complessivamente indicati come **indeterminazione ottica**.

3.1.1. Indeterminazione geometrica e fotometrica

L'*indeterminazione* che grava sull'input ottico è di due tipi: *geometrica* (a carico di proprietà quali distanza, forma, grandezza, velocità) e *fotometrica* (a carico del colore delle superfici).

Per quanto riguarda l'indeterminazione geometrica, l'argomento classico sull'insufficienza dell'informazione ottica come base per la valutazione della distanza venne sviluppato dal filosofo e teologo irlandese George Berkeley (1685-1753) [Riquadro 3.1]. In sostanza, Berkeley mise in evidenza l'*indeterminazione radiale*, cioè l'impossibilità di determinare quanto distante sia la causa remota di un raggio elementare che arriva all'occhio; questo limite elementare – relativo cioè a un singolo raggio di luce – è alla base di molti fenomeni cospicui e talora curiosi.

Un effetto molto istruttivo riguarda la *grandezza dell'immagine consecutiva*, un oggetto illusorio il cui fondamento sensoriale è stato discusso nella **sezione 2.6.2**. Adattate selettivamente il vostro sistema visivo fissando il centro del disco rosso della **figura 3.1** per almeno 30 s, dato che poi non ci saranno i contorni a tratto che nella **figura 2.14** facilitavano l'emergere della colorazione illusoria. Preparatevi inoltre a guardare l'immagine consecutiva su una parete vuota lontana, dove avrete collocato un punto di fissazione. Se state tenendo il libro a circa 50 cm, l'immagine consecutiva vista sulla parete a 5 m di distanza avrà un diametro 10 volte superiore.

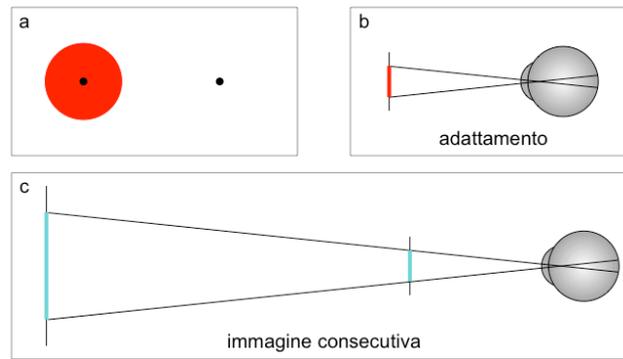


Figura 3.1. Immagine consecutiva e *legge di Emmert*. (a) Fissate il centro del disco rosso per 30 s, senza muovere gli occhi e la testa. Completato l'adattamento, fissate il punto a destra (sulla pagina del libro) e vedrete comparire un disco color ciano grande quanto il disco rosso. Se invece spostate lo sguardo su un punto al centro di una parete vuota lontana, vedrete comparire un *grande* disco color ciano. (b) Sezione laterale di un occhio che fissa il disco rosso vicino, ricavandone uno sbilanciamento della sensibilità nella zona retinica corrispondente all'angolo ottico sotteso dal disco. (c) L'*invarianza grandezza/distanza* fa sì che l'immagine consecutiva appaia grande quanto il disco rosso inducente se vista alla stessa distanza, e proporzionalmente più grande se vista a una distanza maggiore.

La proporzionalità diretta tra grandezza dell'immagine consecutiva e distanza di osservazione viene chiamata *legge di Emmert*, dal nome dello studioso svizzero che la descrisse nel 1881, e si basa sull'**invarianza grandezza/distanza**, una relazione geometrica che derivabile dall'argomento di Berkeley sull'indeterminazione radiale: lo stesso angolo ottico (quello definito dal disco rosso inducente) può corrispondere a un oggetto piccolo vicino o a un oggetto grande lontano. L'angolo ottico resta invariante perché ancorato alla zona adattata; per cui la grandezza dell'immagine consecutiva viene a dipendere dalla sua distanza.

I curiosi effetti dell'invarianza grandezza/distanza non si osservano soltanto con l'immagine consecutiva. Essi sono onnipresenti nella percezione pittorica. Nella **Figura 3.2** le tre automobili appaiono di grandezza diversa pur essendo identiche sulla carta (e quindi sottendendo lo stesso angolo visivo). L'effetto illusorio dipende dalle informazioni prospettiche che suggeriscono una distanza dall'osservatore maggiore per l'auto più in alto, intermedia per quella in mezzo e minore per quella in basso. In base al principio dell'invarianza grandezza/distanza la grandezza percepita delle tre auto è direttamente proporzionale alla distanza specificata dalle informazioni prospettiche.

Si tenga presente che il principio dell'invarianza grandezza/distanza (e quindi la legge di Emmert, se ci riferiamo all'immagine consecutiva) va contro l'aspettativa ingenua che un oggetto, allontanandosi, appaia più piccolo. Ciò accade per gli oggetti

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 fisici di grandezza costante, che proiettano un'immagine retinica via via minore all'aumentare della distanza dal punto di vista, fino alla loro sparizione. Ma gli esempi illustrati nelle **Figure 3.1 e 3.2** si riferiscono al caso opposto: cioè alla situazione in cui l'angolo ottico è costante, mentre l'oggetto può apparire più o meno lontano. In questo caso la grandezza percepita è direttamente (non inversamente) proporzionale alla distanza.



Figura 3.2. Tutti vedono tre automobili di grandezza diversa. Misurandole con un righello si constata che la loro grandezza pittorica è identica.
<http://richardwiseman.wordpress.com/>

L'*indeterminazione fotometrica* riguarda invece il colore delle superfici, che la quantità e tipo di luce riflessa verso l'occhio non può specificare completamente. Il colore superficiale è parzialmente indeterminato, perché l'intensità della luce che arriva al punto di vista da una superficie materiale è il prodotto della riflettanza (la proporzione di luce incidente riflessa verso l'occhio) e della quantità di illuminazione: quindi la stessa intensità di stimolazione può derivare da una superficie a bassa riflettanza ma molto illuminata, oppure da una superficie ad alta riflettanza ma poco illuminata. Una conseguenza emblematica dell'indeterminazione fotometrica è illustrata in una famosa dimostrazione di Edward H. Adelson (<http://persci.mit.edu/people/adelson>; **Figura 3.3**). Se la scena è sufficientemente articolata e sono disponibili informazioni sulla presenza di zone diversamente illuminate, il sistema visivo percepisce colori superficiali diversi anche quando le corrispondenti regioni dell'immagine hanno luminanza uguale (nella Figura 3.3 la luminanza di A è uguale a quella di B). In questo caso vale l'*invarianza riflettanza × illuminazione*: la stessa luminanza (derivante dalla stessa quantità d'inchiostro sulla pagina del libro) appare come un grigio scuro e molto illuminato (tassello A) oppure come un grigio chiaro e in ombra (tassello B).

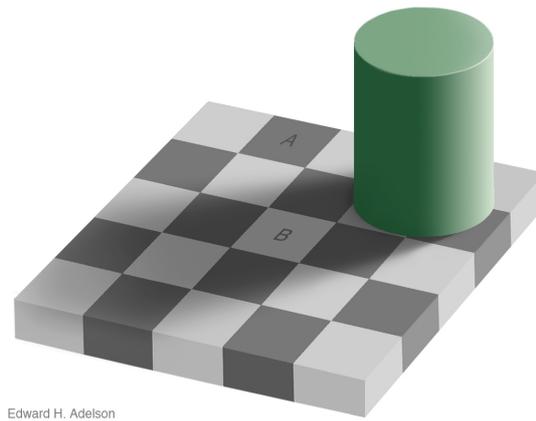


Figura 3.3. Le zone A e B appaiono molto diverse, ma sulla carta il grigio è il medesimo. Verificate l'uguaglianza sovrapponendo uno stesso grigio campione su ciascuna delle due regioni. L'immagine riproduce una circostanza assai comune nella vita quotidiana: un cartoncino chiaro ma in ombra può riflettere verso l'occhio la stessa intensità riflessa da un cartoncino scuro in piena luce.

3.1.2. La catena psicofisica

Per essere completa, una teoria della visione deve collegare appropriatamente le entità schematizzate nella **Figura 3.4**: lo stimolo distale (gli oggetti fisici che postuliamo come esistenti e dotati di proprietà materiali anche in assenza di luce); lo stimolo prossimale (le immagini o proiezioni di tali oggetti che possono essere raccolte in un qualsiasi punto di un ambiente percorso dalla luce e privo di ostacoli come la nebbia, ovviamente); gli eventi neurali (dal pattern di attivazione dei fotorecettori retinici fino alla distribuzione di attività cerebrali ritenuta isomorfa all'esperienza percepita); l'esperienza percepita. Queste entità sono i costituenti fondamentali della **catena psicofisica** che, nel bene e nel male, vincola l'osservatore al mondo esterno.

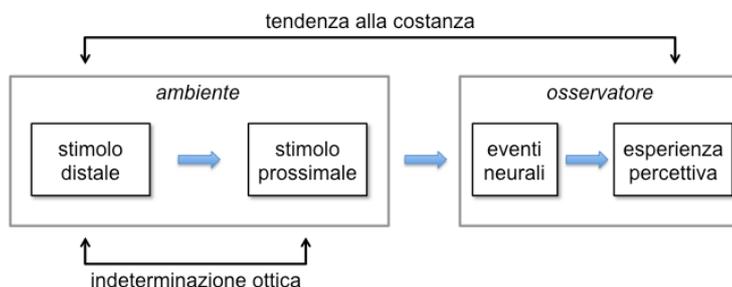


Figura 3.4. La catena psicofisica nella visione. Lo stimolo distale sono delle proiezioni ottiche parzialmente indeterminate, che non possono specificare totalmente lo stimolo distale. Gli oggetti disponibili nell'esperienza percettiva tendono a corrispondere agli stimoli distali, e quindi a rimanere costanti nonostante la variabilità degli stimoli prossimali.

Si tratta tuttavia di una catena unidirezionale, come indicano le frecce nella Figura

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

3.4. Rispetto a un dato punto di vista, l'immagine (la proiezione ottica che costituisce lo stimolo prossimale) è determinata in modo univoco dall'oggetto 3D (lo stimolo distale), mentre non è vero l'inverso. Partendo dall'immagine non è possibile ricostruire totalmente l'oggetto che l'ha determinata. Su questa impossibilità logica si fondano le ambiguità della grandezza percepita, già illustrate nelle **Figure 3.1 e 3.2**. Immaginiamo che qualcuno ci chieda di guardare attraverso il buco della serratura di una porta che mai abbiamo aperto. Chi ci garantisce che quella che stiamo vedendo nel buco della serratura sia una stanza vera, con mobili della grandezza giusta e sedie buone per sedersi, o non sia piuttosto una stanzetta in miniatura? Proprio nessuno. Stanzette del genere vennero realizzate, nel Seicento, per far riflettere sul potere illusionistico dell'arte e sui limiti della percezione. La grandezza dell'immagine che cade sul fondo dell'occhio non può dire nulla sull'effettiva grandezza dell'oggetto che l'ha determinata, se non si conosce la sua distanza. L'armadio di una casa di bambola, alto 10 cm, proietta alla distanza di 30 cm la stessa immagine proiettata da un armadio alto un metro, alla distanza di 3 m.

In generale l'osservatore, essendo a contatto con le immagini e non direttamente con gli oggetti (**Figura 3.4**), si deve confrontare con vari problemi riconducibili al concetto di *ottica inversa*, nei quali la soluzione (cioè, risalire dall'immagine all'oggetto che l'ha generata) dipende dalla possibilità di superare in qualche modo l'indeterminazione intrinseca alle proiezioni ottiche. Su questo punto le teorie della percezione si dividono [**Riquadro 3.2**].

I problemi di ottica inversa sono «mal posti» (in senso matematico), come emerge dall'analogia con una semplice operazione aritmetica. Dati due fattori, il loro prodotto è determinato in modo univoco. Il problema $[6 \times 4 = ?]$ è ben posto e contempla un'unica soluzione. Il problema inverso, invece, è mal posto: che senso ha infatti chiedersi $[24 = ? \times ?]$. Anche vincolando il problema al prodotto di due interi e trascurando l'ordine dei fattori, restano non una ma più soluzioni: $[24 \times 1]$, $[12 \times 2]$, $[8 \times 3]$, $[6 \times 4]$. Come scegliere una piuttosto che un'altra?

Ebbene, in molti casi l'immagine ottica contiene proprietà che risultano dalla combinazione di più condizioni fisiche, univocamente determinate dall'ottica; ma il tentativo di ricostruire il mondo fisico a partire dall'immagine è una “missione impossibile”, in assenza di vincoli o principi aggiuntivi.

3.1.3. Energia luminosa e informazione ottica

Chi ragiona sulla percezione visiva può trovarsi a pensare che la presenza di luce

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 sia una condizione indispensabile per la visione degli oggetti. Questa convinzione è addirittura rafforzata dall'uso dell'illuminazione artificiale. Basta pigiare un interruttore e la stanza nella quale fino a un attimo prima era possibile distinguere una varietà di oggetti colorati sprofonda nel buio. Stiamo ovviamente parlando di una situazione controllata, in cui porte e finestre siano ben chiuse. Sulla base di questa semplice osservazione, possiamo concludere che l'assenza di luce precluda la visione degli oggetti? Certamente sì. Ma possiamo anche concludere che la presenza di luce sia una condizione sufficiente affinché un osservatore normale veda degli oggetti? Certamente no. La seconda conclusione non discende logicamente dalla prima, nel senso che non è un'inferenza valida (cfr. cap. 20). Invece il nostro pensiero, cioè la logica ingenua, viene facilmente catturato dall'idea che «se non-luce allora non-visione» implichi «se luce allora visione».

Azionando l'interruttore noi operiamo quella che per il metodo sperimentale è una fatale confusione tra due variabili: l'energia luminosa e l'informazione veicolata dalla sua distribuzione spaziale. Tale confusione è eliminata in un esperimento naturale molto istruttivo: la nebbia. Chi va in montagna può ritrovarsi immerso in una nuvola, circondato da un latte luminosissimo nel quale è tuttavia costretto a muoversi alla cieca, come nel buio. La nebbia dimostra che la luce, anche molta luce, non serve a nulla se è omogenea: perché sia possibile vedere degli oggetti, l'immagine che arriva agli occhi deve contenere delle disomogeneità, dei dislivelli di luce corrispondenti ai bordi oggettuali. La nebbia non elimina l'energia luminosa, utilissima per l'abbronzatura; distrugge l'informazione ottica, indispensabile per la visione. Deviando in tutte le direzioni i raggi riflessi dalle superfici, la nebbia elimina quella preziosa relazione di corrispondenza tra bordi degli oggetti fisici e dislivelli nella distribuzione della luce che costituisce l'informazione normalmente utilizzata dal sistema occhio-cervello per costruirsi, appunto, una visione del mondo. Quando azioniamo l'interruttore e sopprimiamo l'illuminazione, facciamo succedere due cose in un solo momento: all'occhio non arriva più luce (cioè, il flusso di energia va bruscamente a zero); ma arriva anche la stessa quantità di luce – una quantità nulla – da ogni direzione dello spazio circostante la pupilla. La nebbia dimostra invece che la variabile importante per la visione degli oggetti non è l'*energia luminosa* ma l'**informazione ottica** [Riquadro 3.3].

In questo capitolo tratteremo soltanto la percezione visiva, per poter ancorare i nostri ragionamenti alle dimostrazioni presentate nelle figure. In particolare, discuteremo i processi corrispondenti alla visione di livello intermedio, grazie ai quali

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15
 l'informazione ottica registrata e codificata a livello della visione di basso livello (trattata nel cap. 2) viene organizzata e interpretata, prima di diventare l'input per i processi di riconoscimento (cap. 3). La visione di livello intermedio corrisponde, sostanzialmente, all'emergere di entità strutturate. Parleremo quindi dell'intelaiatura dello spazio visivo, della distinzione tra figura e sfondo, del modo in cui l'occhio integra le parti nascoste e riconduce a un piccolo numero di unità percettive l'infinità di stimoli locali contenuti nelle immagini ottiche. La nostra sarà solo un'incursione nel grande territorio della percezione; ma che forse vi farà venire la voglia di esplorarlo in modo sistematico.

3.1.4. Codificazione e organizzazione

Informazione è un concetto *relazionale*: non ha senso pensarla come un'entità materiale indipendente da un osservatore in grado di utilizzarla. Ma non tutti gli osservatori sono eguali. Quanto meno, dobbiamo tenere ben distinto l'**osservatore ideale** (in grado di utilizzare tutta l'informazione disponibile) dall'**osservatore reale** (un organismo come noi, in grado di utilizzarne soltanto una parte).

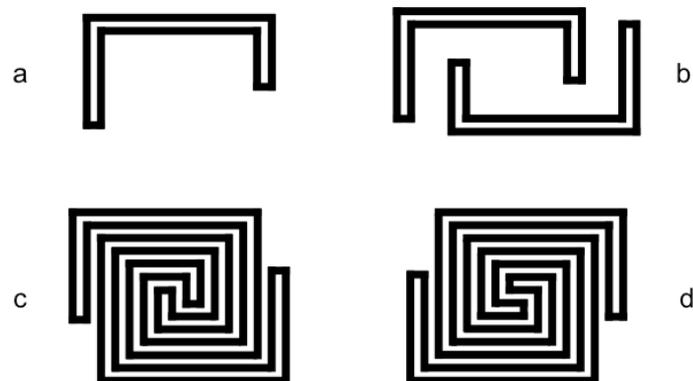


Figura 3.5. Le configurazioni *a* e *b* sono chiaramente discriminabili: in *a* c'è un oggetto; in *b* due. Le configurazioni *c* e *d* sono invece inafferrabili: usando solo lo sguardo è impossibile capire se si tratta di un oggetto oppure di due.

La **figura 3.5** esemplifica questa distinzione. La differenza tra *a*, in cui si vede una forma, e *b*, in cui si vedono due forme, è del tutto evidente: sembra che per l'occhio *codificare* la differenza tra un unico oggetto connesso e due oggetti sconnessi sia banale. Qualunque sia il codice utilizzato dal sistema visivo (una variazione nella scarica di neuroni specializzati o un altro parametro del linguaggio del cervello) una differenza topologica tra le due immagini sembra conservata anche nella codifica visiva. Chiediamoci ora cosa sono le spirali *c* e *d*. Sono entrambe delle configurazioni

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 sconnesse, ciascuna paragonabile a due serpenti che hanno deciso di avvolgersi in spire uno accanto all'altro, o due configurazioni connesse, ciascuna composta da un unico serpente di lunghezza doppia? Oppure una delle due spirali è connessa e l'altra è sconnessa (ma quale delle due è connessa e quale sconnessa)? L'osservazione a occhio nudo non consente di rispondere con sicurezza. Naturalmente possiamo scoprire come stanno le cose partendo dalle estremità esterne e seguendo le spire verso l'interno, magari aiutandoci con una matita. Ma resta il fatto che questa operazione analitica richiede un'attenzione e un controllo inusuali, del tutto estranei al modo in cui normalmente guardiamo il mondo. E anche quando, grazie all'aiuto della matita, riusciamo a scoprire quale configurazione è connessa, basta distogliere lo sguardo perché tutto torni come prima. La **figura 3.5** va contro l'ipotesi che la connessione, pur essendo una proprietà topologica fondamentale, sia sempre rilevabile. Al contrario, se la densità dei bordi è eccessiva la connessione/sconnessione non viene rilevata e la percezione rimane indeterminata. In sostanza, siamo davanti a qualcosa che c'è ma non si vede.

La ricerca scientifica sulla percezione sfrutta spesso la contrapposizione tra quello che c'è nell'immagine (e che per definizione è colto dall'osservatore ideale, capace di vedere tutto quanto è specificato dall'informazione ottica) e ciò che viene effettivamente percepito da un osservatore reale. Si noti che nelle configurazioni *cd* la discriminazione percettiva fallisce non per la cattiva qualità ottica delle immagini, che sono ottime (sensazioni sopra soglia, contrasto elevato), ma per qualche limite del processo di organizzazione delle informazioni presenti nelle configurazioni *cd*, che non è in grado di generare rappresentazioni distinte per una o due entità percettive. La cecità percettiva alla connessione/sconnessione è un caso limite. Sono invece comuni le situazioni in cui la *ricerca visiva* di un *target* (il "bersaglio" dell'attenzione) è possibile, ma con maggiore o minore facilità (**cfr. cap. 6, par. 2.2**).

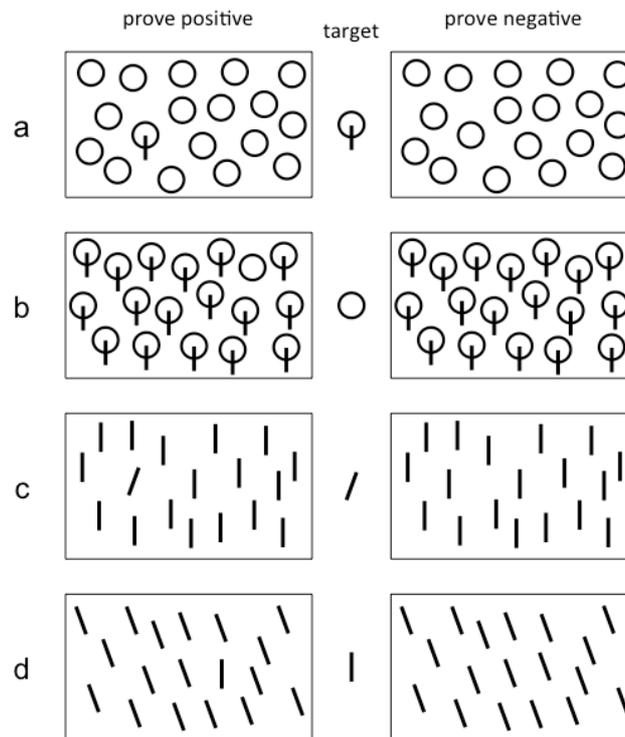


Figura 3.6. Nei compiti di ricerca visiva l'osservatore deve dire che il target (lo stimolo riprodotto al centro) è presente nelle prove positive (con il target in mezzo a parecchi distrattori, colonna di sinistra) ma non nelle prove negative (solo distrattori e niente target, colonna di destra). Il confronto fra le condizioni *ab* dimostra l'*asimmetria della ricerca visiva*. Trovare un cerchio barrato fra i cerchi (a) è più facile che trovare un cerchio fra i cerchi barrati (b). Analogamente, trovare una barretta obliqua fra tante barrette diritte (c) è più facile che trovare una barretta diritta fra tante barrette oblique (d).

La **figura 3.6** ci permette di apprezzare quando il *target* salta all'occhio (il termine inglese usato in letteratura è *pop out*) fra i “distrattori” e quando invece la discriminazione della sua presenza/assenza richiede una scansione sistematica. Nella condizione **a** è facile decidere se fra tanti cerchi non barrati c'è anche un cerchio barrato (prova positiva a sinistra) oppure no (prova negativa a destra). Non altrettanto facile è trovare, nella condizione **b**, il cerchio non barrato in mezzo ai cerchi barrati. In altri termini, rilevare l'*aggiunta* di una caratteristica elementare (la barretta verticale) è più facile che rilevarne l'*eliminazione*, a parità di differenza locale tra target e distrattore. Il corrispettivo, nella vita quotidiana, sarebbe l'asimmetria tra la facilità con cui troviamo un uomo con il cappello in una folla di uomini senza cappello e la relativa difficoltà con cui troviamo un uomo senza cappello in una folla di uomini con il cappello. Nel caso appena illustrato, l'effetto *odd man out* (“il tipo strano si nota”) dipende dall'esistenza di un livello zero (il semplice cerchio) rispetto al quale l'aggiunta della barretta crea una condizione atipica. Non conta soltanto la diversità dalla maggioranza, ma la direzione di tale diversità.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

Sempre **nella figura 3.6** il confronto tra le condizioni **c** e **d** dimostra che nel continuum degli orientamenti la verticale funge da riferimento (da “ancora”). Anche qui l'asimmetria consiste nella maggiore individuabilità dell'elemento atipico, deviante dalla norma. Trovare un segmento obliquo in mezzo a quelli verticali **c** è più facile che trovare un segmento verticale in mezzo a quelli obliqui **d**, nonostante la differenza di inclinazione tra *target* e distrattore sia di 20 gradi in entrambi i casi.

L'**asimmetria della ricerca visiva** per i segmenti diversamente orientati è un'ottima dimostrazione del ruolo svolto dagli assi cardinali dello spazio visivo. Nella figura 3.6c l'allineamento lungo l'asse verticale rappresenta la norma, che rende facilmente individuabile l'elemento deviante. Ciò indica che le configurazioni percettive sono *organizzate* e possiedono una *struttura gerarchica*, nel senso che le parti più importanti (o semplicemente “portanti”) fungono da riferimento per le altre. Nel dominio dell'orientamento, il ruolo di sistema di riferimento è giocato dagli assi verticale e orizzontale. Una linea collocata sulla verticale o sull'orizzontale viene vista diritta, e ogni piccola deviazione appare come un disturbo da correggere, un'anomalia che in molti suscita la voglia di raddrizzare le cose. Per i vari gradi di obliquità, invece, la nostra percezione è meno selettiva; come dimostra tra l'altro la presenza, in molte lingue, di un unico termine (equivalente all'italiano *obliquo* oppure *inclinato*) per indicare una infinità di direzioni, accanto a termini dedicati alle sole direzioni cardinali (*verticale* e *orizzontale*).

L'ancoramento agli assi cardinali influenza la percezione degli angoli (**figura 3.7**). Mostrate ad alcuni osservatori la configurazione **a**, dicendo più o meno così: «Guarda l'angolo in alto e poi dimmi quale angolo, fra i tre in basso, gli somiglia di più». Tutti sceglieranno l'angolo centrale. Prendete adesso la configurazione **b**, in cui gli stessi angoli presenti in **a** sono stati ruotati in modo da avere il vertice in alto, e ponete la stessa domanda a un altro gruppo di osservatori. Qui le risposte si distribuiranno a caso fra i tre angoli. Il risultato appare logico se si nota che in **b** i tre angoli in basso appaiono molto simili, mentre in **a** è evidente che uno è acuto, uno è l'angolo retto e uno è ottuso. Ma allora dobbiamo concludere che non è sufficiente che un angolo abbia 90 gradi per apparire retto. Perché ciò accada l'angolo deve avere i lati allineati con gli assi cardinali dello spazio visivo. Non è una questione di ampiezza angolare ma di coincidenza dei lati con le direzioni cardinali.



Figura 3.7. In *a* è evidente che l'angolo in alto somiglia di più a quello centrale che non agli altri due. In *b* i tre angoli in basso si somigliano molto: nessuno dei tre appare più simile degli altri all'angolo in alto.

L'effetto illustra una chiara differenza tra **mondo fisico** e **mondo fenomenico**. Nel mondo fisico l'oggetto non cambia solo perché è ruotato. Invece nel mondo fenomenico uno stesso oggetto (l'angolo retto, con un'ampiezza di 90 gradi) può apparire speciale, oppure no, in funzione dell'orientamento. Nella figura 3.7*b* abbiamo a che fare soltanto con segmenti obliqui, i quali non godono di una specifica identità. In quanto tali, differenze nell'ordine dei 3 gradi sono troppo piccole per produrre delle identità distinte lungo il continuum dell'orientamento e quindi dell'ampiezza angolare. Tuttavia, angoli che differiscono anche solo di 3 gradi possono essere facilmente discriminati uno dall'altro se i loro lati sono prossimi agli assi cardinali, che fungono da schema di riferimento mentale per le forme percepite. Quando i lati coincidono con la verticale e l'orizzontale, allora l'angolo retto non solo è ma anche appare "singolare", come chiari in bellissimo saggio sulla somiglianza percepita Erich Goldmeier (1910-1989), allievo di Wertheimer.

3.2. Unificazione/segregazione

Proprio allo psicologo praghese Max Wertheimer (1880-1943) dobbiamo la formulazione di uno dei problemi classici riconducibili all'indeterminazione che grava sulla relazione fra stimoli prossimali e stimoli distali. Due elementi vicini dello stesso oggetto fisico saranno sicuramente vicini anche nell'immagine ottica. Ma non è vero l'inverso: due elementi vicini nell'immagine non necessariamente provengono dallo stesso oggetto fisico. Il **raggruppamento** di elementi o parti dell'immagine in unità organizzate è un fenomeno che, secondo gli psicologi della Gestalt, rivela l'azione di fattori riconducibili al principio di minimo [Riquadro 3.2].

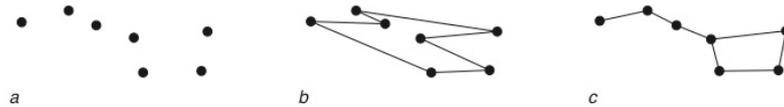


Figura 3.8. Organizzazione percettiva in una costellazione bidimensionale. I punti in *a* non vengono mai unificati nel modo indicato in *b*. La soluzione che si impone spontaneamente è quella in *c*.

Prendiamo in considerazione un'immagine statica contenente elementi che – isolatamente considerati – possiedono soltanto una posizione spaziale. In natura, tale condizione è ben approssimata dalle costellazioni celesti, composte da punti più o meno uguali (**fig. 3.8**). Un importante fattore coinvolto nella percezione delle costellazioni è la **prossimità**: a parità di altre condizioni, vengono unificati gli elementi più vicini. Soluzioni che potrebbero apparire molto sensate (come quella delineata in *b*) vengono rifiutate dal sistema visivo perché implicano l'unificazione di elementi lontani e la segregazione di elementi vicini. Il rapporto con il principio di minimo è chiaro: poiché i raggruppamenti percettivi equivalgono alla creazione di collegamenti virtuali tra elementi, il sistema visivo tende a minimizzare il costo associato a tali collegamenti, che è funzione della loro lunghezza complessiva.

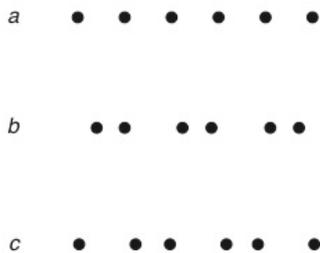


Figura 3.9. La fila di punti in *a* è instabile: l'unificazione/segregazione percettiva è molto influenzata dallo sforzo attentivo. In *b* e *c*, invece, i raggruppamenti sono guidati dalla sinergia tra prossimità e articolazione senza resti (*b*) o dalla prevalenza di un fattore contro l'altro (*c*).

Il fattore della prossimità spiega anche i raggruppamenti osservabili quando gli elementi sono distribuiti lungo una sola dimensione (**figura 3.9**). La fila *a* appare come un tutto unico, eventualmente segmentabile con l'attenzione in più modi; la fila *b* appare invece articolata in tre coppie di punti vicini e la fila *c* in due coppie di punti vicini più due punti isolati agli estremi. Quest'ultimo raggruppamento non è del tutto stabile, poiché comporta dei resti; la tendenza a produrre un'**articolazione senza resti** (altro fattore di unificazione) può far emergere in *c* una soluzione in contrasto con la prossimità, cioè l'unificazione in tre coppie di punti lontani.

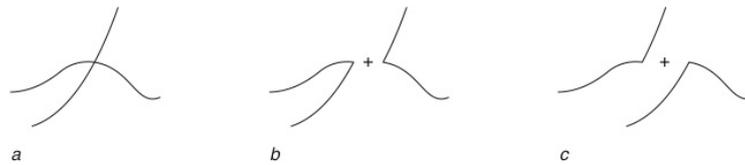


Figura 3.10. La configurazione *a* viene percepita come un intreccio di due linee. Solo con uno sforzo attento è possibile percepire la giustapposizione di due angoli dai lati curvi, come in *bc*.

La **figura 3.10** riguarda l'unificazione dei tratti di linea, elementi più complessi dei singoli punti in quando dotati di proprietà quali lunghezza, orientamento e tipo di curvatura. La percezione dell'incrocio tra due linee, che in *a* si impone spontaneamente, riflette la preferenza del sistema visivo per la **buona continuazione**: cioè per un'unificazione che minimizzi i cambiamenti di direzione. Sotto questo profilo è interessante la **figura 3.11**, che gran parte degli osservatori descrivono come un arco *ac* dal cui centro si diparte il segmento di tangente rettilinea *b*. La possibilità di percepire il segmento *c* come qualcosa che si diparte dalla linea *ab* appare piuttosto innaturale. La buona continuazione include sia l'assenza di discontinuità locali (i bruschi cambi di direzione illustrati nella **figura 3.10bc**) sia la conservazione dello stesso tipo di curvatura lungo l'intera linea (**figura 3.11**).

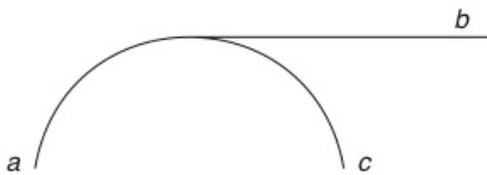


Figura 3.11. La configurazione viene percepita preferibilmente come un arco *ac* con la tangente *b* vissuta come aggiunta, che non come un gancio *ab* con l'aggiunta *c*.

La formazione di unità percettive è un fenomeno molto robusto, che può addirittura mascherare la connessione fisica presente nell'immagine ottica. Per esempio, nella figura 3.12a la prossimità e l'allineamento dei segmenti paralleli fanno sparire i lati obliqui dell'ottagono, che esistono fisicamente nello stimolo prossimale ma vengono "assorbiti" nella retinatura dei quattro quadrati ruotati.

Spesso i fattori di unificazione/segregazione vengono messi uno contro l'altro intenzionalmente, per produrre effetti di instabilità che rendono più attraente una particolare soluzione grafica. La **figura 3.12b** riproduce il marchio di un noto editore internazionale, che con pochi segni rende accattivante e dinamica l'idea di una fila di libri alla quale alcuni (più alti o più bassi) cercano di sottrarsi. La prossimità tra i vari segmenti tende a farli sparire come elementi dotati di una propria identità e li fa percepire semplicemente come il materiale di cui è fatto il rettangolo centrale, al quale i

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15
due segmenti prolungati cercano di sottrarsi.

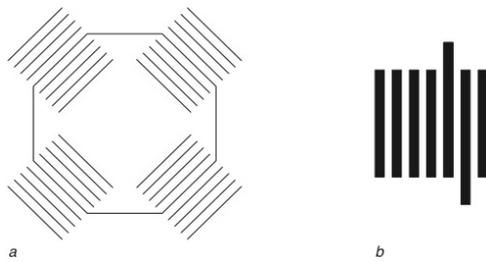


Figura 3.12. In *a* l'assorbimento dei lati obliqui nei reticoli fa sparire l'ottagono, presente nello stimolo ma non nella percezione. In *b* l'instabilità percettiva del marchio di MIT Press deriva dal conflitto tra prossimità e allineamento degli estremi dei segmenti (un aspetto della buona continuazione).

Non possiamo trattare qui gli altri fattori di unificazione. Ma è importante ricordare che quasi tutti sono riconducibili alla tendenza alla semplicità (seguendo l'approccio gestaltista) o all'eliminazione delle interpretazioni meno probabili perché non generali (secondo l'approccio helmholtziano). Il *principio di genericità* (frequentemente utilizzato nelle teorie che si richiamano a Helmholtz) si applica bene alla **figura 3.10**, dove rappresentarsi la configurazione come coppia di angoli con vertice coincidente (le soluzioni *bc*) corrisponde a un'interpretazione generica che non rende conto dell'allineamento dei lati, presente nello stimolo.

3.3. Articolazione figura/sfondo

Nella comune osservazione visiva, gli spazi vuoti tra gli oggetti non vengono notati. Noi vediamo gli oggetti come entità dotate di forma, mentre gli spazi intermedi ne sono privi, salvo nel momento in cui, esercitando un piccolo sforzo, riusciamo a portare la nostra attenzione sui vuoti e a vederli come figure. In natura, questa inversione di ruolo si può ottenere con il profilo del cielo modellato dalle colline; di solito il cielo appare come lo sfondo collocato dietro al paesaggio (colline, alberi, case); ma per l'osservatore esperto non è difficile scambiare intenzionalmente i ruoli di figura e di sfondo, magari facendo emergere nella regione del cielo, vista come figura, un profilo umano. Le inversioni figura/sfondo osservabili in natura sono meglio riproducibili in contesti controllati, utilizzando figure stampate su carta o generate su un monitor. Tutte derivano dalla sostanziale ambiguità dei contorni tra regioni dell'immagine che – anche quando restringiamo l'interpretazione al caso dell'occlusione figura-sfondo – lasciano localmente indeterminato il verso dell'occlusione.

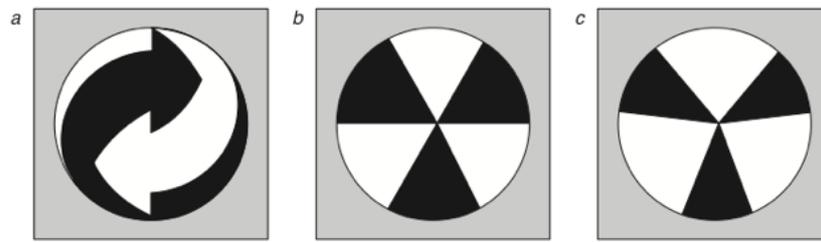


Figura 3.13. Il marchio del riciclo in *a* deve la sua efficacia all'instabilità percettiva derivante dall'alternanza tra due articolazioni figura/sfondo tra loro incompatibili: freccia nera su sfondo bianco e viceversa. La configurazione radiale in *b* è totalmente equilibrata e molto instabile. La configurazione radiale in *c* è più stabile perché prevalgono come figure le regioni nere, di ampiezza minore.

Il problema venne inquadrato da Edgar Rubin (1886-1951), psicologo e fenomenologo danese. Prendiamo uno dei marchi utilizzati per contrassegnare i materiali riciclabili (fig. 3.13*a*), in cui un unico contorno bianco-nero divide il cerchio in due regioni adiacenti: terra e acqua di un globo idealmente bipartito. Poiché queste regioni sono identiche per forma, potremmo pensare che la configurazione sia perfettamente bilanciata e che l'occhio sia indotto a vederle semplicemente come tasselli accostati. Tuttavia, la configurazione è bilanciata a livello globale ma sbilanciata a livello locale. L'andamento del contorno tra regione nera e regione bianca è tale che in alto il nero (la zona che diventa la punta della freccia) è incluso dal bianco, mentre in basso è il bianco a essere incluso dal nero. L'**inclusione** è un potente fattore di articolazione figura/sfondo: a parità di altre condizioni, tende a diventare figura la regione inclusa e a diventare sfondo la regione includente.

Il marchio del riciclo – oltre a invitarci a un comportamento ecosostenibile – ci ricorda un aspetto importante dei processi percettivi di base. Le leggi di organizzazione visiva agiscono in un ambito relativamente locale. Quando dalla loro applicazione emerge un'incompatibilità globale, questa non viene risolta con una rinuncia, abbandonando il campo. Guardando la **figura 3.13*a*** l'occhio non dice: «Beh, evidentemente si tratta di due regioni accostate». Piuttosto mantiene in vita una combinazione di soluzioni locali (freccia nera in alto e freccia bianca in basso) che appare assurda agli occhi della logica, anche se non proprio aliena al pensiero ingenuo (come indicano alcuni fenomeni, altrettanto paradossali, discussi nei **capp. 20 e 21**). Il contorno, lungo la sua traiettoria, delimita in alto la zona nera e in basso la zona bianca: si fa insomma servitor di due padroni, rovesciando la propria funzione in base all'inclusione locale.

Attenzione a non confondere l'inversione della funzione di contenimento del

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15
 contorno con la sua ambivalenza geometrica. A livello geometrico il confine fra due
 regioni appartiene a entrambe. Nell'organizzazione figura/sfondo non è così: a livello
 percettivo il contorno delimita soltanto la figura e ne definisce la forma, mentre lo
 sfondo, privo di contorno, è amorfo. Anche la tendenza alla *funzione unilaterale dei*
bordi è interpretabile come espressione del **principio di minimo**, che nei processi di
 organizzazione percettiva agirebbe come regolatore, spingendo al risparmio sui costi di
 rappresentazione degli oggetti.

I casi *b* e *c* della figura 3.13 illustrano il ruolo dell'area. L'inclusione è
 perfettamente bilanciata: poiché tutti i contorni bianco/nero sono radiali, il bianco è
 incluso nel nero quanto il nero è incluso nel bianco. Ma in *b* i settori hanno ampiezza
 eguale, rendendo equivalenti le soluzioni «bianco su nero» e «nero su bianco»; in *c*,
 invece, prevale la soluzione «nero su bianco» perché – a parità di altre condizioni –
 tendono a essere percepite come figure le regioni di **area minore**.

Il fattore dell'area minore è interpretabile come applicazione del principio di
 minimo, in base al seguente ragionamento. In casi come quelli della **figura 3.13** si
 realizza una *doppia rappresentazione* di alcune regioni, che vengono rappresentate
 come figure ma anche come lo sfondo, dato che questo continua percettivamente dietro
 alla figura, nella forma del **completamento amodale**, un tipo di esistenza percettiva
 non accompagnata dalla qualità *modale* per la visione, il colore. La tendenza delle aree
 minori a diventare figura è coerente con l'ipotesi che il sistema visivo tenda a
 minimizzare i costi relativi alla rappresentazione degli oggetti. Poiché l'articolazione
 figura/sfondo comporta la doppia rappresentazione di alcune regioni, percepite
 modalmente come figure in primo piano e amodalmente come sfondo retrostante, ne
 consegue che il sistema eviterà di rappresentare due volte le regioni più ampie se può
 limitarsi a rappresentare due volte quelle più piccole.

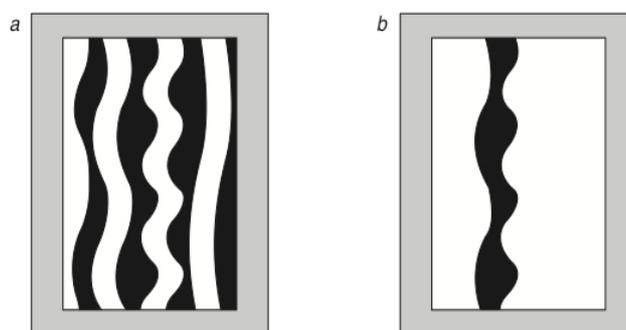


Figura 3.14. In *a* il fattore della larghezza costante fa prevalere le figure bianche
 su sfondo nero. In *b* inclusione e area minore favoriscono la percezione di una
 figura nera.

Il richiamo al principio di minimo è utile anche in altri casi di articolazione figura/sfondo. A una prima occhiata, nella figura 3.14a si vedono delle forme bianche ad andamento sinuoso, su sfondo nero. Se contassero solo l'inclusione e l'area minore, la configurazione sarebbe abbastanza ben bilanciata: bianco e nero si includono a vicenda (salvo per le bande estreme) e hanno un'area circa eguale. Tuttavia, una proprietà formale produce un netto vantaggio per le regioni bianche: la **larghezza costante**. Sempre ragionando sui costi della rappresentazione sottostante, possiamo ipotizzare che codificare una forma regolare, definita da un allargamento costante rispetto a un asse curvilineo centrale, sia meno costoso che codificare una forma di larghezza variabile. Naturalmente si tratta soltanto di un costo relativo. Nella figura 3.13b senza alcuna difficoltà appare come figura la forma nera di larghezza variabile: inclusione e area minore cooperano nel determinare un risultato stabile.

La figura 3.15 illustra un altro effetto della larghezza costante, evidenziato dallo psicologo e filosofo italiano Paolo Bozzi (1930-2003). In teoria, sia a sia b potrebbero corrispondere a un buco quadrato ritagliato in un quadrato nero oppure alla sovrapposizione di un quadrato bianco su un quadrato nero. Eppure l'occhio preferisce vedere in a una cornice nera e in b due quadrati sovrapposti. Niente di strano per chi ormai sa che le regioni a larghezza costante (come la fascia nera in a) tendono a diventare figure. In b la regione nera è irregolare: il che lascia libero gioco all'inclusione, che fa prevalere un quadrato bianco in primo piano, sovrapposto a un quadrato nero più grande. Nella figura 3.15 la stratificazione figurale ha un paio di interessanti conseguenze. Una riguarda il colore della superficie centrale, che quando assume il ruolo di figura in primo piano appare di un bianco più deciso, più brillante. L'altra riguarda la quantità di nero, che appare maggiore in b nonostante l'area nera sia geometricamente invariata, risultando dalla semplice rotazione del quadrato centrale bianco. Ciò indica che l'area percepita dipende non soltanto dalla corrispondente area geometrica, ma anche dall'eventuale completamento amodale, che rende funzionalmente presenti anche parti dell'oggetto non direttamente visibili.



Figura 3.15. La tendenza a vedere come figure le zone a larghezza costante spiega perché in a si vede una cornice nera e in b un quadrato bianco sovrapposto a uno nero. Si noti in b l'illusione di inclinazione già illustrata nella figura 2.16b.

3.4. Completamento amodale e organizzazione figurale

La continuazione dello sfondo dietro alla figura rappresenta soltanto il caso più semplice di completamento amodale. Basta guardarsi intorno per vedere superfici che sembrano continuare dietro ad altre superfici, in un gioco di piani multipli. Consideriamo due altri casi di completamento amodale, che ci permetteranno di trarre alcune conclusioni generali sui processi mediante i quali il sistema visivo costruisce un modello del mondo a partire da immagini ottiche parziali e frammentarie.

La **figura 3.16a** appare come una composizione ordinata di cinque quadrati. Con una certa sorpresa, scopriremo che gli stessi quadrati danno luogo alla percezione di forme molto diverse, nel momento in cui possono unificarsi completandosi dietro a delle figure ocludenti. Il confronto tra le situazioni *b-f* è istruttivo: emergono forme inattese, cui si associa una quantità percepita di grigio incredibilmente diversa. Particolarmente significativa per i processi di riconoscimento (**cap. 4**) è il caso **c**. Grazie alla giustapposizione strategica di tre rettangoli ocludenti, emerge una E parzialmente oclusa. Ma allora, perché non si vede la E anche in *a*, dato che i pezzi modalmente presenti (i cinque quadrati grigi) sono sempre gli stessi?

I fenomeni illustrati nella **figura 3.16** sono coerenti con l'idea generale che la percezione visiva vada tenuta ben distinta dall'immagine ottica, la quale fornisce solo il materiale per la costruzione di una rappresentazione del mondo. Un'immagine con cinque quadrati grigi non garantisce che essi siano percepiti come tali, posto che la percezione dipende anche (e soprattutto) dalle leggi di organizzazione. Quanto accade nella **figura 3.16** è sempre riconducibile alla tendenza alla *funzione unilaterale dei bordi*. Nel momento in cui i rettangoli neri si costituiscono come forme in primo piano, i contorni tra grigio e nero non appartengono più ai quadrati grigi, che restano funzionalmente aperti e disponibili a collegarsi con altri quadrati grigi altrettanto aperti. Il tutto avviene senza l'intervento della *familiarità* con forme note, come dimostra la mancata percezione della E (sagoma ben nota e molto familiare) nella **figura 3.16a**.

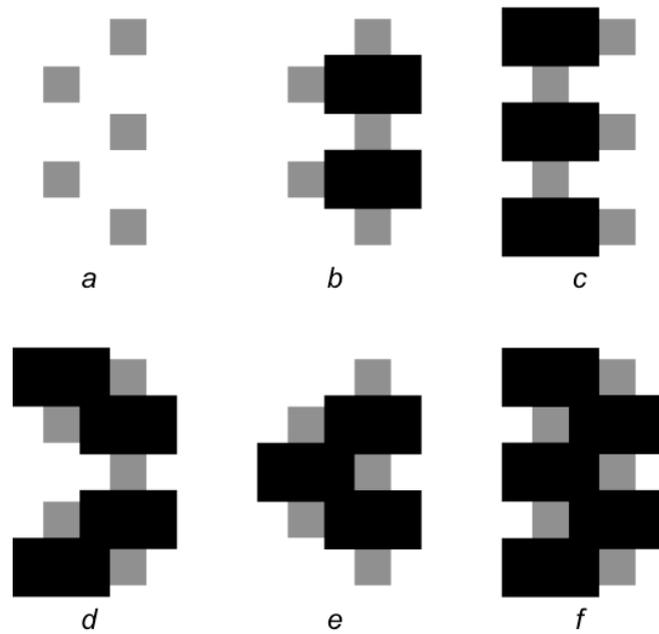


Figura 3.16. Gli stessi quadrati grigi, visti come tali in *a*, danno luogo a forme percettivamente diverse, con una specifica identità e con una crescente quantità apparente di grigio, in funzione del tipo di copertura e quindi di completamento amodale.

La figura 3.16, costruita secondo i principi utilizzati dallo psicologo italiano Gaetano Kanizsa (1913-1993) in molte dimostrazioni analoghe, dimostra l'importanza della disposizione locale di contorni, in particolare delle giunzioni a T. Analizziamo l'andamento dei contorni nei punti in cui si incontrano regioni bianche, nere e grigie. Ciascuna giunzione a T pone un problema di segmentazione che il sistema visivo risolve localmente, assegnando al nero il bordo che nella T corrisponde al segmento superiore e lasciando localmente indeterminata la continuazione amodale del bordo bianco-grigio. I vari frammenti di bordo bianco-grigio vengono poi interpolati o lungo una linea retta o in modo da formare un angolo parzialmente coperto, dando luogo alla presenza amodale di superfici dotate di una forma specifica.

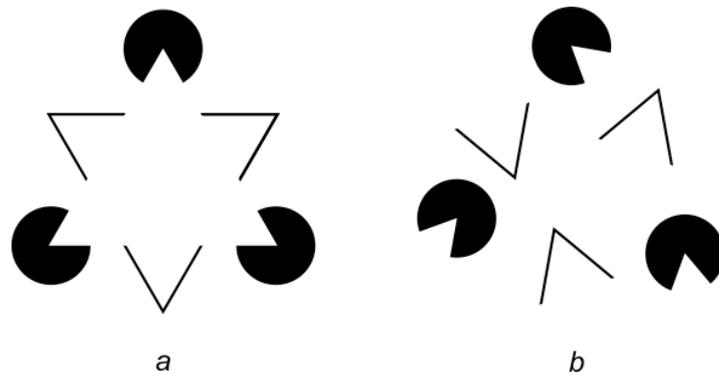


Figura 3.17. Secondo Kanizsa il triangolo più chiaro dello sfondo, visibile in *a*, è il risultato di processi mediati dal completamento amodale dei settori neri e degli angoli a tratto, che grazie alla generazione di una superficie illusoria in primo piano assumono un'organizzazione più semplice del mosaico di sei pezzi tra loro indipendenti; soluzione inevitabile in *b*, dove non sono disponibili gli allineamenti critici.

Qualcosa di simile accade in uno dei fenomeni più clamorosi della scienza della visione, il *triangolo di Kanizsa* (figura 3.17*a*). Secondo la spiegazione proposta originariamente dal suo inventore, il triangolo illusorio deve la sua esistenza alla tendenza al completamento amodale degli inducenti, i tre settori neri e i tre angoli a tratto che diventano tre dischi completi e un unico triangolo chiuso se rappresentati come forme parzialmente coperte. Quando la geometria è sfavorevole – quando cioè mancano gli allineamenti tra le varie porzioni di contorno, come in *b* – il triangolo non emerge. L'illusione trova facile spiegazione anche nella cornice teorica che fa riferimento a Helmholtz: davanti all'alternativa fra sei pezzi con tratti di contorno casualmente allineati e una figura di colore uguale allo sfondo, che occlude quattro forme intere, il cervello helmholtziano-bayesiano opta per la seconda soluzione, ritenuta più probabile concausa dell'immagine in *a* (cap. xx-Zorzi-Girotto su Bayes).

3.5. Struttura dello spazio visivo

I fenomeni e processi descritti nelle sezioni **dalla 3.2 alla 3.4** riguardano il costituirsi degli oggetti percettivi. Nella **sezione 3.1** avevamo però anticipato l'idea che lo spazio fenomenico entro cui trovano posto gli oggetti percettivi non sia equivalente allo spazio geometrico. In modo molto schematico, discuteremo separatamente il campo visivo bidimensionale e la profondità, cioè la distanza dall'osservatore.

3.5.1. Assi cardinali del campo visivo

La figura **3.18**, ideata dal fisico e filosofo viennese Ernst Mach (1838-1916), mostra una delle più semplici e convincenti dimostrazioni della strutturazione dello

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 spazio fenomenico intorno ad assi percettivamente privilegiati. La forma che nell'orientamento normale chiamiamo quadrato assume, se ruotata di 45 gradi, uno aspetto diverso, al punto che molti sono tentati di chiamarla "rombo". Ma allora è giusto dire che è la stessa forma? Non proprio. Dovremmo dire che si tratta della stessa *forma geometrica* ma non della stessa **forma percepita**, distinguendo – per usare i costrutti prima introdotti – tra forma riferita all'osservatore ideale (in grado di rappresentarsi tutte le proprietà geometriche) e forma riferita a un osservatore reale (il quale, a causa dei propri limiti e delle proprie specificità, si rappresenta soltanto una parte di tali proprietà).

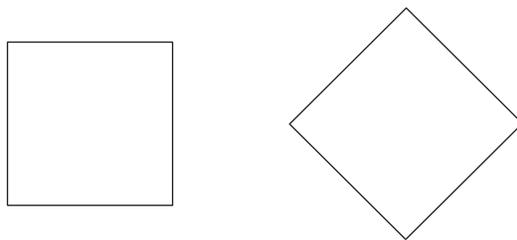


Figura 3.18. L'orientamento spaziale influenza la forma percepita, anche quando la forma geometrica (intesa come insieme di relazioni tra le parti) rimane identica.

Per la geometria, il quadrato ha molte proprietà: quattro angoli retti, quattro lati uguali, diagonali perpendicolari e di uguale lunghezza, angoli opposti uguali, quattro assi di simmetria bilaterale. Il *quadrato di Mach* (così viene chiamato l'effetto illustrato nella **figura 3.18**) dimostra che non tutte queste proprietà sono contemporaneamente rappresentate nella mente dell'osservatore. La forma di sinistra viene chiamata quadrato, ma dal punto di vista percettivo non possiede tutte le proprietà del quadrato geometrico. La presenza di quattro angoli retti è ben evidente, mentre non è altrettanto evidente la perpendicolarità delle diagonali. Al contrario, nella forma di destra è proprio la perpendicolarità delle diagonali a essere saliente, mentre passa sotto silenzio la perpendicolarità degli angoli racchiusi dai lati. La ragione di tale differenza è analoga a quella che ci ha permesso di spiegare la dimostrazione di Goldmeier sulla singolarità dell'angolo retto (**figura 3.7**). La percezione è una rappresentazione dotata di struttura, che non corrisponde alla codificazione di tutti gli aspetti di una determinata configurazione, in modo indipendente dal contesto spaziale in cui essa è inserita; è quindi una rappresentazione selettiva, che esaltando alcuni aspetti a scapito di altri può compromettere l'identità geometrica.

Possiamo discutere la **figura 3.18** anche in maniera un po' diversa. Il quadrato è doppiamente singolare perché, in uno spazio astratto, rappresenta l'intersezione tra due

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15
insiemi di parallelogrammi – quelli con gli angoli retti (i rettangoli) e quelli con le diagonali perpendicolari (i rombi) – all'interno di ciascuno dei quali esso è singolare: fra i rettangoli perché è quello con base e altezza eguali; fra i rombi perché è quello con le diagonali uguali. La figura 3.18 anticipa quanto diremo sugli aspetti **categoriali** dell'esperienza diretta (cap. 4) ed evidenzia il carattere selettivo dell'*appartenenza psicologica*. Il quadrato, che per la geometria è – ribaltando una metafora – «sia carne sia pesce», appartiene percettivamente o all'insieme dei rettangoli o all'insieme dei rombi; mai a entrambi simultaneamente. Come in altri campi della percezione, anche qui sembra operare un principio economico: se una forma viene rappresentata mediante la caratteristica che è distintiva per una data categoria (l'uguaglianza di base e altezza per l'insieme dei rettangoli), allora non serve rappresentarla anche mediante la caratteristica che è distintiva per una diversa categoria (l'uguaglianza delle diagonali per l'insieme dei rombi). La scelta del riferimento categoriale dipende comunque dalla struttura dello spazio fenomenico, che è organizzato intorno alla verticale come principale asse cardinale.

La verticale dello spazio percepito non possiede soltanto un orientamento; possiede anche una **polarità**. A questo proposito il termine «diritto» è rivelatore: in italiano si contrappone sia a obliquo, sia a capovolto. La dicotomia «diritto-obliquo» fa riferimento all'orientamento; mentre la dicotomia «diritto-capovolto» fa riferimento alla polarità sopra-sotto. Le dimostrazioni del ruolo della polarità sopra-sotto sono numerosissime e riguardano quasi sempre forme complesse di cui il sistema visivo fornisce – per così dire – una lettura parziale.

Per esempio, nella figura 3.19 si vedono un cosacco e un asino. Capovolgendo la pagina ci si accorge che si tratta della stessa immagine, che a seconda della polarità assume un significato molto diverso. Prodiggi della percezione, direte voi. Certamente, ma forse anche limiti della percezione. La doppia identità non è rappresentabile contemporaneamente, perché lo schema mentale della testa (sia di un essere umano sia di un animale) è mono-orientato, come dimostrano l'effetto inversione e altri effetti relativi al riconoscimento dei volti, trattati nel cap. 4. La percezione va diritta alla fisionomia, che deriva dalla particolare conformazione delle parti, dando comunque per scontato che gli occhi stiano sopra alla bocca e che tra occhi e bocca ci sia un naso. Capovolgendo la pagina alcuni particolari grafici (come quelli che definiscono gli occhi) mantengono la propria funzione, anche se il cambio di polarità ne muta l'espressione; altri cambiano funzione. Ma nell'insieme il capovolgimento fa emergere una nuova identità.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

Quando una configurazione è influenzata dalla polarità alto/basso ci si deve chiedere se lo schema di riferimento dominante è egocentrico o ambientale. Nella **figura 3.19** a sinistra si vede il cosacco perché la direzione cappello-baffi è coerente con la direzione testa-piedi dell'osservatore oppure perché è coerente con la direzione soffitto-pavimento (ammesso che stiate tenendo la pagina in verticale)? Per dissociare i due schemi di riferimento dovete mettervi a testa in giù, guardando la pagina verticale tra le vostre gambe. Scoprirete che l'orientamento importante è quello egocentrico (o retinico, come a volte si dice).

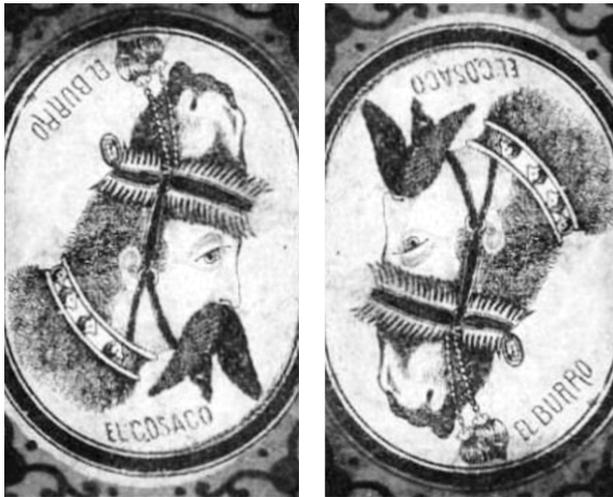


Figura 3.19. La stessa immagine può corrispondere a diverse identità, quando l'oggetto (come nel caso della fisionomia) è mono-orientato, cioè è ancorato a una specifica polarità alto-basso. Capovolgendo la pagina la stessa immagine passa da cosacco a asino, e viceversa.

3.5.2. Informazioni sulla profondità

Come si diceva nella **sezione 3.1**, la tridimensionalità dello spazio percepito rappresenta un problema classico, ricollegabile all'indeterminazione radiale e secondo Berkeley irresolubile sulla base delle sole informazioni ottiche [Riquadro 3.1]. La ricerca ha messo in evidenza che la profondità effettivamente percepita dagli osservatori dipende sia da molteplici informazioni ottiche di livello non elementare sia dall'interazione con informazioni extra-ottiche, prime fra tutte quelle intrinseche al sistema di esplorazione visiva che coinvolge i movimenti degli occhi e della testa. Siccome l'essere umano ha due occhi con campi visivi largamente sovrapposti, è utile suddividere le informazioni sulla profondità in quattro tipi, derivanti dalla combinazione di natura (ottica vs. extra-ottica) e numero (monoculare vs. binoculare) delle fonti. In molti casi è utile anche considerare separatamente le informazioni statiche e dinamiche.

Nella normale esplorazione visiva dell'ambiente da parte di un osservatore in movimento e con due occhi, le informazioni sulla tridimensionalità sono ricche e convergenti, e il problema scientifico è quello della loro integrazione. Il sistema visivo le usa tutte o ne usa solo alcune? E come combina i vari contributi? In generale, ciascuna delle informazioni che passeremo rapidamente in rassegna costituisce una condizione sufficiente ma non necessaria per la visione tridimensionale. Stabilirne il peso relativo non è facile: azzerando tutte le fonti di informazione sulla profondità eccetto una – quella che si vuole studiare in isolamento – non si eliminano le informazioni azzerate; esse specificano l'assenza di profondità, fungendo da indizi di piattezza (*cue to flatness*).

Informazione ottica monoculare

Rientrano in questa categoria i molti accorgimenti pittorici scoperti e utilizzati per generare uno spazio 3D, in modo più o meno efficace, manipolando le proprietà geometriche e fotometriche dell'immagine. Il paradosso della percezione pittorica sta nel fatto che l'osservatore tratta l'immagine artificiale – dipinta o catturata con una macchina fotografica – come un surrogato dello stimolo distale e non dello stimolo prossimale, a dispetto dell'ammonimento “*Ceci n'est pas une pipe*” del pittore surrealista belga René Magritte (1898-1967) apposto sotto l'accurato ritratto di una pipa (**figura 3.20**).



Figura 3.20. “Il tradimento delle immagini” di René Magritte, dipinto nel 1929.

In altri termini, la percezione visiva ignora il suo fondamento diretto (l'informazione retinica), a favore della presunta causa originaria (l'oggetto esterno). Ciò è dimostrato molto efficacemente dall'*illusione dei parallelogrammi* inventata da Roger Shepard all'inizio degli anni Ottanta (**figura 3.21**). Viene chiamata così una curiosa immagine che evoca una scena 3D sulla base di poche ma efficacissime informazioni ottiche monoculari: l'obliquità delle linee, che tende a farle percepire inclinate in profondità (mentre le linee verticali e orizzontali appaiono collocate sul piano frontale); la posizione rispetto all'orizzonte, che fa vedere gli elementi più bassi sul piano pittorico come vicini all'osservatore e quelli più alti come lontani; la tendenza degli angoli acuti e ottusi, contenuti per esempio nelle giunzioni a Y, ad apparire retti. Questi semplici indizi suggeriscono una scena contenente tre blocchi poggiati su un piano che recede in profondità. Ciò che l'occhio non riesce proprio a vedere è che i due parallelogrammi in basso (le proiezioni delle facce superiori di due blocchi diversamente orientati) sono identici, ma ruotati. Quando prevale un'organizzazione tridimensionale – e non ci vuole molto perché ciò accada – le relazioni geometriche disponibili nell'immagine pittorica e nell'immagine retinica risultano inaccessibili.

Per capire perché il blocco a sinistra in basso **nella figura 3.21** appaia così lungo e stretto si consideri il suo lato corto: essendo orizzontale viene visto sul piano frontale, con la sua lunghezza effettiva. Ma nel blocco a destra lo stesso lato è obliquo, e quindi viene percepito come più lungo in base all'invarianza grandezza/distanza (una stessa proiezione retinica, se percepita come un segmento esteso in profondità, apparirà comunque più lunga rispetto a quando appare come un segmento frontale). Viceversa, il lato lungo del blocco a sinistra in basso, percepito come inclinato in profondità, appare molto più lungo del suo equivalente frontale nel blocco a destra.

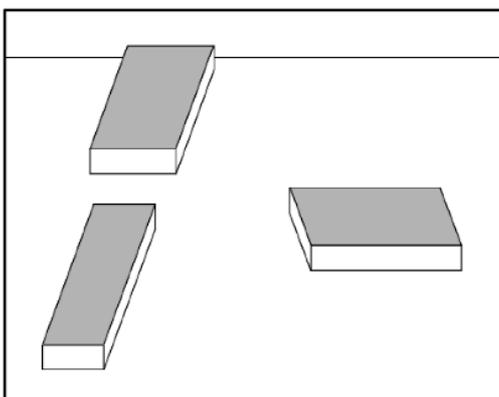


Figura 3.21. Guardate i due blocchi a sinistra. Quale dei due somiglia di più, per forma, al blocco di destra? Tracciate il contorno della faccia superiore del blocco di destra su un pezzo di carta trasparente e scoprirete che si

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

sovrappone esattamente al blocco di sinistra *in basso*, quello che sembra tanto lungo e stretto.

Particolarmente efficaci nel determinare l'inclinazione 3D sono i *gradienti prospettici* relativi a grandezza, forma e densità microstrutturale, informazioni che normalmente covariano nelle proiezioni che l'osservatore ottiene da una superficie come il terreno o il pavimento di un interno. Gioca un ruolo importante anche il gradiente di chiaroscuro (*shading*), cioè la variazione graduale di luminanza dipendente dall'angolo di incidenza dell'illuminazione direzionale sulle superfici.

La **Figura 3.22** dimostra che l'informazione disponibile nel *chiaroscuro* – una tecnica pittorica molto utilizzata per conferire rilievo 3D alla rappresentazione degli oggetti – è in realtà incompleta, nel senso che può specificare la quantità di curvatura locale di una superficie, ma non il verso concavo/convesso. Ciascuno dei 25 elementi della configurazione è un disco con il medesimo gradiente di chiaroscuro compatibile con una superficie concava/convessa illuminata dall'alto o dal basso. La percezione mostra una netta preferenza per il rilievo coerente con l'illuminazione proveniente dall'alto: appaiono quindi come sfere i dischi che hanno la parte chiara verso l'alto. Come nella Figura 3.19, anche in questo caso conta lo schema di riferimento egocentrico. Ma qui la cosa è ben più sorprendente, giacché la preferenza per l'illuminazione proveniente dall'alto potrebbe essere facilmente associata a un fattore ambientale. Quindi avremmo potuto attenderci di vedere sempre la lettera H formata da sfere in rilievo, anche quando ci mettiamo con la testa in giù, tenendo la pagina normalmente orientata rispetto all'ambiente. E invece no: per la mente l'alto che conta – per l'illuminazione come per gli asini e i cosacchi – è quello dello spazio interno, che gira con la nostra testa.

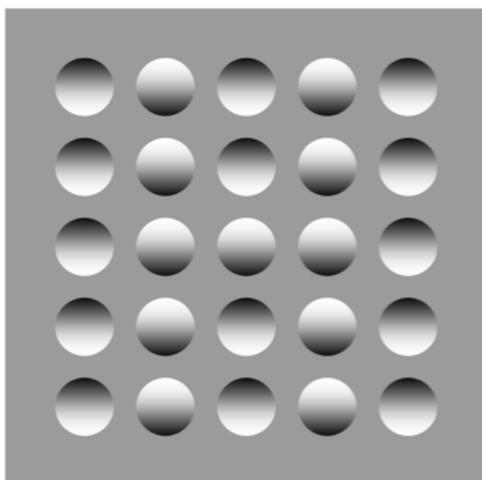


Figura 3.22. Nell'orientamento normale si vede una H composta da sfere sospese, risultanti dal completamento amodale di superfici convesse (rispetto allo sguardo), coerenti con un'illuminazione proveniente dall'alto. Se si capovolge la pagina la H risulta composta da incavature concave. Interessante anche osservare la figura con la pagina ruotata di 90 gradi: la polarità alto-basso del chiaroscuro non ha un equivalente nella polarità sinistra-destra.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

I gradienti prospettici e di chiaroscuro sono informazioni monoculari di tipo metrico, nel senso che il loro valore specifica il grado di inclinazione in profondità. Un'informazione puramente ordinale è invece quella veicolata dalla forma delle regioni dell'immagine e dall'andamento dei contorni, fattori che (come abbiamo discusso nella sezione 3.4) portano alcune regioni a completarsi dietro ad altre. Il completamento amodale è quindi il meccanismo che media l'*interposizione*, cioè la relazione davanti/dietro fra superfici, anche in assenza d'informazioni sull'entità della loro distanza relativa.

Appartiene sempre alle informazioni ottiche monoculari la *prospettiva di movimento* e altre informazioni dinamiche ad essa collegate. Quando l'osservatore si muove rispetto al mondo statico, i cambiamenti delle proiezioni ottiche trasmettono informazione sulla distanza degli oggetti dal punto di vista. Il cinema di animazione, anche quello più semplice, fa ampio uso della ricchezza informativa del movimento relativo tra parti dell'immagine per evocare uno scenario 3D. La Figura 3.23 mostra una sequenza di 3 fotogrammi tratti dall'episodio "The New Car" di Peppa Pig. Con l'aiuto della linea bianca verticale si nota che nei tre fotogrammi l'auto occupa la stessa posizione, simulando uno spostamento del punto di vista solidale con l'auto; la prima collina si sposta molto verso sinistra, in direzione opposta al movimento simulato dell'auto; la seconda collina si sposta verso sinistra ma molto meno, simulando una maggiore lontananza dalla strada; la nuvola si sposta di pochissimo, sempre verso sinistra, simulando un oggetto quasi all'infinito. Durante l'osservazione del filmato, l'insieme di questi movimenti relativi – coerente con le informazioni statiche sulla stratificazione fra la prima e la seconda collina, disponibili in ogni fotogramma – produce una convincente percezione di tridimensionalità.

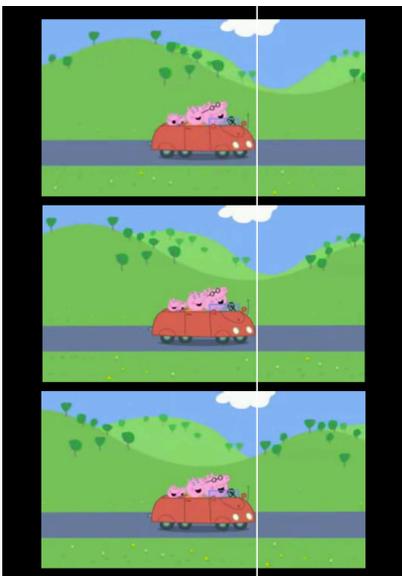


Figura 3.23. Osservate la posizione relativa delle due colline e della nuvola nei tre fotogrammi. Guardando "The New Car" in rete vi convincerete dell'efficacia della prospettiva di movimento quale informazione 3D.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

Informazione ottica binoculare

Avere due occhi con campi visivi sovrapposti permette a molti animali (umani inclusi) di usare le disparità tra le immagini proiettate in due punti di vista diversi come fonte d'informazione sulla profondità. La dimostrazione sperimentale dell'efficacia delle *disparità binoculari* è relativamente recente. Si dovette attendere il 1838 e l'invenzione dello stereoscopio da parte di Sir Charles Wheatstone (1802-1875) per capire che coppie d'immagini piatte (gli stereogrammi) ma contenenti piccole differenze di dislocazione orizzontale potevano costituire uno stimolo sufficiente a evocare una suggestiva esperienza di profondità, la *stereopsi binoculare*.

Provate a tenere due matite verticali davanti a voi, a distanze diverse dal naso. Chiudendo alternativamente un occhio e l'altro vi renderete conto che in ciascuna immagine monoculare la posizione relativa delle due matite è diversa. Se è più vicina la matita di sinistra allora la separazione tra le proiezioni delle due matite sarà maggiore nell'occhio destro, e viceversa. Questa in sostanza è l'informazione utilizzata in tutti i dispositivi di visione stereo.

Informazione extra-ottica monoculare

Questo tipo di informazione è esemplificato dall'accomodazione, il meccanismo che fino a una certa età consente di variare la curvatura del cristallino e di mettere a fuoco oggetti distanti da qualche decina di centimetri a un paio di metri. Spostare la messa a fuoco da un punto vicino a uno lontano comporta il rilassamento dei muscoli che controllano il cristallino, mentre lo spostamento opposto (da un punto lontano a uno vicino) comporta una contrazione. In condizioni dinamiche, quindi, l'accomodazione fornisce un'informazione attendibile sulla profondità relativa di due punti, a condizione che almeno uno dei due si trovi entro due metri di distanza dall'osservatore.

Informazione extra-ottica binoculare

L'angolo di vergenza e soprattutto le sue variazioni, convergenza e divergenza, forniscono un segnale relativo alla distanza del punto in cui si incrociano gli assi ottici relativi ai due occhi (il vertice dell'angolo di vergenza). L'angolo di vergenza è nullo per una mira all'infinito; è pari a circa 18 gradi se la mira binoculare è a 20 cm. Lo sforzo muscolare connesso con la convergenza (l'aumento dell'angolo di vergenza) fornisce un segnale univoco sulla riduzione di distanza dall'osservatore.

3.6. Epilogo

In questo capitolo abbiamo definito alcuni concetti generali (informazione, codificazione, organizzazione, catena psicofisica, rappresentazione), applicandoli ai fenomeni percettivi fondamentali: la salienza preattentiva di alcuni oggetti in mezzo ad altri, l'unificazione/segregazione tra parti dell'input ottico, l'articolazione figura/sfondo, il completamento amodale, la struttura dello spazio visivo.

L'esposizione è stata guidata dall'idea che il sistema visivo fornisce all'organismo una *rappresentazione* del mondo (basata sulle immagini ottiche, ma sostanzialmente diversa da queste) che non dipende dai processi cognitivi intenzionali. Ci siamo quindi occupati soltanto dei processi percettivi di base, preattentivi e in larga misura estranei all'influenza dell'esperienza passata, rinviando ad altri capitoli (anzitutto quello sul riconoscimento) il rapporto con i cosiddetti processi cognitivi superiori.

Va anche detto che la disponibilità di una conoscenza diretta (la percezione) è il presupposto per lo sviluppo di una conoscenza indiretta (il pensiero), consistente in concezioni anche molto lontane dall'esperienza sensibile. Studiare percezione aiuta a comprendere la complessa e talora contraddittoria fenomenologia dell'esperienza diretta; ma anche a sviluppare la consapevolezza che i processi percettivi consegnano all'osservatore una rappresentazione del mondo esterno non del tutto veridica, anche se sufficientemente adeguata alle necessità fondamentali dell'osservatore, come il controllo dell'azione e la comunicazione con i conspecifici.

Riquadro 3.1. Indeterminazione radiale.



George Berkeley (1685-1753)

”Tutti, credo, sono concordi nel ritenere che la Distanza, di per sé e immediatamente, non possa essere vista. Poiché la Distanza, essendo una Linea che punta in direzione dell’Occhio, proietta solo un Punto sul Fondo dell’Occhio. Punto che rimane invariabilmente lo stesso, indipendentemente dalla maggiore o minore Distanza.” In questo passo del *Saggio su una nuova teoria della visione* (1709) Berkeley demolisce l’idea che il senso della profondità – cioè il vissuto di distanza dall’osservatore che accompagna la percezione di ogni oggetto e quindi, in definitiva, l’intera struttura dello spazio tridimensionale – possa essere un prodotto genuinamente visivo, considerato che per la profondità non ci può essere uno stimolo locale sulla retina. La

stella da cui proviene un dato raggio di luce può essere vicina o lontana. L’indeterminazione radiale deriva appunto dall’impossibilità di determinare la distanza dello stimolo distale dal punto di vista, sulla base dell’immagine ottica elementare.

Berkeley propose una soluzione extra-ottica per la profondità visiva, che a suo parere sarebbe importata dal tatto attivo, cioè dalla coordinazione fra input retinici 2D e azioni esplorative compiute dal corpo nel mondo 3D. In verità, analizzando l’informazione ottica a un livello non elementare si scoprono molte informazioni utili alla ricostruzione della struttura degli oggetti distali. Ciononostante l’idea di Berkeley è ancora ben presente nell’ambito della *visione attiva*, che valuta il contributo delle azioni esplorative alla percezione.

Riquadro 3.2. Teorie della percezione.



Hermann von Helmholtz

Un’autorevole soluzione ai problemi posti dall’indeterminazione ottica venne fornita dal fisico, matematico e fisiologo tedesco Hermann von Helmholtz (1821-1894), il quale ipotizzò che la percezione sia il risultato di giudizi inconsci che porterebbero l’osservatore a scegliere – tra quelli possibili – l’oggetto esterno che più verosimilmente rappresenta la causa del dato sensoriale parzialmente indeterminato. L’**approccio probabilistico** helmholtziano si basa sull’esperienza passata dell’osservatore e, nella ricerca attuale, si pone spesso a fondamento dei modelli bayesiani della percezione [cap. xx].

Il presupposto dell’indeterminazione ottica è condiviso dalla teoria della Gestalt, sviluppata da studiosi – in primo luogo Max Wertheimer (1880-1943), Wolfgang Köhler (1887-1967) e Kurt Koffka (1886-1941) – che si occuparono in modo specifico di percezione. Anche nei suoi sviluppi più recenti la teoria della Gestalt (termine tedesco normalmente tradotto con «forma») si richiama al **principio di minimo**, un’idea guida presente in molti ambiti scientifici, secondo la quale l’organizzazione percettiva tenderebbe all’uso del minima quantità di risorse mentali (o di costi di elaborazione neurale), nelle condizioni date. Per i gestaltisti la chiave per il superamento dell’indeterminazione ottica è quindi la *tendenza alla semplicità*: la preferenza per le soluzioni semplici e ordinate deriverebbe dall’auto-organizzazione dei sistemi percettivi e sarebbe relativamente indipendente dall’esperienza passata dell’osservatore.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

I fenomeni centrali per l'approccio helmholtziano e per quello gestaltista sono spesso gli stessi, e non mancano i tentativi di integrazione, basati sulla constatazione che, in molti casi, ciò che più probabile nell'ambiente è anche strutturalmente più semplice. In questo campo, costruire un esperimento cruciale non è affatto facile.

Una vivace critica a tutte le teorie focalizzate sull'indeterminazione ottica, e volte a superarla, è venuta dall'**approccio ecologico** alla percezione, proposto dallo psicologo americano James J. Gibson (1904-1979). Gibson riteneva che la ricerca si fosse concentrata su situazioni iper-semplificate e poco rappresentative, mentre in condizioni normali i sistemi percettivi hanno a disposizione informazione sovrabbondante, piuttosto che indeterminata. Il serbatoio di informazione per la visione sarebbe il *flusso ottico*, termine introdotto da Gibson per indicare gli aspetti statici e dinamici delle immagini raccolte nel punto di vista in movimento, ritenuti sufficienti a spiegare la struttura percepita dell'ambiente e degli oggetti. All'osservatore resterebbe il compito di apprendere, interagendo con il mondo esterno, come sfruttare al meglio il sistema visivo, estraendo il massimo di informazione dal serbatoio disponibile.

Riquadro 3.3. Informazione ottica e *Ganzfeld*.



La percezione della nebbia è riproducibile in laboratorio senza riempire una stanza di vapore acqueo. Per ottenere il *Ganzfeld* (in tedesco «campo totale») – condizione limite in cui l'energia luminosa che arriva all'occhio è identica in tutte le direzioni – lo psicologo tedesco Wolfgang Metzger (1899-1979) poneva l'osservatore a poco più di un metro da un'ampia parete liscia e omogenea, illuminata in modo uniforme; in alternativa si può chiedere

all'osservatore di guardare dentro una semisfera omogenea, o di coprirsi ciascun occhio con mezza pallina di ping-pong o con un cucchiaino di plastica (fatti di materiale che lascia passare la luce ma ne distrugge la struttura). Indipendentemente dal metodo usato, l'osservatore si sentirà immerso in un ambiente lattiginoso, in cui lo sguardo sembra penetrare al massimo fino a 2-3 metri. Se la luce viene colorata, per esempio collocando un filtro rosso tra la lampada e i cucchiaini sugli occhi) la tonalità cromatica tenderà a svanire e, dopo qualche minuto di immersione nel *Ganzfeld*, l'osservatore si sente immerso in una nebbia acromatica.

4. Riconoscimento

4.1. Un concetto eterogeneo

Il riconoscimento è un po' come il *Boletus luridus*, commestibile solo cotto. In questo capitolo cercheremo di portare a giusta cottura il concetto, per evitare i danni che deriverebbero da un'ingestione affrettata.

Nel linguaggio quotidiano, parole come “riconoscere” e “riconoscimento” ricorrono con sfumature variamente collegabili a competenze date quasi per scontate, per la loro automaticità e fluidità. Ritrovare persone e luoghi noti, avvertire che un'azione è collaudata, leggere (quello che state facendo adesso, cercando di catturare i significati senza badare a come vengono elaborati i segni) sono attività su cui non ci soffermiamo a ragionare, salvo quando ci troviamo a interagire con una persona che una malattia o un incidente hanno privato, appunto, della capacità di riconoscere. L'analisi dei vari tipi di *agnosia* (si veda il cap. XX, sul linguaggio) – inclusa la *prosopagnosia*, cioè l'incapacità di riconoscere i volti di persone note – fornisce un importante contributo alla costruzione di modelli dell'esperienza diretta degli osservatori.

In più ambiti scientifici (dalla microbiologia all'ingegneria informatica) è diffusa l'espressione *pattern recognition*, fortunata combinazione di due termini ombrello forse solo metaforicamente collegata ai fenomeni mentali di nostro interesse, che sono sostanzialmente due: (i) il vissuto di familiarità (alcuni oggetti/eventi appaiono come “ri-presentazioni” di oggetti/eventi già incontrati, contrapposti al mero “presentarsi” di altri) che si produce quando i processi percettivi di base entrano in contatto con la memoria (cap. 8 e altri?); (ii) gli effetti sperimentali che si producono anche in assenza del vissuto di familiarità, ma che vanno attribuiti all'influenza della memoria sull'esperienza diretta.

Oggetto di riconoscimento può essere qualunque tipo di materiale: fonemi (per il linguaggio orale), grafemi (per il linguaggio scritto), immagini 2D, oggetti 3D, volti, andature, scene naturali, tracce di animali, manufatti umani. Ma le immagini pittoriche hanno, per ragioni che ora discuteremo, un'importanza speciale.

4.1.1. “Vedere come”

Nel cap. 3 abbiamo toccato un terreno tradizionalmente considerato di confine fra percezione e cognizione, e quindi particolarmente fertile per il concetto di riconoscimento: la percezione pittorica. Le figure 3.19 e 3.20 riguardavano il cosiddetto “vedere come” (*seeing as*), in cui l'osservatore è portato a distinguere fra i *dati* che

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 arrivano ai suoi occhi, nella forma di materiale grezzo plasmabile in più modi, e le loro *interpretazioni* alternative. La percezione pittorica è ideale per lo studio del “vedere come” perché l'immagine, che nella catena psicofisica della visione naturale è un costrutto geometrico, un po' difficile da afferrare, nella percezione pittorica è un insieme di tracce materiali collocate su una superficie tangibile.

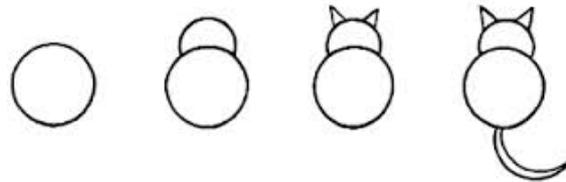


Figura 4.1. Un gioco grafico discusso da Gombrich. Il cerchio a sinistra può essere visto come una fetta di pane; la prima aggiunta lo trasforma in una borsa della spesa, al cui manico possiamo aggiungere dei particolari, trasformandola in una borsetta. Ma l'ultima aggiunta produce una radicale metamorfosi: improvvisamente compare un gatto.

Per Sir Ernst Gombrich (1909–2001), grande storico dell'arte e studioso della percezione visiva, l'improvvisa comparsa di un gatto in uno schizzo a tratto era un mistero che l'aveva accompagnato fin da bambino. Gombrich si riferiva a un giochino grafico (**figura 4.1**), in cui la scoperta visiva del gatto chiude un processo “per aggiunta di parti” che produce un'autentica metamorfosi dell'esperienza diretta. La linea che nel secondo stadio è il manico di una borsa diventa, dopo l'aggiunta finale della coda, il contorno della testa di un micino di spalle. Ciò che giustamente sorprendeva Gombrich è la rapidità con cui la mente s'impossessa di pochi tratti dando vita a un oggetto dotato di significato, che coesiste con la consapevolezza di avere davanti un materiale, il foglio tracciato, diversissimo da un vero gatto con peli e fusa.

Il “vedere come” va comunque distinto dal riconoscimento degli oggetti normali, che non raffigurano altri oggetti. Giustamente viene trattato così, in ambito evolutivo o comparato, quando ci si chiede se riconoscere oggetti comuni nelle raffigurazioni – che sono un prodotto della cultura umana – richieda un apprendimento paragonabile a quello che ci porta, nei primi anni di vita, a leggere le parole scritte. Avendo in mente l'analogia tra il linguaggio delle immagini e linguaggio verbale, Julian E. Hochberg e Virginia Brooks (a quel tempo alla Cornell University) condussero un esperimento classico, pubblicato nel 1962, allevando un bambino in un mondo senza immagini fino all'età di 19 mesi. La coppia di ricercatori voleva verificare se – al termine dell'inusuale periodo di deprivazione – un bambino esposto a semplici raffigurazioni fosse comunque in grado di riconoscere gli oggetti raffigurati. La procedura di deprivazione fu interrotta

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 quando il piccolo aveva 19 mesi, perché era diventato praticamente impossibile evitare che questi denominasse gli oggetti visti in immagini incontrate per caso. A quel punto, una prova controllata di denominazione dimostrò che il riconoscimento di oggetti in raffigurazioni mai viste prima avveniva in modo spontaneo. Hochberg e Brooks conclusero che almeno alcune componenti dei processi sottostanti il riconoscimento di oggetti 3D nei disegni a tratto e nelle fotografie sono innate (Figura 4.2).

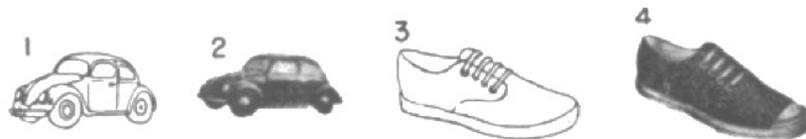


Figura 4.2. Alcune immagini usate da Hochberg e Brooks per verificare se un bambino di 19 mesi, deprivato di ogni esperienza con materiale pittorico, riconosceva gli oggetti raffigurati. Anche i disegni a tratto, presentati sempre prima delle fotografie, furono denominati con successo.

La conclusione è importante, ma rinvia alla natura di queste abilità, che potrebbero essere comuni all'elaborazione di stimoli artificiali (disegni, dipinti, fotografie) e stimoli naturali; quelli che Gombrich, nel saggio *Arte e illusione* (1960), aveva descritto come casi di "immagine nelle nuvole", cioè di falso riconoscimento.

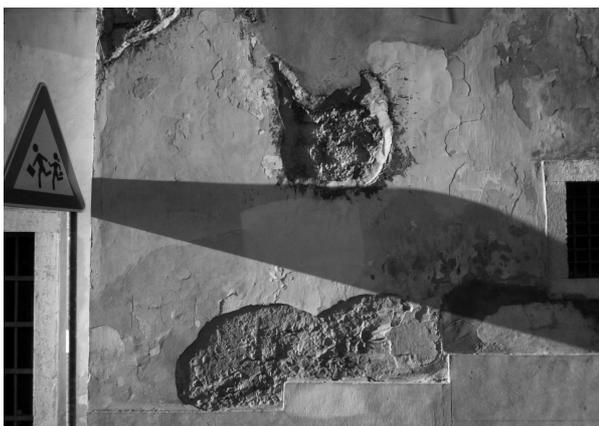


Figura 4.3. Un muro rappezzato in cui un bambino di 21 mesi spontaneamente riconosce un gatto.

Era un tardo pomeriggio estivo e portavo a cavalluccio lungo le strade silenziose del centro storico la mia piccola Emma, che allora aveva 21 mesi. Improvvisamente la sentii gridare contenta "Ioio!", una delle poche parole del suo vocabolario, usata per altro in modo assai accurato per indicare un gatto (grande o piccolo, vivo o di peluche, stampato sulle magliette o su qualche accessorio, ma gatto). Non c'erano gatti vivi in giro ed Emma insisteva con "Ioio", mostrando il muro della casa che stavamo

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 costeggiando (Figura 4.3). Alla fine il gatto sul muro lo vidi pure io: anche la mia mente si era impossessata di un pezzo di muro sbrecciato, malamente coperto di malta grezza, e l'aveva trasformato in una testa tonda con due inconfondibili orecchie da gatto.

La comparsa di oggetti significativi nelle nuvole, nelle macchie di umidità sui muri o nella corteccia dei tronchi, l'emergere di frammenti di parlato nel rumore o nelle registrazioni riprodotte all'indietro, e altri fenomeni simili sono casi di *pareidolia*, termine che suona come una sorta di rimprovero mosso ai meccanismi di riconoscimento, colpevoli di aver prodotto un esito positivo in circostanze inappropriate. Protagoniste della pareidolia sono soprattutto le facce, in particolare se attribuibili a entità sovranaturali, riconosciute nelle circostanze più improbabili; ma la possibilità di oggettivare le forme più diverse va ben oltre le facce.

Sentiamo in proposito la testimonianza di Leonardo da Vinci (1452-1519), nel *Trattato della pittura*: “E questo è: se tu riguarderai in alcuni muri imbrattati di varie macchie o pietre di vari misti, se arai a inventionare qualche sito, potrai lì vedere similitudine de' diversi paesi, ornati di montagnie, fiumi, sassi, albori, pianure, grandi valli e colli in diversi modi; ancora vi potrai vedere diverse battaglie e atti pronti di figure, strane arie di volti e abiti e infinite cose, le quali tu potrai ridurre in integra e bona forma. E interviene in simili muri e misti come del sono di campane, che ne' loro tocchi vi troverai ogni nome e vocabulo che tu imaginerai”. In sostanza, Leonardo raccomandava l'esercizio della pareidolia come antecedente della pittura, legando in modo molto stretto percezione pittorica e produzione pittorica, nella consapevolezza – ben presente in chi lavora con le immagini pittoriche – che la distinzione fra riconoscimento vero e falso è assai problematica (e forse inutile).

Nella figura 4.4 tutti vedono il simpatico volto di Albert Einstein (1879-1955), anche se la qualità della foto potrebbe dirsi scadente. Se qualcuno vi dicesse che “in realtà” sulla pagina ci sono soltanto 3600 quadrati, ciascuno con un diverso livello di grigio, e che riconoscervi un volto (e in particolare quello dello spirito geniale di Einstein) è un caso di pareidolia, trovereste la tesi piuttosto bizzarra. Dove finisce il “vedere” e dove comincia il “vedere come”?

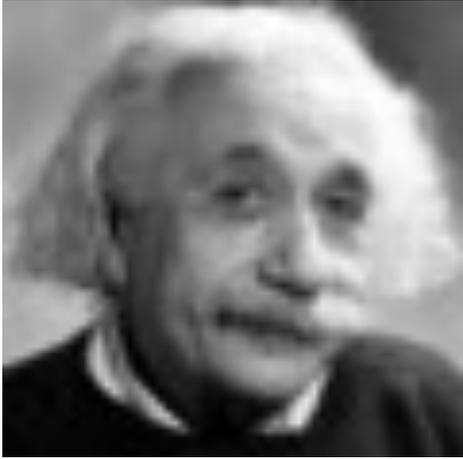


Figura 4.4. Albert Einstein in un'immagine contenente "soltanto" 60×60 quadrati con diversi livelli di grigio.

Lasciando aperta la difficile questione del confine tra "vedere" e "vedere come", chiariamo perché tutta la percezione pittorica può essere considerata un caso di riconoscimento. Secondo un'ipotesi molto diffusa, la percezione pittorica implica una *doppia consapevolezza*: (i) quella relativa a una superficie materiale – per esempio la pagina del libro – percepita nel mondo reale; (ii) quella relativa all'oggetto raffigurato, percepito come virtualmente presente. Quando si realizza la consapevolezza (ii) – che potrebbe non realizzarsi se l'osservatore fosse incapace di leggere le immagini e rimanesse al livello (i) – l'immagine, prima "conosciuta" solo come superficie materiale, viene "ri-conosciuta". La pareidolia è interessante, a questo proposito, poiché dimostra che l'oggetto raffigurato, virtualmente presente, emerge come risultato di una riorganizzazione percettiva tipicamente irreversibile (figura 4.5).



Figura 4.5. Riconoscete qualcosa? Se non trovate nulla capovolgete la pagina. Se ora avete riconosciuto ciò che prima era nascosto, guardate di nuovo la pagina dritta: l'immagine non vi apparirà più priva di senso come all'inizio. Qualora l'immagine vi sembri ancora priva di senso, leggete la chiusa di questo capitolo.

4.1.2. Ritrovare il già noto

In psicologia sperimentale, un'importante *definizione operativa* della prestazione di riconoscimento è quella fornita dal paradigma *old/new* (vecchio/nuovo), che prevede due fasi separate da un intervallo temporale variabile da pochi secondi (riconoscimento immediato) a ore o giorni (riconoscimento differito). Nella fase di

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 apprendimento/addestramento lo sperimentatore presenta al partecipante una serie d'immagini o parole da mantenere in memoria, controllando il tempo di esposizione, la complessità del materiale e altri fattori che influenzano la codifica dell'informazione in entrata. In fase di test il partecipante esegue molte prove, in ciascuna delle quali lo sperimentatore presenta una coppia di stimoli, uno vecchio (incluso nella fase di apprendimento) e uno nuovo, bilanciandone la posizione sinistra/destra o prima/dopo. Il partecipante sa che in ogni prova c'è sempre uno stimolo *old* e deve indicarlo allo sperimentatore fornendo la risposta "sinistra/destra" o "prima/dopo", come previsto dal compito 2AFC [Riquadro 2.6].

In un famoso lavoro, intitolato *Learning 10000 pictures*, pubblicato nel 1973, lo psicologo canadese Lionel Standing ottenne dai suoi partecipanti, due giorni dopo l'esposizione a un blocco di 1000 stimoli, il seguente numero di errori (scelte dello stimolo *new* nel compito 2AFC): 5 errori per le immagini vivide (per esempio, un cane con la pipa in bocca); 9 errori per le immagini normali; 15 errori per le parole. Per ciascuna delle tre categorie di stimoli la piccola proporzione di errori rimaneva costante al variare del numero di immagini apprese (da 20 a 1000 per blocco, in più blocchi comprendenti complessivamente oltre 10 mila immagini): un risultato sorprendente, che portò Standing a concludere che la nostra capacità di riconoscere immagini pittoriche è praticamente illimitata, non soltanto migliore di quella per il materiale verbale.

Questa straordinaria prestazione nel paradigma *old/new*, che operationalizza il riconoscimento, non va confusa con il vissuto di familiarità che, per l'analisi fenomenologica, rende distintivo l'atto del riconoscere qualcosa. Un'elevata proporzione di scelte corrette operate da un essere umano o da un animale (cap. 18) è compatibile con vari processi, le cui differenze verranno illustrate nella sezione 4.3.

4.1.3. Illusioni di riconoscimento

I falsi ricordi di materiale verbale includono un'illusione di riconoscimento assai istruttiva, che si ottiene usando il cosiddetto *paradigma DRM*, acronimo che identifica una particolare procedura sperimentale, affinata in anni diversi dagli psicologi nordamericani James Deese, Henry L. Roediger III e Kathleen McDermott. Il partecipante apprende una lista di parole (per esempio, *foglia acqua verde albero radice pioggia terra cielo fiore erba vegetazione flora*) tra loro associate; poi, in test di riconoscimento del tipo *yes/no* (analogo al paradigma *old/new*), il partecipante deve dire se una certa parola era inclusa nella lista studiata. In tali condizioni i falsi riconoscimenti di una parola come *natura*, mancante nella lista studiata ma fortemente

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 associata alle parole presentate, sono incredibilmente frequenti (fino all'80% di risposte *yes*). Non solo, le persone riferiscono di ricordare – in perfetta buona fede, dicono le ricerche sperimentali – specifici aspetti della presentazione di fatto mai avvenuta, come la voce di chi avrebbe letto la parola.

L'*effetto DRM* è robusto e ottenibile anche con materiale pittorico, dopo l'esposizione a sequenze d'immagini appartenenti a una specifica categoria di cui l'intrusa costituisce un elemento centrale (**cap. 8 su memoria e categorizzazione?**). Non rappresenta tuttavia l'unico caso di aggiunta, nel vissuto di ri-presentazione, di elementi obiettivamente assenti nella presentazione originaria. Dobbiamo specificare "obiettivamente assenti" perché non si può escludere che l'intrusione di un contenuto mentale fortemente associato si produca (talvolta almeno) già durante la presentazione del materiale da apprendere.



Figura 4.6. *Boundary extension*. La scena della fotografia *a* viene rappresentata mentalmente come più estesa della parte delimitata dai bordi effettivi. In una prova di riconoscimento i partecipanti scelgono *b* (e non *a*) come copia di ciò che gli è stato mostrato in precedenza.

Un altro caso di aggiunta di parti obiettivamente assenti è quello osservabile nell'effetto di *boundary extension* (ampliamento dell'immagine oltre i bordi), scoperto e studiato in modo approfondito dalla psicologa statunitense Helene Intraub (**figura 4.6**). Dopo aver visto la fotografia *a*, in un compito 2AFC i partecipanti tendono a preferire *b* e non *a* come corrispondente a ciò che gli è stato presentato originariamente.

In sostanza, come esistono le illusioni percettive così esistono le illusioni di riconoscimento, autentiche creazioni soggettivamente indistinguibili dai riconoscimenti veridici. Questa scoperta solleva il problema dei processi sottostanti le prestazioni di riconoscimento, su cui finora non abbiamo detto nulla. Infatti, sostenere che il riconoscimento deriva dal confronto tra l'informazione disponibile nel momento dato e qualcosa che abbiamo in memoria non spiega né come avvenga l'operazione di confronto (il *matching*) né quale sia il formato dei due insiemi di dati da confrontare.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15
Entrambe le questioni verranno trattate nella prossima sezione.

4.2. Modelli di riconoscimento

Nel caso del riconoscimento l'esigenza di sviluppare dei modelli nasce da due interessi: (i) comprendere come molti organismi, inclusi gli esseri umani, riescano con grande naturalezza a riconoscere oggetti per loro attraenti o pericolosi; (ii) costruire macchine che facciano altrettanto e, se possibile, di più. Questa combinazione d'interessi, relativi sia alla spiegazione dei fenomeni biologici sia alle tecnologie dell'artificiale, qualifica l'approccio interdisciplinare della **scienza cognitiva**. La motivazione alla ricerca su *pattern recognition* e *object recognition* è molto forte, perché le prestazioni di riconoscimento di cui sono capaci i sistemi biologici (esseri umani e altri animali) sono ancora complessivamente insuperate, anche se è ovvio che in compiti circoscritti un sistema artificiale come un *word processor* è più efficiente di un essere umano (per esempio, se si tratta di trovare tutte le ricorrenze di una data parola in un testo).

In un buon sistema di riconoscimento – e quello umano lo è – si combinano le seguenti proprietà.

- a) *Selettività*. Il sistema deve discriminare bene, cioè evitare di produrre la risposta “uguale” in presenza di stimoli che vanno classificati come diversi.
- b) *Invarianza*. Il sistema deve generalizzare bene, cioè evitare di produrre la risposta “diverso” in presenza di stimoli che vanno classificati come uguali.
- c) *Velocità*. La risposta deve emergere rapidamente.

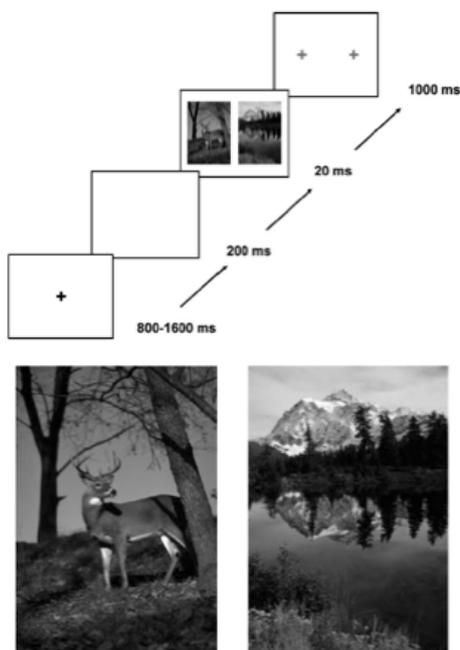


Figura 4.7. Holle Kirchner e Simon J. Thorpe, in un esperimento pubblicato su *Vision Research* nel 2006, chiedevano al partecipante di spostare lo sguardo dalla posizione centrale verso la scena contenente un animale. Le coppie di immagini erano presentate solo per 20 ms. Nelle prove più veloci, il tempo necessario a innescare il movimento oculare nella direzione giusta era di 120 ms (valore medio per i 15 partecipanti).

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

Non è facile quantificare la potenza del sistema umano rispetto a queste tre proprietà, ma è un luogo comune affermare che è molto elevata. Riguardo alla selettività, compiti complessi come il riconoscimento visivo delle impronte digitali in ambito forense è tuttora svolto da operatori umani. Riguardo all'invarianza, riconoscere uno stesso oggetto in movimento come autoidentico, a dispetto del cambiamento delle proiezioni retiniche, è una capacità altamente sviluppata negli organismi biologici. Riguardo alla velocità, in un compito di classificazione bastano 600 ms per decidere che lo stimolo presentato è l'immagine di animale; e in un compito in cui si devono solo spostare gli occhi verso sinistra o verso destra bastano 120 ms perché un osservatore allenato inizi a spostare gli occhi verso l'immagine che contiene un animale, piuttosto che un paesaggio (figura 4.7). I modelli presentati nelle sezioni 4.3.1-3 vengono valutati tenendo conto anche delle straordinarie prestazioni raggiungibili dagli esseri umani nei compiti di riconoscimento.

4.3.1. Confronto di sagome (*template matching*)

Se dovete scegliere un paio di scarpe per un bambino piccolo e non vi fidate delle sua capacità di rispondere a tono alla domanda “Ti stanno giuste?”, un buon modo è quello di estrarre la soletta e di verificare la corrispondenza tra piede e soletta, per evitare almeno delle eccessive discrepanze in più o in meno. L'espressione *template matching* definisce una tipologia di modelli che utilizza per il riconoscimento la sovrapposizione *diretta* tra una sagoma (*template*) depositata in memoria e l'immagine in input. In una versione elementare, se la discrepanza è inferiore a un valore soglia il confronto ha esito positivo. In biologia meccanismi del genere vengono chiamati “chiave-serratura” (*lock-and-key*).

Il *template matching* è utilizzabile in due operazioni base:

- i) rilevazione di un oggetto di interesse (*object-of-interest detection*), consistente nell'identificare, all'interno dell'immagine, la posizione spaziale per la quale la corrispondenza fra il pattern di valori di intensità del *template* e il pattern di valori dell'immagine esplorata è massima (Figura 4.8);
- ii) riconoscimento dell'oggetto maggiormente corrispondente al *template*, data una molteplicità di oggetti (come nel riconoscimento di caratteri, in cui un sistema deve decidere quale, fra un insieme di lettere, è, poniamo, la R). (vedi cap. XX – linguaggio??)

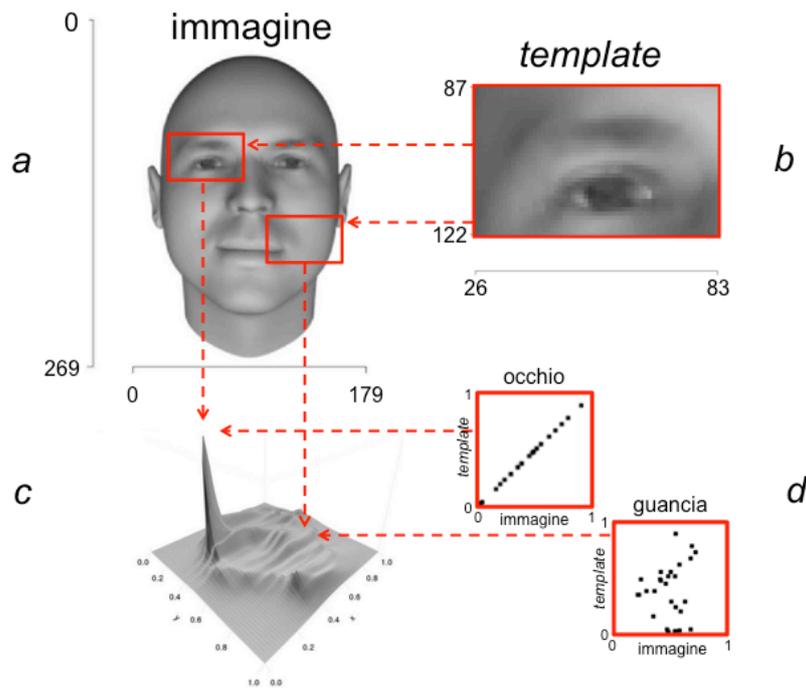


Figura 4.8. Schematizzazione di un'operazione di *template matching*, volta a identificare la posizione dell'occhio destro. L'immagine *a* viene “filtrata” facendo scorrere su di essa il *template* dell'occhio destro (ingrandito in *b*) e calcolando per ogni punto dell'immagine la correlazione tra i valori di grigio del *template* e quelli della porzione di immagine cui esso è sovrapposto. La mappa 3D in *c* mostra la distribuzione dei valori della correlazione (in altezza) per ogni punto dell'immagine; valori esemplificati in *d* dalla correlazione perfetta (pari a 1) per il *template* esattamente sovrapposto all'occhio destro e dalla correlazione vicina a 0 per il *template* sovrapposto alla guancia.

La **Figura 4.8** mostra l'output dell'esplorazione completa di un'immagine (il “filtraggio” della faccia) mediante scorrimento del *template* dell'occhio destro. In ogni posizione viene effettuato un confronto “punto a punto” fra gli n valori di intensità del *template*, riducibili alla sequenza $\mathbf{T}(x_1, x_2, \dots, x_n)$, e gli n valori di intensità della porzione di immagine su cui il *template* è sovrapposto, riducibili alla sequenza $\mathbf{I}(x_1, x_2, \dots, x_n)$. La correlazione fra $\mathbf{T}(x_1, x_2, \dots, x_n)$ e $\mathbf{I}(x_1, x_2, \dots, x_n)$, o altro parametro analogo, rappresenta la misura locale di somiglianza/dissomiglianza in base alla quale il sistema di riconoscimento decide sulla posizione più probabile dell'oggetto di interesse.

La **figura 4.8** mette anche in evidenza che l'output del *template matching* può essere molto distante dai suggerimenti della fenomenologia percettiva. Guardate la mappa in *c*: salvo il picco che identifica in modo univoco la posizione spaziale dell'occhio destro, la correlazione con il *template* è ovunque bassa, anche nella zona corrispondente all'occhio sinistro, che per qualunque osservatore umano è invece molto

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 simile all'occhio destro. Ma ciò che rende simile l'occhio destro e l'occhio sinistro non è la corrispondenza tra i valori di grigio rinvenibili nell'immagine, che possono essere piuttosto diversi. Il *template matching* – almeno nella forma semplificata illustrata nella **figura 4.8** – è cieco alla struttura degli oggetti.

Alcuni aspetti dell'operazione (ii) sono illustrati nella **figura 4.9**, che mostra l'esito di un confronto fra caratteri molto semplificati, inseribili in una matrice 7×4. Sovrapponendo il *template* della lettera A si ottengono i seguenti valori di dissomiglianza (espressi mediante la proporzione di celle discrepanti, una metrica molto semplice): 0 per la lettera A, ovviamente, e 0,21 per le lettere U e F. Quella che ai nostri occhi è un'evidente differenza – la diversa struttura della U e della F, l'una rispetto all'altra, e di ciascuna rispetto alla A – si traduce in un valore di dissomiglianza che in sostanza le tratta come uguali. Nella sua schematicità la **figura 4.9** mostra che *il template matching*, almeno in questo esempio, non è un buon modello del riconoscimento umano dei caratteri, anche se può servire a identificare l'oggetto di interesse, trascurando l'analisi degli oggetti non di interesse.

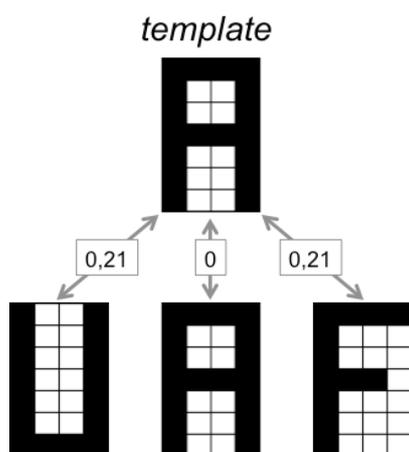


Figura 4.9. Il *template* di una A, sovrapposto a tre diverse lettere, può produrre due esiti soltanto: corrispondenza perfetta per la A (dissomiglianza 0) e discrepanza uguale per U e F (6 celle su 28 con valori diversi fra *template* e immagine).

Un sistema di riconoscimento basato su *template* deve comunque prevedere alcune operazioni di *normalizzazione* dell'immagine; altrimenti la sovrapposizione fra *template* e immagine produrrebbe valori di dissomiglianza elevati anche quando l'immagine è strutturalmente identica al *template*, ma ha grandezza o orientamento diversi. Con tale accorgimento un modello a *template* può avere una buona valenza esplicativa e rivelarsi compatibile con i dati sperimentali relativi al riconoscimento umano [**Riquadro 4.1**].

Il limite forse più grosso del *template matching* – ciò che lo rende inadeguato come modello generale del riconoscimento – deriva dalla variabilità praticamente infinita degli oggetti da classificare e identificare. L'idea che ogni oggetto distinguibile

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 da un altro e ogni vista prospettica di ciascun oggetto debbano trovare nella mente un *template* in grado di rappresentarlo, dando luogo al corrispondente riconoscimento, appare biologicamente implausibile. I modelli a rete, descritti nella sezione seguente, cercano di superare questa difficoltà usando la strategia componenziale che sappiamo funzionare nel linguaggio: la combinazione di poche lettere, a loro volta riducibili a combinazioni di pochi tratti grafici, dà luogo a una molteplicità di parole. Ciò che funziona nella produzione del linguaggio potrebbe funzionare anche nella percezione.

4.3.2. Reti di caratteristiche (*feature nets*)

Nel *template matching* (sezione 4.3.1) l'input ottico viene confrontato in blocco con uno o più forme conservate in memoria, dopo le operazioni di normalizzazione (rotazione, traslazione, cambiamento di scala) che comunque coinvolgono l'intera immagine. Al contrario, i modelli a rete basati sull'analisi delle caratteristiche (*feature analysis*) mirano a ottenere un'elaborazione più efficiente sfruttando la divisione del lavoro fra più rilevatori (*feature detectors*), ciascuno dei quali funziona, nel suo piccolo, in base al modello del *template matching*.

Questa organizzazione schiavistica del riconoscimento è incorporata nel *pandemonium*, proposto dal padre della "percezione artificiale" Oliver G. Selfridge (1926-2008) come modello del riconoscimento dei caratteri e, salendo lungo la gerarchia, delle parole. In verità, il riferimento a un mondo di immaginari demoni mentali – immaginari sia perché non dovete pensare di averli davvero nella testa sia perché operanti sulle immagini in input – cattura non tanto l'aspetto gerarchico del loro passaparola, quanto il fatto che il loro linguaggio obbedisce al *principio dell'univarianza* (sezione 2.6.2). La risposta di ciascuna unità del *pandemonium* è fatta solo di urla più o meno forti, metafora acustica della quantità di attivazione di ciascun demone delle caratteristiche e di ciascun demone cognitivo deputato a segnalare un aggregato di caratteristiche (figura 4.10).

Il *pandemonium* di Selfridge condivide molti dei difetti del *template matching*, posto che – come abbiamo già detto – i demoni di livello più basso riconoscono le caratteristiche sempre in base alla sovrapposizione di sagome; ma la sua architettura – cioè la distribuzione della funzione di riconoscimento fra unità variamente connesse – ha avuto grande influenza, anche per la convergenza con i dati della ricerca neurofisiologica sull'organizzazione gerarchica del sistema visivo, in cui è possibile trovare cellule che rispondono a proprietà degli stimoli via via più complesse man mano che ci si muove dalle aree visive primarie della corteccia cerebrale verso aree di livello

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 più elevato. A rendere il *pandemonium* psicologicamente poco plausibile è la rigida organizzazione *bottom-up*, inadatta a spiegare i molti fenomeni *top-down* – come l'effetto di superiorità della parola (cap. xx) – misurabili in ambito sperimentale.

I modelli del tipo *feature net* (incluso il *pandemonium* di Selfridge) funzionano bene in domini limitati come l'insieme dei caratteri alfabetici, prodotto di un processo selettivo di tipo culturale che ha fatto sopravvivere un numero limitato di simboli contraddistinti dalla presenza/assenza di poche caratteristiche ben discriminabili. È abbastanza naturale, quindi, constatare che questa tipologia di modelli si è rivelata molto utile negli studi sul linguaggio scritto e parlato (cap. 15). Un importante sviluppo in questo ambito riguarda la compresenza di collegamenti *bottom-up* e *top-down*, necessari a spiegare i fenomeni contestuali, come quelli in cui l'elaborazione locale risulta influenzata dai fattori globali.

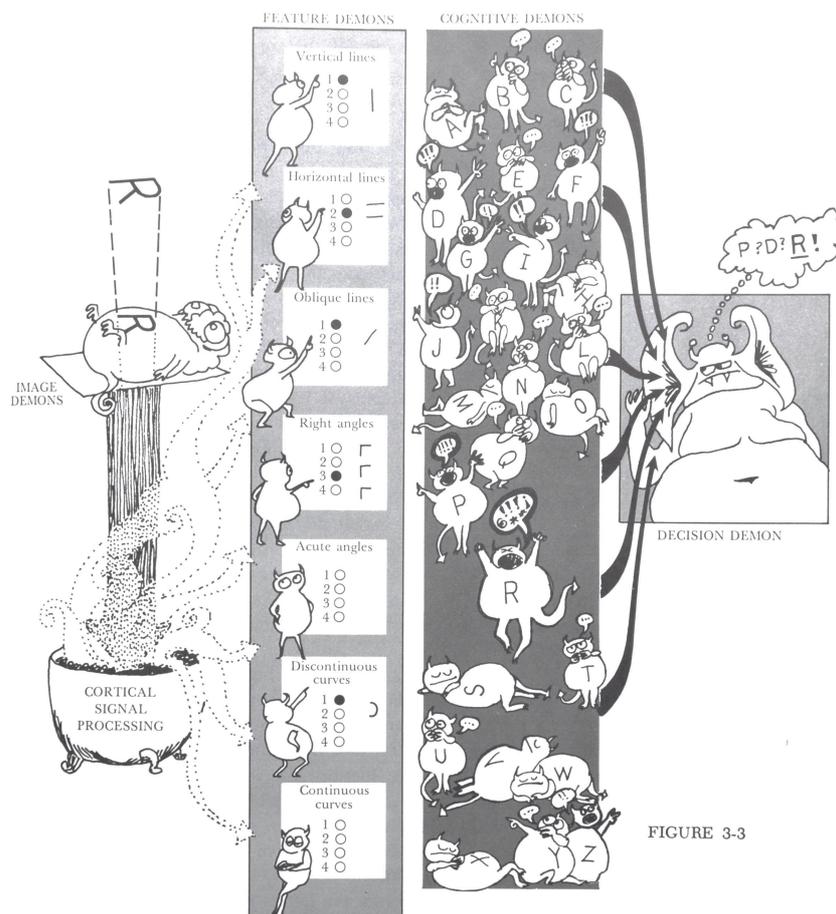


FIGURE 3-3

Figura 4.10. Il *pandemonium* di Selfridge in uno dei fumetti che popolavano il fortunato manuale di Peter H. Lindsey e Donald A. Norman intitolato *Human Information Processing: An Introduction to Psychology*, apparso nel 1972. L'informazione fluisce da sinistra a destra: l'input **R** stimola i demoni dell'immagine (i recettori retinici) e avvia l'analisi del segnale neurale consistente nella scomposizione dell'immagine in tipo e numero di caratteristiche presenti (un tratto verticale, due tratti orizzontali, un tratto obliquo, tre angoli retti, una curva discontinua) e assenti (nessun angolo acuto, nessuna curva continua). I demoni

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

cognitivi (uno per ciascuna lettera dell'alfabeto) si attivano in funzione delle caratteristiche presenti/assenti: il demone della **R** urla come un ossesso, e pure quelli della **F**, della **P** e della **D** si sentono un po' chiamati in causa. Al demone decisionale spetta identificare il demone cognitivo che urla di più.

4.3.3. Descrizioni strutturali (*structural descriptions*)

Come abbiamo cercato di chiarire nella sezione 4.3.1, un limite dei modelli a *template* è la loro cecità alla struttura, cioè l'incapacità di rappresentare le relazioni tra le parti dell'oggetto, trattato come un blocco di informazione da analizzare punto a punto. Sotto questo profilo i *feature net* non sono molto diversi: un'unità sovraordinata come un demone cognitivo risponde in funzione della mera presenza/assenza delle caratteristiche che congiuntamente definiscono il suo oggetto preferenziale, non della loro relazione reciproca.

Evidenziare le *relazioni* tra le parti (proprietà che la teoria della Gestalt ha sempre considerato cruciali) è l'obiettivo dei modelli basati sulle *descrizioni strutturali*, a partire da quello incorporato nella teoria dell'elaborazione visiva proposta dallo psicologo e neuroscienziato britannico David Marr [Riquadro 4.2]. Lo stadio finale dell'elaborazione visiva prevede, secondo Marr, la costruzione di un modello 3D con coordinate centrate sull'oggetto e basato sull'utilizzo del *cilindro generalizzato* come primitiva volumetrica (figura 4.11). Elementi fondamentali del modello 3D proposto da Marr sono l'asse di allungamento di ciascun cilindro e la disponibilità di rappresentazioni a più livelli di risoluzione (la figura umana rappresentata in prima approssimazione come un cilindro unico, e poi come struttura articolata in più cilindri legati da specifiche relazioni spaziali, via via più dettagliate). Il riconoscimento avverrebbe soltanto dopo la costruzione di questo modello, sulla base di un confronto con le descrizioni strutturali degli oggetti conservate in memoria.

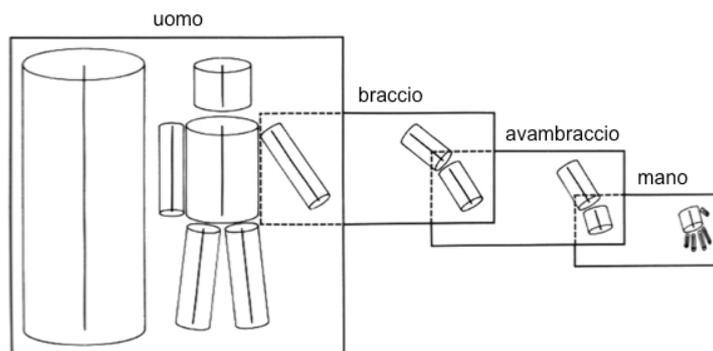


Figura 4.11. Il modello 3D di David Marr incorpora una rappresentazione a scale diverse basata sulla descrizione delle relazioni strutturali tra parti, ciascuna rappresentata mediante un cilindro specificato da vari parametri (raggio, lunghezza dell'asse, orientamento 3D dell'asse). Il modello 3D è l'input per il riconoscimento.

Le descrizioni strutturali sono un elemento costitutivo anche del modello RBC (*Recognition-By-Components*) di Irving Biederman, che si differenzia dal modello di Marr in alcuni aspetti rilevanti. A partire dal 1987 Biederman ha elaborato un modello

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 del riconoscimento degli oggetti basato su una trentina di primitive volumetriche, chiamate *geoni*, un neologismo che sta per “ioni geometrici” (figura 4.12).

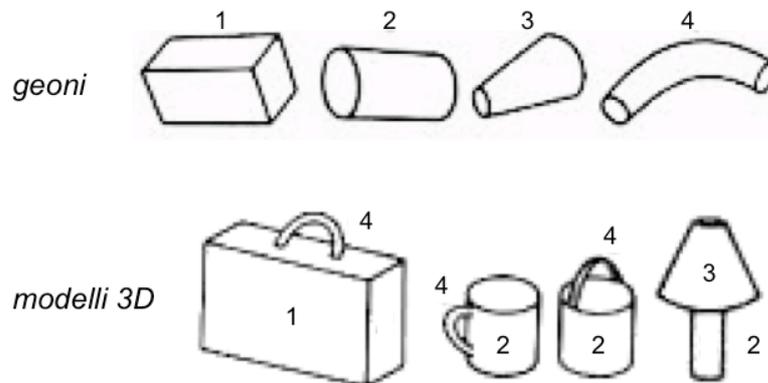


Figura 4.12. Quattro geoni proposti da Irving Biederman come primitive volumetriche (in alto). In moltissimi casi i modelli 3D usati nel riconoscimento degli oggetti sono combinazioni di due soli geoni (in basso).

I geoni sono i costituenti elementari delle rappresentazioni degli oggetti, caratterizzate dal tipo di ioni e dalle relazioni geometriche che li legano. Nel modello RBC i geoni vengono identificati grazie alla presenza delle cosiddette *proprietà non accidentali*, caratteristiche dell'immagine con elevato valore diagnostico. Il concetto di proprietà non accidentale è strettamente legato all'idea helmholtziana che la percezione sia una sorta di ipotesi sulla causa esterna che ha prodotto lo stimolo prossimale. Se l'immagine contiene una configurazione di tre segmenti di contorno che convergono in un punto (per esempio, formando una Y), è improbabile che essa si sia prodotta per caso, come conseguenza dell'accidentale convergenza di tre segmenti indipendenti in un unico punto dello spazio; è molto più probabile che essa derivi dalla proiezione di un triedro solido (come l'angolo convesso di un tavolo o l'angolo concavo formato da due pareti e dal soffitto). Biederman ha selezionato i geoni in base delle proprietà non accidentali più stabili, quelle che con maggiore probabilità troviamo nelle immagini.

4.3.4. Quale modello per il riconoscimento umano?

Lo sviluppo dei modelli di riconoscimento, legato anche alla possibilità concreta di costruire macchine che sostituiscano l'uomo in molti compiti di rilevazione di oggetti di interesse, svolge un'importante funzione nella comprensione del riconoscimento umano. Confrontando la prestazione umana con la prestazione attesa in base a un modello (meglio se osservata in una simulazione al computer) possiamo quanto meno rigettare i modelli manifestamente inadeguati.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

Cerchiamo allora di capire se, a grandi linee, i processi di riconoscimento nell'osservatore umano sono meglio modellati da un sistema visivo che utilizza il confronto con l'immagine in input – o nella sua totalità (*template matching*) o dopo una scomposizione in caratteristiche 2D (quindi sempre riferite all'immagine) – oppure da un sistema visivo in cui il confronto avviene tra una descrizione strutturale delle relazioni fra primitive volumetriche e il modello interno (sempre 3D) degli oggetti di interesse. È importante tener presente che l'interpretazione dei dati empirici non è facile. Proprio l'esame delle varie tipologie di modelli descritti nelle tre sezioni precedenti dovrebbe aver chiarito che un riconoscimento non ottimale potrebbe dipendere dall'inadeguatezza della rappresentazione dell'input (a livello dell'immagine o di strutture da essa derivate), del modello interno dell'oggetto di interesse oppure del processo di confronto; per non parlare dell'inadeguatezza di più di uno di questi elementi.

Paradossalmente, il limite principale dei modelli basati sulle descrizioni strutturali centrate sull'oggetto è il fatto che – in una versione radicale – essi funzionano addirittura troppo bene. Se la capacità di generare un modello 3D a partire dalle immagini in input fosse perfetta, la prestazione umana nei compiti di riconoscimento non dovrebbe dipendere dalla vista disponibile in fase di test. Invece il riconoscimento umano, pur esibendo una buona invarianza prospettica (dopo un'occhiata a un oggetto sconosciuto siamo in grado di riconoscerlo anche ruotato), non è del tutto indifferente al punto di vista. Com'è giusto – direte voi – giacché per un oggetto sconosciuto l'unica fonte di informazione è la sua immagine.

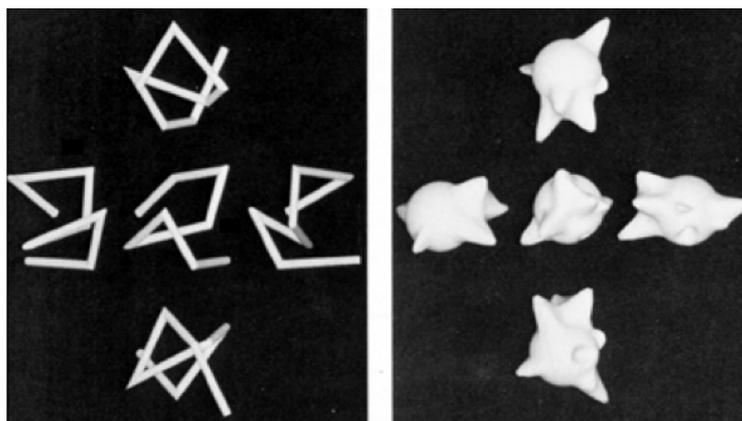


Figura 4.13. Bülthoff e Edelman utilizzarono due tipi di oggetti 3D non familiari: fermagli con piegature casuali e amebe con protuberanze casuali. Le quattro viste intorno alla vista centrale corrispondono a rotazioni di ± 75 gradi rispetto a un asse verticale o orizzontale.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

La questione era affrontata in un lavoro di Heinrich H. Bülthoff e Shimon Edelman pubblicato nel 1992. In fase di addestramento il partecipante vedeva due viste dinamiche di uno stesso oggetto separate di 75 gradi (figura 4.13): per esempio, il fermaglio centrale e quello a sinistra, in oscillazione da -15 a +15 gradi (una presentazione ricca di informazioni sulla struttura 3D). In fase di riconoscimento compariva una nuova vista statica dello stesso oggetto (prove positive con risposta corretta *yes*) oppure una vista statica di un altro fermaglio con piegature diverse (prove negative con risposta attesa *no*). Nelle prove positive la vista poteva essere: (i) intermedia fra le viste oscillanti usate in fase di addestramento; (ii) esterna rispetto alle viste oscillanti, come nel fermaglio a destra di quello centrale; (iii) ortogonale rispetto alle viste oscillanti, come nel fermaglio sopra a quello centrale.

Il riconoscimento era ottimo quando la vista test era intermedia, a riprova del fatto che è facile interpolare la forma di un oggetto 3D, date due sue viste; era peggiore quando la vista test era esterna alle due viste date; e ancora peggiore quando era ortogonale. La superiorità dell'interpolazione sull'estrapolazione, a parità di distanza angolare assoluta dalle viste di addestramento, mette in difficoltà qualunque modello basato sulla ricostruzione perfetta della struttura 3D. In fase di addestramento, sembra che l'osservatore riempia il vuoto fra le viste fornite, più che ricostruire l'intero oggetto.

4.3. Effetto inversione

Studiando la sezione 4.2 potreste aver ricavato l'impressione erronea che l'aspetto cruciale della dicotomia fra modelli basati sull'immagine e modelli basati sulle descrizioni strutturali riguarda il numero di dimensioni spaziali coinvolte: cioè il fatto che il confronto avvenga tra rappresentazioni 2D piuttosto che 3D. Non è questo il punto. La differenza riguarda il contenuto delle due rappresentazioni: nei modelli basati sull'immagine il confronto avviene *punto a punto*; nei modelli basati sulle descrizioni strutturali contano le *relazioni* tra le parti, anche sul piano frontale (senza tirare in ballo la profondità e quindi la ricostruzione volumetrica dell'oggetto).

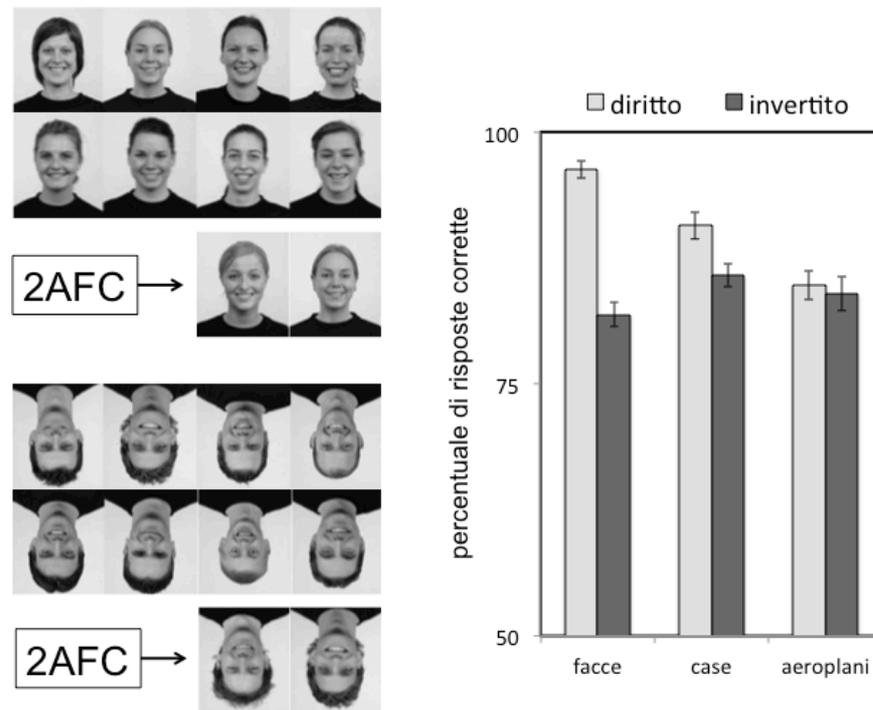


Figura 4.14. Effetto inversione in un compito di riconoscimento *old/new*. L'effetto è molto più forte nelle facce e in generale per il materiale con cui l'osservatore ha molta familiarità (case ma non aeroplani). Il grafico a destra mostra i dati ottenuti da Yin nel primo degli esperimenti riportati in un famoso articolo del 1969. La barretta associata a ogni percentuale media di risposte corrette è una misura di variabilità della media. Siccome la variabilità è sempre piccola, le differenze tra le medie sono molto affidabili.

Per valutare l'importanza delle relazioni tra le parti si utilizza l'inversione sul piano frontale (cioè la rotazione di 180 gradi dell'immagine). Questa operazione lascia inalterati tutti i parametri locali dell'immagine, ma incide a livello della descrizione strutturale, se questa include (come comunemente si ritiene) descrittori del tipo "sopra/sotto", che fanno riferimento alla relazione reciproca delle parti rispetto alla polarità alto-basso (la cui importanza è stata descritta nella [sezione 3.5.1](#)).

L'esperimento classico, che aprì un fruttuoso filone di ricerca, è quello di Robert K. Yin, pubblicato nel 1969 e condotto al MIT sotto la guida di Hans-Lucas Teuber (1916-1977), neuropsicologo fortemente legato alla tradizione gestaltista. La [figura 4.14](#) illustra: a sinistra, il tipo di compito utilizzato nelle due sessioni con facce rispettivamente diritte e capovolte; a destra, la percentuale di risposte corrette nel riconoscimento *old/new* quando il materiale è costituito da facce, case, aeroplani. Due sono i risultati notevoli: (i) apprendere e riconoscere forme mono-orientate (cioè dotate di una polarità alto/basso) è più difficile quando queste sono invertite; (ii) il deterioramento della prestazione (*effetto inversione*) è massimo per le facce e si riduce quanto più il materiale è poco familiare.

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15

L'importanza della familiarità e dell'*expertise* (l'essere esperti) è confermata da altri esperimenti dai quali emerge che l'entità dell'effetto inversione aumenta, per lo stesso materiale, man mano che si diventa esperti. In un compito come quello illustrato nella **figura 4.14** a sinistra, gli esperti di razze canine mostrano un effetto inversione *superiore* agli inesperti. Diventare esperti produce un vantaggio nel riconoscimento di forme presentate nell'orientamento canonico, ma produce anche un effetto negativo: aumenta l'entità dell'effetto inversione, minimo negli inesperti. Se volessimo adottare il *template matching* come modello di riconoscimento umano, dovremmo comunque concludere che la capacità umana di effettuare l'operazione di normalizzazione richiesta dal compito di Yin (la rotazione sul piano frontale) è fortemente limitata.

L'effetto inversione non è una prerogativa umana. Un bellissimo esperimento condotto dalla psicologa comparata Jennifer J. Pokorny e collaboratori, pubblicato nel 2011, ha dimostrato che nelle scimmie cappuccine l'effetto inversione è massimo per le facce dei conspecifici e forte per quelle umane (con cui le scimmie cappuccine hanno buona familiarità); mentre scompare per le facce di scimpanzé e per le immagini di veicoli (entrambe non familiari).

4.4. Percezione categoriale

L'azione *top-down* della familiarità ha un'importante conseguenza sulla somiglianza percepita e quindi sulla discriminabilità: se sono familiari, stimoli appartenenti a un medesimo continuum fisico possono essere percepiti in modo categoriale, con conseguenti *assimilazione interna* (elementi appartenenti alla stessa categoria appaiono più simili di quanto apparirebbero se la rappresentazione fosse non-categoriale) e *contrasto esterno* (elementi appartenenti a categorie diverse appaiono meno simili di quanto apparirebbero se la rappresentazione fosse non-categoriale).

Un effetto di percezione categoriale delle facce venne documentato da James M. Beale e Frank C. Keil nel 1995. **La figura 4.15** illustra l'effetto base, riferito a una coppia di presidenti USA più recenti di quelli del lavoro originario (Kennedy e Clinton). Beale e Keil aveva utilizzato la tecnica del *morphing* (modificazione controllata della forma), che consente di miscelare due forme diverse (come le foto frontali di due persone diverse) nelle percentuali desiderate. Se per voi i volti di Bush e Obama sono abbastanza familiari, osservando le quattro facce in basso dovrete avvertire come più diverse tra loro le due facce centrali, appartenenti a categorie diverse (rispettivamente, "Bush" per la faccia 60% Bush – 40% Obama e "Obama" per la faccia 40% Bush – 60% Obama), rispetto alle facce che si trovano dallo stesso lato del confine categoriale

materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 (la faccia 50% Bush – 50% di Obama). Nell'esperimento di discriminazione venivano usate solo le facce in basso, in tre blocchi di prove, utilizzando il compito *AB-X*, in cui si presenta prima una coppia di stimoli leggermente diversi (l'intervallo lungo il continuum *morph* delle facce era pari al 20%) e poi una dei due. Il partecipante deve decidere se X è uguale a A oppure a B. Le percentuali di accuratezza era più elevata (85%) per la coppia centrale, a cavallo del *confine categoriale*. La discriminazione era più difficile (65% e 75%) quando entrambe le facce AB appartenevano alla stessa categoria: erano cioè percepite entrambe come “Bush” o come “Obama”, in base alla prevalenza nel *morph*.

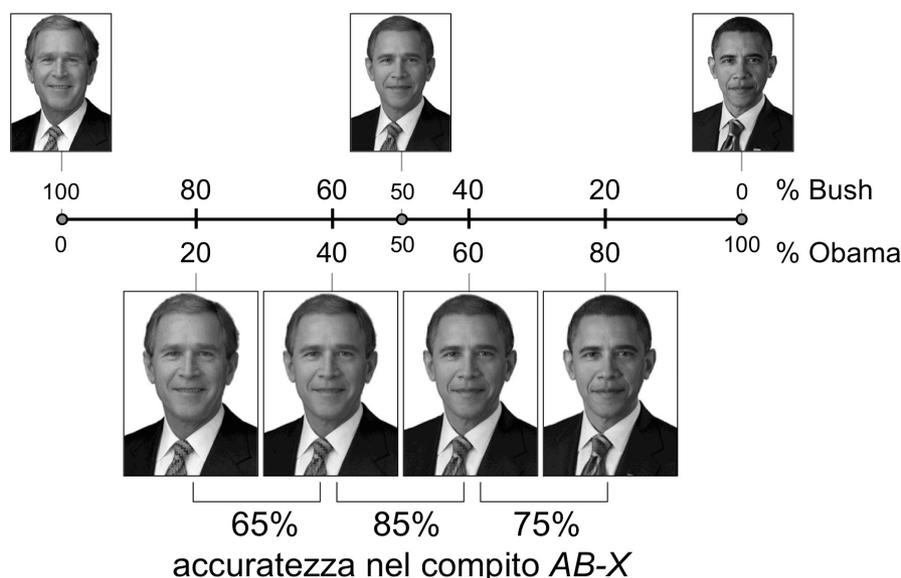


Figura 4.15. Aggiornamento dell'esperimento di Beale e Keil, che nel 1995 avevano utilizzato il continuum *morph* Kennedy-Clinton. Le quattro facce in basso sono equidistanti sul continuum *morph*.

Il bello è che le facce a cavallo del confine categoriale non sono discriminate meglio se il continuum ha per estremi degli sconosciuti. In quel caso la discriminabilità degli elementi di ogni coppia di facce dipende soltanto dalla percentuale di mescolanza: due facce distanti il 20% sul continuum *morph* sono discriminate con la stessa facilità, indipendentemente dalla posizione rispetto al confine categoriale.

La segmentazione categoriale è stata dimostrata in vari ambiti, dalla percezione dei fonemi (**cap. XX**) a quella del colore cromatico. In generale, si è osservato che la posizione del confine categoriale lungo il continuum è influenzata da vari fattori, anche culturali, che comunque portano la mente ad accentuare le differenze tra gli stimoli

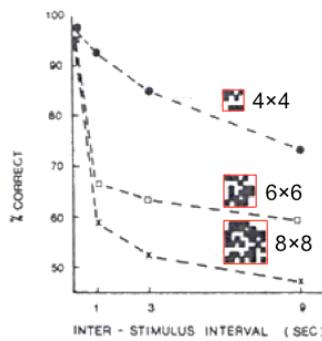
materiale fornito dall'autore come preprint agli studenti del primo semestre 2014/15 separati dal confine categoriale e a ridurre le differenze tra gli stimoli inclusi nella medesima categoria.

4.5. Epilogo

A questo punto potremmo avere dei dubbi sulla possibilità di separare in modo netto percezione e riconoscimento. Ci siamo resi conto che il termine riconoscimento può essere usato, in un'accezione onnicomprensiva, per indicare qualunque processo sottostante una risposta agli stimoli, che per essere “digeriti” dalla mente hanno bisogno di essere comunque ricondotti a rappresentazioni confrontabili con i modelli interni (continui o categoriali, questa è una questione a se stante). In questa accezione estesa tutta la percezione diviene una forma di riconoscimento.

Ma usare “riconoscimento” in modo onnicomprensivo, come inevitabile approdo di ogni processo che porta al costituirsi dell'esperienza diretta, sarebbe inopportuno. Perderemmo la possibilità di isolare i casi in cui appare il vissuto di familiarità e quelli in cui si può dimostrare in laboratorio che la familiarizzazione è cruciale (come l'effetto inversione e la percezione categoriale). Va infine tenuto presente che il termine riconoscimento indica anche – in psicologia sperimentale – un modo per saggiare la memoria, contrapposto tipicamente alla rievocazione libera (**cap. xx – memoria**). L'importante – perdonerete il gioco di parole – è saper riconoscere i diversi significati di riconoscimento e non farsi ingannare dal primo significato colto dalla mente, come nel caso della mucca nascosta nella **figura 4.5**, resa famosa dallo psicologo americano Karl Dallenbach (1887-1971).

Riquadro 4.1. *Template matching* e complessità.



Il grafico a sinistra è tratto da un lavoro emblematico dell'approccio *human information processing* (elaborazione dell'informazione nell'uomo) – dominante nella psicologia sperimentale degli anni Sessanta e Settanta – che utilizzava in modo sistematico l'analogia mente-computer. Nel 1974 Bill Phillips, che lavorava all'Università di Stirling, in Scozia, pubblicò una serie di esperimenti sul confronto successivo tra pattern casuali, matrici di celle metà nere e metà bianche.

Il grafico riassume i risultati del primo esperimento: la percentuale di risposte corrette nel compito stesso/diverso (ispirato al *template matching*) diminuisce all'aumentare dell'intervallo inter-stimolo (*inter-stimulus interval*, ISI; da 0 a 9 secondi) e della complessità della matrice (4×4, 6×6, 8×8). Le curve coprono tutto l'intervallo fra la prestazione perfetta (100% di riconoscimenti corretti per intervalli inferiori ai 300 ms, indipendentemente dalla complessità) alla prestazione casuale (50% di risposte corrette per le matrici più complesse, con ISI lungo).

Negli esperimenti successivi Phillips dimostrò che la prestazione di riconoscimento misurata in questo compito stesso/diverso è eterogenea: quando l'ISI è inferiore ai 300 ms il partecipante *vede* – e riferisce di vedere con chiarezza – la permanenza della medesima matrice o l'eventuale cambiamento, che si manifesta come *movimento apparente* di tipo stroboscopico delle celle cambiate; quando l'ISI è più lungo, invece, il partecipante *ricorda* che la matrice presentata per prima era diversa dalla seconda. Nel primo caso è operante il meccanismo percettivo di “trasformazione continua nello spazio-tempo”, che ci fa vedere il movimento fluido anche quando gli stimoli sono statici e in successione discreta (come al cinema e in molte insegne luminose); nel secondo è operante la memoria visiva a breve termine (cap. 7). Phillips mise in evidenza che sia la fenomenologia (ciò che il soggetto riferiva di vedere in modo saliente) sia l'andamento della prestazione in condizioni controllate dimostravano come un compito formalmente descrivibile come “riconoscimento” possa corrispondere a processi psicologicamente diversi.

Riquadro 4.2.

David Marr (1945-1980) fornì contributi fondamentali alla scienza cognitiva, integrando neurofisiologia, psicofisica e visione artificiale. Il testo intitolato *Vision*, pubblicato nel 1982 dopo la sua prematura scomparsa, contiene una strategia per lo studio dei sistemi di elaborazione dell'informazione, biologici e artificiali, basata su tre livelli complementari di analisi: computazionale, algoritmico/rappresentazionale, fisico/implementativo.

- i. Al livello superiore viene formulato il *problema computazionale* che qualunque sistema deve risolvere (per la visione, costruire un modello 3D utilizzando le immagini 2D) e l'esplicitazione del suo significato adattivo.
- ii. L'individuazione di un *algoritmo* adeguato riguarda il modo in cui sono rappresentate le informazioni oggetto del problema computazionale e le operazioni necessarie alla sua soluzione, espresse in un linguaggio formale.
- iii. Il livello più specifico riguarda l'*implementazione fisica* dell'algoritmo individuato dalla teoria, in un sistema biologico (mediante meccanismi basati sulla comunicazione neuronale) o in un sistema artificiale (per esempio, un dispositivo elettronico o elettromeccanico).

Una teoria completa della visione conterrà enunciati che collegano un livello e l'altro e che tengono conto del legame asimmetrico fra visione artificiale e visione biologica. Quest'ultima costituisce – essendo il prodotto della selezione naturale – un modello di buon funzionamento. Come sottolinea Shimon Edelman – altro grande della visione computazionale – i sistemi di elaborazione dell'informazione ricordano le famiglie di cui parla Lev Tolstoj all'inizio di *Anna Karenina*: “Tutte le famiglie felici sono simili le une alle altre; ogni famiglia infelice è infelice a modo suo.”

Ispirandosi alla visione biologica Marr formulò alcuni principi che – a suo avviso – debbono valere per ogni sistema visivo. Uno di questi, il principio del “minimo impegno” (*principle of least commitment*) riguarda in modo specifico il riconoscimento: se l'elaborazione dell'informazione presente nello stimolo prossimale viene pensata come una catena di processi guidati dall'obiettivo di riconoscere la causa esterna dello stimolo prossimale (idea helmholtziana che domina gran parte della visione artificiale), allora vanno evitate le decisioni precoci basate su rappresentazioni parziali o incomplete. Meglio attendere di avere acquisito e integrato tutta l'informazione disponibile.

Il *principio del minimo impegno* ispira la distinzione in stadi di elaborazione dell'input ottico proposta da Marr, che è motivata dall'esigenza di “spremere” tutta l'informazione dallo stimolo prima di attivare il riconoscimento (cioè il confronto con i contenuti depositati in memoria). I principali stadi sono i seguenti:

- *primal sketch* (schizzo primario), una rappresentazione dei contorni presenti nell'immagine;
- *2,5 D sketch* (schizzo a 2 dimensioni e mezzo), una rappresentazione in rilievo delle sole superfici direttamente visibili (senza un completamento volumetrico 3D delle superfici auto-occluse) ottenuta integrando i vari indizi di profondità in una distribuzione di vettori normali all'orientamento locale della superficie (come illustrato nella rappresentazione del doppio cilindro qui accanto);
- *modello 3D*, una rappresentazione centrata sull'oggetto (mentre le due precedenti sono centrate sull'osservatore) che ne esplicita la struttura gerarchica (Figura 4.11).

