

The title

First Author¹ & Ernst-August Doelle^{1,2}

¹ Wilhelm-Wundt-University

² Konstanz Business School

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

The authors made the following contributions. First Author: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; Ernst-August Doelle: Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to First Author, Postal address. E-mail: my@email.com

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline. Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines. One sentence clearly stating the **general problem** being addressed by this particular study. One sentence summarizing the main result (with the words “**here we show**” or their equivalent). Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge. One or two sentences to put the results into a more **general context**. Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: keywords

Word count: X

The title

Introduzione

Il sistema visivo umano (e, in generale, biologico) deve ricostruire la forma tridimensionale dell'ambiente a partire dall'informazione sensoriale. Ci sono tanti "indizi" sensoriali che forniscono informazioni sulla forma 3D dell'ambiente e del movimento relativo tra l'ambiente e l'osservatore. Inoltre, l'ambiente non è rigido, ma si deforma (cambia forma) nel tempo, il che rappresenta un'ulteriore complicazione. Un modo per leggere il fenomeno stereocinetico di cui ci occupiamo qui è di pensare che esso isoli uno specifico "indizio" di profondità che può essere utile per il sistema visivo per ricostruire la forma 3D degli oggetti. Proffitt, Rock, Hecht, and Schubert (1992) e Caudek and Proffitt (1993) hanno mostrato che, infatti, lo stimolo che suscita l'effetto stereocinetico (i cerchi quasi concentrici che ruotano rigidamente sul piano) forniscono un sottoinsieme delle informazioni che verrebbero fornite dalla proiezione polare "corretta" (ovvero, completa) di un oggetto rigido, ovvero un cono, con degli anelli dipinti sulla sua superficie, che oscilla in relazione ad un osservatore (o viceversa). L'articolo seminale di Musatti (1924) si pone il problema di come spiegare, in termini di un modello "di proiezione inversa" il percepito che questo stimolo suscita. La proposta specifica fornita da Musatti (1924) per "spiegare" questo fenomeno percettivo ha un interesse più che altro storico. Invece, ciò che ancora suscita il nostro interesse, non sono tanto le caratteristiche specifiche del modello geometrico che Musatti (1924) propone, quanto il problema che Musatti solleva. Il problema è il seguente: come è possibile per un osservatore percepire un oggetto tridimensionale con caratteristiche molto ben definite e abbastanza costanti tra osservatori diversi, sulla base di questo stimolo? Il problema potrebbe sembrare di poco interesse: "è solo un'illusione (in realtà, quello che succede è che c'è un pattern bidimensionale che ruota rispetto all'osservatore). Nel mondo naturale queste cose non succedono, quindi non c'è nulla da spiegare." Questa spiegazione, però, è troppo affrettata e, in realtà, gli stimoli stereocinetici di Duchamp e di Benussi sono degli oggetti fisici presenti nel mondo empirico (non sono solo delle "illusioni" che sperimentiamo soltanto

nei sogni o quando il nostro sistema cognitivo è in qualche modo alterato). Pertanto, una teoria che vuole spiegare come funziona il sistema visivo non può ignorarli. In questo articolo mostreremo come, in realtà, è stato proprio così: questi fenomeni, e il “problema” che Musatti (1924) ha posto, sono stati *ignorati*. E questo è un problema che ci accompagna. Questi fenomeni “illusori” sono stati ignorati dalla teoria dominante perché tale teoria non riusciva a spiegarli. Il punto è che, ignorando tali fenomeni, la teoria dominante (che qui chiameremo il Modello Standard) si è condannata a non riuscire a spiegare né i fenomeni che intendeva ignorare, né i fenomeni (tutto il resto) che in realtà intendeva spiegare. Questa è la “lezione” che questo breve articolo vuole suggerire e che dimostra quanto sia ancora attuale il “problema” che Musatti (1924) ha sollevato. Ma per articolare questa “lezione” è necessario entrare nei dettagli. Il diavolo, sempre, si nasconde nei dettagli.

Perché l'effetto stereocinetico è importante?

lo stimolo stereocinetico si distingue da una proiezione polare ideale, generata dal movimento relativo tra un osservatore e un cono, poiché, a livello locale e istantaneo sulla retina, esso fornisce informazioni esclusivamente riguardanti il campo di velocità, omettendo dettagli sulle accelerazioni. Secondo la terminologia introdotta da James J. Gibson, ciò significa che il flusso ottico derivante da tale stimolo include dati relativi alle velocità relative degli oggetti nello spazio, ma non incorpora informazioni sulle loro accelerazioni.

Se le informazioni sull'accelerazione fossero presenti, utilizzando principi di “geometria inversa”, sarebbe teoricamente possibile ricostruire la forma tridimensionale dell'oggetto distale con un determinato grado di precisione, a meno di un fattore di scala indeterminato. L'assenza di dati relativi alle accelerazioni implica che la configurazione tridimensionale dell'oggetto viene definita in termini affini, il che significa che essa è conciliabile con un'infinità di possibili oggetti distali che variano nella profondità z da 0 a infinito. Tuttavia, questo non corrisponde a ciò che effettivamente gli osservatori percepiscono. Gli osservatori riportano una profondità ben determinata, non “indeterminata”. La discrepanza tra

l'informazione visiva fornita dallo stimolo stereocinetico e la percezione tridimensionale che suscita solleva dunque questioni fondamentali sulla natura dei meccanismi attraverso i quali il sistema visivo interpreta e integra le informazioni di velocità senza dati espliciti sull'accelerazione.

Come risolvere il paradosso?

Ci sono state due “scuole” di pensiero a riguardo. La prima è già stata descritta. Il problema si ignora e quindi scompare. Ci concentreremo qui sul secondo punto di vista. In una serie di lavori, Cadek e Domini hanno mostrato come sia una pessima idea proporre che i fenomeni stereocinetici siano delle “illusioni” che non hanno nulla a che fare con il modo in cui il sistema visivo elabora le informazioni “serie” (non quelle illusorie da laboratorio psicologico), ovvero quelle fornite nel mondo visivo naturale. È una pessima idea pensare questo perché si possono mostrare tre cose: (1) la profondità percepita (e la rotazione percepita) che vengono riportate dagli osservatori nel fenomeno stereocinetico sono molto ben predicibili con un modello che si basa sull'informazione disponibile (come aveva fatto Musatti (1924)); (2) lo stesso modello spiega anche molto bene le percezioni che gli osservatori riportano quanto tutte le informazioni sono disponibili (e, quindi, quando in principio una soluzione “corretta” è possibile in base alla geometria inversa); (3) il modello usato per i fenomeni stereocinetici ben predice “le distorsioni sistematiche” che gli osservatori manifestano “in condizioni di visione naturali” (lasciamo da parte gli effetti stereocinetici), qualcosa che il Modello Standard non è in grado di fare. Per il principio del rasoio di Ockam, dunque, sembra che il Modello Standard sia in difficoltà. Ma così non è. E questa è la lezione più importante che nasce da questa “rivisitazione” di “un'illusione percettiva” che potrebbe sembrare avere un interesse solo “storico” – ma che in realtà è ben altro.

Il modello alternativo

Iniziamo ad esplorare in dettaglio e a fornire qui una spiegazione del modello proposto da Cadek e Domini per risolvere l'apparente paradosso del fenomeno stereocinetico. Le

informazioni retiniche offerte dagli stimoli stereocinetici si configurano come un “campo di velocità” istantaneo. Se analizzato localmente, può essere inteso come la somma di quattro variazioni elementari. Due di queste influenzano unicamente l’orientamento (‘curl’) e la dimensione (‘div’) di una “patch” proiettata; le altre due modificano la forma della proiezione bidimensionale e sono perciò denominate **shear**. Le due componenti di **shear** rappresentano le uniche trasformazioni dell’immagine che forniscono informazioni sulla forma tridimensionale e possono essere riassunte in un’unica grandezza denominata **def** (Domini & Caudek, 2003). Reiteriamo qui: la grandezza denominata “def” specifica, in termini locali, la configurazione affine dell’oggetto distale. Pertanto, essa non descrive la forma euclidea dell’oggetto. Questa distinzione è cruciale per comprendere che, attraverso “def”, otteniamo un’indicazione sulla struttura dell’oggetto che si basa su trasformazioni geometriche affine, le quali mantengono i punti, le rette e i piani, ma possono alterare le distanze e gli angoli, differendo quindi dalla descrizione euclidea che preserva le distanze e le proprietà metriche degli oggetti.

Esiste una relazione specifica tra la grandezza “def” e le proprietà della superficie distale locale che si intende ricostruire attraverso la geometria inversa: $def = \sigma \times \omega$, dove σ rappresenta l’orientamento locale istantaneo della superficie distale e ω indica la quantità di rotazione istantanea di tale superficie. I parametri σ e ω rappresentano le variabili incognite a livello distale, mentre “def” agisce come lo stimolo a livello prossimale. La questione emergente è: come è possibile determinare i valori di σ e ω a partire dalla conoscenza di “def”? La soluzione diretta attraverso la geometria inversa non è praticabile, in quanto σ e ω appartengono a una famiglia di soluzioni possibili caratterizzata da un unico parametro. Questo sottolinea la difficoltà nel dedurre le proprietà tridimensionali specifiche della superficie distale basandosi unicamente sulle informazioni fornite da “def”.

Domini and Caudek (2003) hanno proposto una soluzione “Bayesiana”: il sistema visivo sceglie i valori $\hat{\sigma}$ e $\hat{\omega}$ che hanno la plausibilità maggiore, data l’informazione fornita da

def. Ci vogliono delle informazioni a priori. Nella proposta di Domini and Caudek (2003), si assumono distribuzioni a priori uniformi per σ e ω . In tali circostanze, il valore più plausibile per entrambi i parametri è uguale a \sqrt{def} .

È facile fare una simulazione in Python che illustra questo punto. Simuliamo 100 valori *def* prodotti da valori “ragionevoli” dei parametri σ e ω . Aggiungiamo una leggera perturbazione alla “misurazione” di *def*.

```
N = 100

sigma_degrees = np.random.uniform(10, 80, size=N)
sigma_radians = np.deg2rad(sigma_degrees)
omega_degrees_per_second = np.random.uniform(0, 180, size=N)
omega_radians_per_second = np.deg2rad(omega_degrees_per_second)
def_values = sigma_radians * omega_radians_per_second
noise_std_dev = 0.1
def_values_perturbed = def_values + np.random.normal(0, noise_std_dev, def_values.shape)
```

Con i vincoli precedenti, il modello per stimare le distribuzioni a posteriori $p(\sigma \mid def)$ e $p(\omega \mid def)$ può essere scritto nel modo seguente in linguaggio Stan.

```
data {
  int<lower=0> N;          // Numero di misurazioni
  vector[N] def;          // Valori di deformazione
}

parameters {
  vector<lower=0, upper=1>[N] s;      // Slant per ogni misurazione
  vector<lower=0, upper=1>[N] omega;  // Velocità di rotazione locale per ogni misurazione
  real<lower=0> sigma;
}
```

```

model {
  // Priori
  s ~ beta(1, 1);
  omega ~ beta(1, 1);

  // Likelihood
  for (i in 1:N) {
    def[i] ~ normal(s[i] * omega[i], sigma); // Assumendo una certa varianza sigma
  }
}

```

In questa simulazione, effettuando un'analisi della correlazione tra la media a posteriori dei parametri σ e ω , calcolata sulla base della deformazione osservata, e il valore di deformazione inizialmente fornito, si ottiene un coefficiente di correlazione pari a 0.95 per entrambi i parametri. Questo risultato conferma quanto proposto da Domini and Caudek (2003), ovvero che \sqrt{def} agisce come il massimo a posteriori per entrambi i parametri incogniti, in accordo con i priori stabiliti.

Nel corso di diversi studi, Caudek e Domini hanno dimostrato che \sqrt{def} rappresenta un predittore affidabile della percezione tridimensionale derivata dal movimento, sia in contesti in cui è possibile una ricostruzione euclidea sulla base delle informazioni disponibili, sia in scenari dove è ammissibile soltanto una ricostruzione affine, come nel caso degli effetti stereocinetici.

bla bla un po' di citazioni

Una compiaciuta ignoranza

Passiamo ora al Modello Standard. Il Modello Standard si fonda su un principio concettualmente intuitivo, ancor più diretto rispetto al Modello Alternativo precedentemente discusso. Tale modello si avvale del concetto delle famiglie coniugate normale-normale, una nozione fondamentale nella statistica bayesiana che facilita il calcolo delle distribuzioni a posteriori quando sia la prior che la likelihood sono distribuzioni normali.

Quando si lavora con famiglie coniugate normale-normale, si assume che la prior per un parametro θ sia una distribuzione normale con media μ_0 e varianza σ_0^2 , e che la likelihood dei dati, dati θ , sia anch'essa normale con varianza nota σ^2 , ma con media che dipende da θ . Questa configurazione permette di esprimere la distribuzione a posteriori di θ , dati i dati osservati, come una nuova distribuzione normale, la cui media e varianza si calcolano secondo formule specifiche che integrano sia le informazioni a priori sia quelle derivanti dai dati.

La media a posteriori, μ_{post} , si ottiene come una somma ponderata tra la media della prior, μ_0 , e la media campionaria dei dati, \bar{x} , dove il peso è determinato dalle varianze della prior e dei dati. La formula per la media a posteriori è data da:

$$\mu_{post} = \frac{\frac{\sigma^2}{\sigma_0^2} \mu_0 + \frac{\sigma_0^2}{\sigma^2} \bar{x}}{\frac{\sigma^2}{\sigma_0^2} + \frac{\sigma_0^2}{\sigma^2}}$$

La varianza a posteriori, σ_{post}^2 , riflette l'incertezza residua su θ dopo aver osservato i dati. Si calcola come l'inverso della somma degli inversi delle varianze della prior e dei dati, ovvero:

$$\sigma_{post}^2 = \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \right)^{-1}$$

Nel contesto del Modello Standard, questi calcoli permettono di aggiornare in modo

efficiente le nostre credenze a posteriori su θ in seguito all'osservazione di nuovi dati. Nello specifico, il teorema normale-normale è stato utilizzato in questo modo per quel che riguarda la percezione della forma 3D. I prior costituiscono le informazioni sulla forma 3D fornite da un determinato “cue” sensoriale – ad esempio, la “structure-from-motion”. La likelihood è fornita dalle informazioni fornite da un secondo “cue”, ad esempio, la stereopsi. La distribuzione a posteriori è il prodotto di questo aggiornamento bayesiano che qui viene chiamato “cue integration”. Il modello fa delle chiare predizioni sulle proprietà della forma 3D che dovrebbero essere percepite, quando due “cue” sono presenti anziché uno solo, come indicato dalle formule precedenti.

Il Modello Standard ammette la possibilità che la percezione tridimensionale emergente dalla presenza di diversi segnali indiziari (o “cue”) possa non corrispondere fedelmente alla realtà, risultando invece sistematicamente alterata. Secondo questo modello, infatti, la stima a posteriori di una determinata dimensione dello stimolo, ad esempio la profondità, viene calcolata come una media pesata delle stime di quella dimensione ottenute presentando ogni segnale indiziario separatamente. I pesi per questa media sono determinati dall'incertezza, ossia dalla varianza, associata a ciascun segnale indiziario.

Tuttavia, una lacuna del Modello Standard risiede nella mancata specificazione della funzione di likelihood, ovvero nel come la forma tridimensionale viene dedotta dalle informazioni sensoriali. Spesso, il Modello Standard presume, esplicitamente o implicitamente, che la ricostruzione tridimensionale basata su ciascun segnale indiziario separatamente sia accurata. Se così fosse, l'unico errore deriverebbe dal rumore sensoriale nella misurazione dei segnali indiziari.

Questa assunzione, però, è falsa: numerose evidenze suggeriscono che i segnali indiziari sensoriali, quando esaminati isolatamente, portano a ricostruzioni tridimensionali sistematicamente distorte. Di conseguenza, definire con precisione la funzione di likelihood per ciascun segnale indiziario si rivela un compito cruciale, anziché un presupposto scontato.

Questa sfida, trascurata dal Modello Standard, è stata invece affrontata dal Modello Alternativo nel suo specifico ambito di applicazione, ovvero nello studio dei fenomeni stereocinetici o, in altre parole, nella struttura derivante dal movimento. Con un pizzico di ironia, si potrebbe sintetizzare affermando che, mentre il Modello Standard presuppone una percezione tridimensionale sempre fedele alla realtà, il Modello Alternativo considera tale percezione come intrinsecamente “illusoria”.

Methods

—>

References

- Caudek, C., & Proffitt, D. R. (1993). Depth perception in motion parallax and stereokinesis. *Journal of Experimental Psychology: Human Perception and Performance*, 19(1), 32–47.
- Domini, F., & Caudek, C. (2003). 3-d structure perceived from dynamic information: A new theory. *Trends in Cognitive Sciences*, 7(10), 444–449.
- Musatti, C. L. (1924). Sui fenomeni stereocinetici. *Archivio Italiano Di Psicologia*, 3, 105–120.
- Proffitt, D. R., Rock, I., Hecht, H., & Schubert, J. (1992). Stereokinetic effect and its relation to the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 3–21.