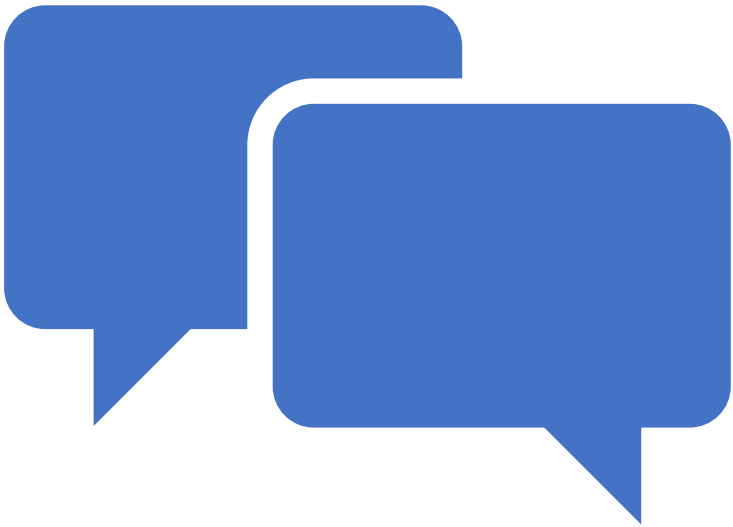


# Does discussion make crowds any wiser?

H. Mercier, N. Claidière (2022)



Presentazione di:

Augello Chiara, Domanti Umberto, Leone Claudia, Savoia Alice

# Scopo della ricerca

Testare se la discussione in grandi gruppi migliora o ostacola la saggezza delle folle, avvicinandole alla risposta corretta in tre tipi di problemi

Dimostrativi

Fattuali

Etici

# Definizioni

- **Folla:** qualsiasi grande gruppo, non necessariamente coeso, di individui che possono anche non conoscersi o non condividere le medesime idee
- **Saggezza delle folle:** capacità di grandi gruppi di fornire risposte molto accurate, raggiungibili attraverso meccanismi di aggregazione (voto di maggioranza, media)



# Background teorico

Partendo da un esempio di decisione democratica senza discussione (antica Atene) ci si è chiesti se la discussione potesse ostacolare la saggezza delle folle.

Infatti, ad Atene, le decisioni venivano prese tramite voto di maggioranza poiché la discussione non era formalmente ammessa.

# Background teorico

---

*Condorcet Jury Theory* (Condorcet, 1785): per una scelta dicotomica, un gruppo di individui dovrebbe raggiungere una decisione tramite il voto di maggioranza

*Data  $p$  come la probabilità che ogni singolo votante selezioni l'opzione migliore, la probabilità che la maggioranza voti per l'opzione migliore supera  $p$  e si avvicina ad 1, man mano che il numero dei votanti aumenta*

# Evidenze contro la discussione

---

- Discussione + aggregazione tramite maggioranza:  
perdita indipendenza dei singoli per influenza reciproca
- Discussione + aggregazione tramite media  
nonostante l'errore della media sia inferiore rispetto all'errore medio individuale, la convergenza su una risposta intermedia porta all'eliminazione delle opinioni più estreme e quindi ad una riduzione di diversità e indipendenza dei singoli
- Decremento di alcuni indicatori della saggezza delle folle (es. grado di diversità delle risposte), a fronte di un miglioramento della performance media dei singoli partecipanti (Lorenz et al., 2011)

# Evidenze a favore della discussione (piccoli gruppi)

---

- Effetto positivo della discussione in piccoli gruppi (N=5), in un tempo limitato (1 min): aumento delle prestazioni individuali e della correttezza della risposta raggiunta tramite la discussione (Navajas, Niella, Garbulsky, Bahrami, & Sigman, 2018)
- Miglioramenti delle prestazioni medie dei membri di piccoli gruppi su un'ampia gamma di problemi, da compiti logici a previsioni politiche (Mellers et al., 2014)

# Mellers et al. (2014) - Psychological Strategies for Winning a Geopolitical Forecasting Tournament

Torneo di previsione di 199 eventi geopolitici in 2 anni

## ANNO 1

2.246 pp (gruppi max 25 membri)

disegno 3 (formazione: nessuna, su scenari, sulle probabilità) x 4 (influenza di gruppo: indipendente, credenza di gruppo, team, previsioni di mercato)

85 domande

## ANNO 2

1.648 pp (gruppi max 15 membri)

disegno 2 (formazione: nessuna, di probabilità) x 3 (influenza di gruppo: indipendente, team, previsioni di mercato)

114 domande



# Mellers et al. (2014) - Psychological Strategies for Winning a Geopolitical Forecasting Tournament

## Variabili indipendenti

### Formazione:

- **scenari:** generazione di nuovi futuri, intrattenere attivamente più possibilità, utilizzare alberi decisionali
- **probabilità:** considerare classi di riferimento, evitare trappole nel giudizio, fare una media di più stime
- **nessuna:** controllo

### Influenza di gruppo:

- **no interazione/indipendenza:** i previsori lavoravano da soli
- **interazione/team:** i previsori discutevano insieme sin da subito formulando un'unica risposta di gruppo
- **credenza di gruppo:** i previsori lavoravano da soli conoscendo le opinioni degli altri membri
- **previsioni di mercato**

# Mellers et al. (2014) - Psychological Strategies for Winning a Geopolitical Forecasting Tournament

## Esempio di quesito sulla previsione degli esiti geopolitici

"L'italiano Silvio Berlusconi si dimetterà, perderà il voto di rielezione/fiducia o comunque lascerà l'incarico prima del 1° gennaio 2012?"

I partecipanti hanno predetto la probabilità che l'evento si sarebbe verificato, su scala da 0% (certo che non si sarebbe verificato) a 100% (certo che si sarebbe verificato).

VD: accuratezza individuale (punteggio Brier medio)

0 → punteggio migliore

2 → punteggio peggiore

# Mellers et al. (2014) - Psychological Strategies for Winning a Geopolitical Forecasting Tournament

## Risultati:

- I previsori nella condizione "team" sono stati più accurati rispetto alla condizione "credenza di gruppo"  $t(1036) = 5.52, p < .001$
  - I previsori nella condizione "credenza di gruppo" sono stati più accurati rispetto alla condizione "indipendente"  $t(1120) = 5.92, p < .001$
- la maggiore accuratezza dei gruppi è dovuta alla comunicazione che ha permesso ai partecipanti di motivarsi a vicenda e discutere problemi e ragioni (altruismo cognitivo)

# Evidenze a favore della discussione (gruppi medi)

---

In gruppi di medie dimensioni (N= da 11 a 25) la discussione portava:

- ad un miglioramento più rapido delle prestazioni rispetto al ragionamento individuale
- ad un miglioramento della capacità del gruppo di selezionare l'alternativa migliore

Limiti: dimensioni modeste del gruppo, inadeguatezza dei problemi per gruppi medi, utilizzato solo il voto maggioranza.

# Does discussion make crowds any wiser?

---

Notte Europea dei Ricercatori

---

N= 1.958 (1.048 femmine)

---

Età compresa tra i 13 e i 60 anni  
(Mediana= 24)

---

33 gruppi da 20 a 208 individui  
(media= 58)

Metodo - Partecipanti

# Metodo - Materiali

---

Sono stati usati 3 tipi di problemi:

- 2 problemi dimostrativi
- 2 problemi fattuali
- 2 problemi etici



Totale= 6 problemi

# Problemi dimostrativi

---

Hanno una risposta corretta che può essere raggiunta tramite la conoscenza condivisa del gruppo

## Problema dimostrativo 1 – Paul e Linda

Paul guarda Linda; Linda guarda John; Paul è sposato; John non è sposato; Qualcuno è sposato e guarda qualcuno che non è sposato?

Risposte: Sì [corretto] / No / Non possiamo dirlo

## Problema dimostrativo 2 – Mazza e palla

Una caramella e una baguette costano insieme 1,10€. La baguette costa 1€ in più della caramella. Quanto costa la caramella? Risposta corretta: 0.05€

# Problemi fattuali

---

Sono problemi che hanno come risposta un dato oggettivo

## Problema fattuale 1 – Coppa del mondo

Quanti goal sono stati segnati nella Coppa del Mondo di calcio del 2010? Risposta corretta: 145

## Problema fattuale 2 – Ascensori

Quanti ascensori ci sono nell'Empire State Building di New York?  
Risposta corretta: 73



# Problemi etici

---

Non hanno una risposta corretta. Sono stati utilizzati come condizione di controllo (risultati non discussi dagli autori poiché non è possibile trarre una conclusione forte)

## Problema etico 1 – Mignolo

Quanto denaro dovrebbe essere stanziato per risarcire chi ha perso un mignolo in un incidente sul lavoro?

## Problema etico 2 – Vermi

Quanto denaro dovrebbe essere assegnato per risarcire qualcuno che scopre di aver mangiato vermi nel suo pasto al ristorante?

# Ipotesi

---

- H1a. Per i problemi dimostrativi, la discussione migliora le prestazioni più del pensiero solitario
- H1b. Per i problemi fattuali, la discussione migliora le prestazioni più del pensiero solitario
- H1c. Per i problemi etici, la discussione non ha un impatto maggiore del pensiero solitario
  
- H2. Per i problemi dimostrativi, la discussione porta a migliori risposte aggregate attraverso la votazione a maggioranza
  
- RQ1. Per i problemi fattuali, la discussione porta a risposte aggregate peggiori, equivalenti o migliori, selezionate attraverso la media?

# Procedura

---

Griglia quadrata

Fogli di risposta con 15 righe → 1 riga per ogni fase temporale

Problema visualizzato su un grande schermo

2 fasi within:

**Fase silenzio** (tot 5 risposte): dopo 20 sec 1° risposta  
a intervalli di 1 min altre 4 risposte

**Fase discussione** (tot 10 risposte): a intervalli di 1 min altre 10 risposte  
consensuali per il gruppo (8 vicini)

## Escluso un totale dell'11% delle risposte

- < 1% Paul e Linda (escluse risposte che non corrispondevano a nessuna delle 3 opzioni di risposta)
- 6% Mazza e Palla (escluse le risposte che non erano di 0.05 e 0.10 €)
- 7% e 12% Ascensori e Coppa del Mondo (escluse le risposte non numeriche o estreme)

# Codifica risposte (problemi dimostrativi)

---

Paul e Linda:

1 alternativa corretta

0 alternative errate

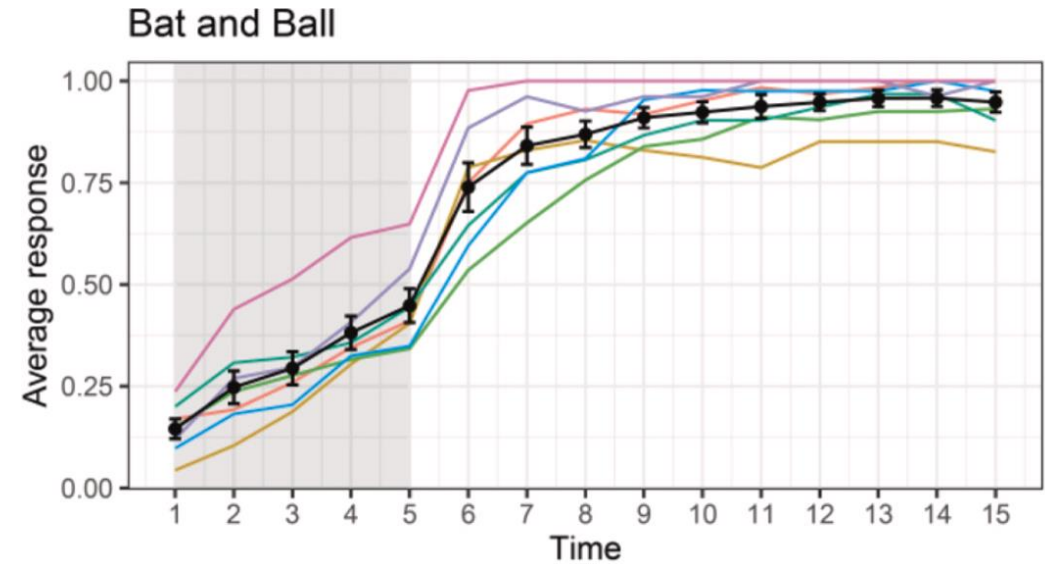
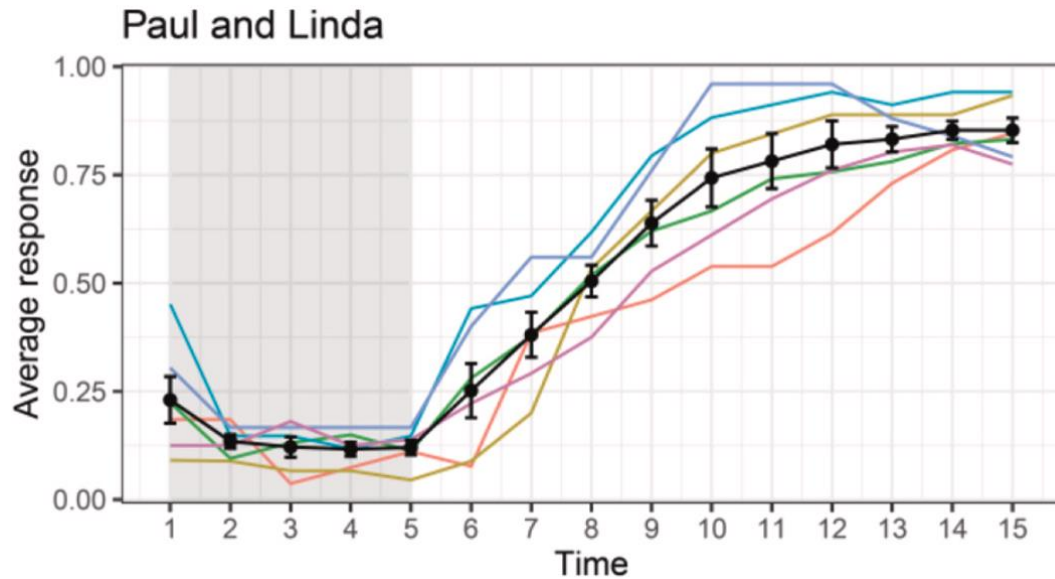
Mazza e palla:

1 alternativa corretta

0 alternative errate

# Risultati – H1a

Per i due problemi dimostrativi, la discussione favorisce la diffusione alla risposta corretta ( $\beta = 0.38$ ,  $SE = 0.04$ ,  $z = 8.37$ ,  $p < .001$ )

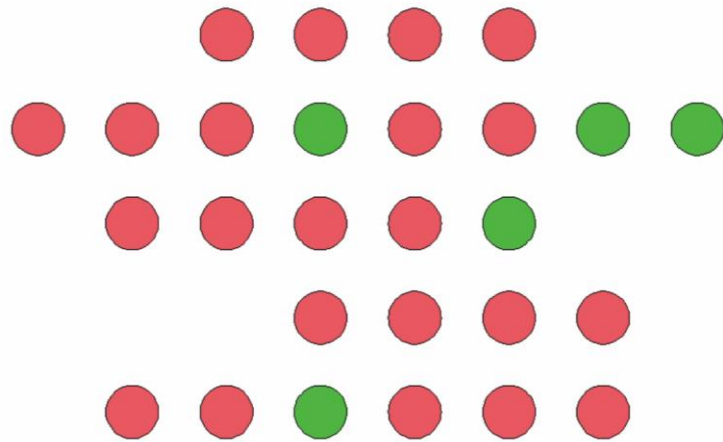


Fase silenzio in grigio. Fase discussione in bianco. Ogni linea colorata rappresenta la media delle risposte di ogni gruppo. La linea nera rappresenta la media tra i gruppi.

# Video - problemi dimostrativi

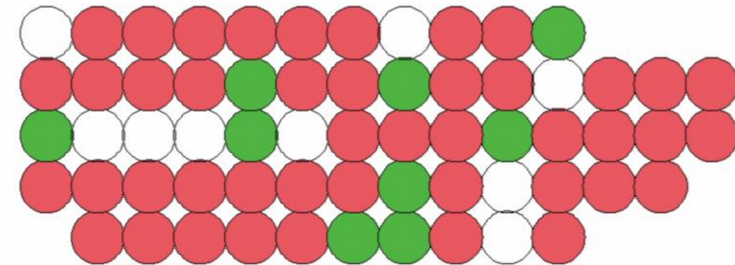
- risposta errata
- risposta corretta
- dati esclusi

Paul & Linda: Silence phase



Time step 1

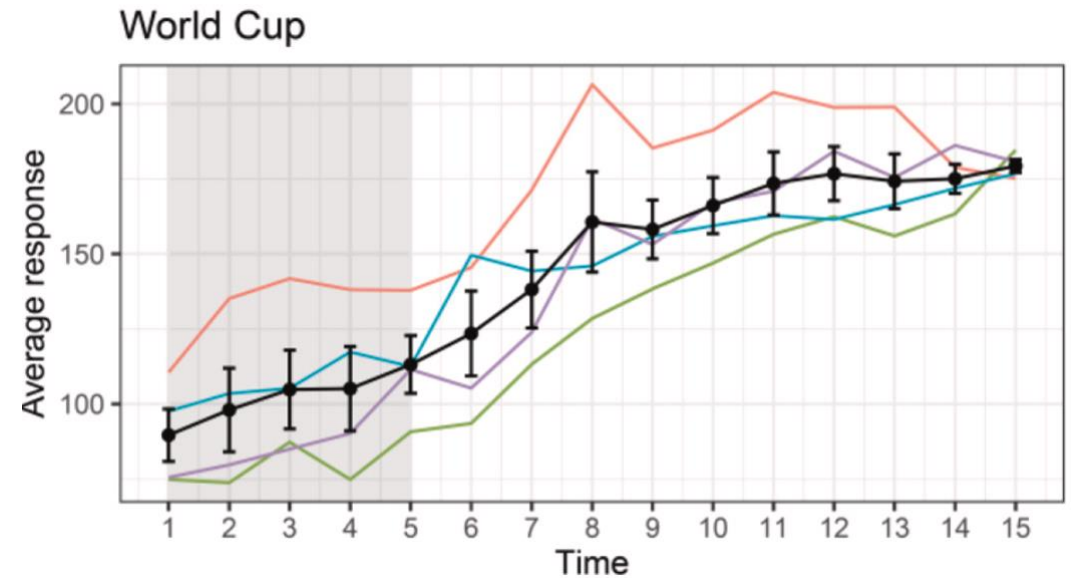
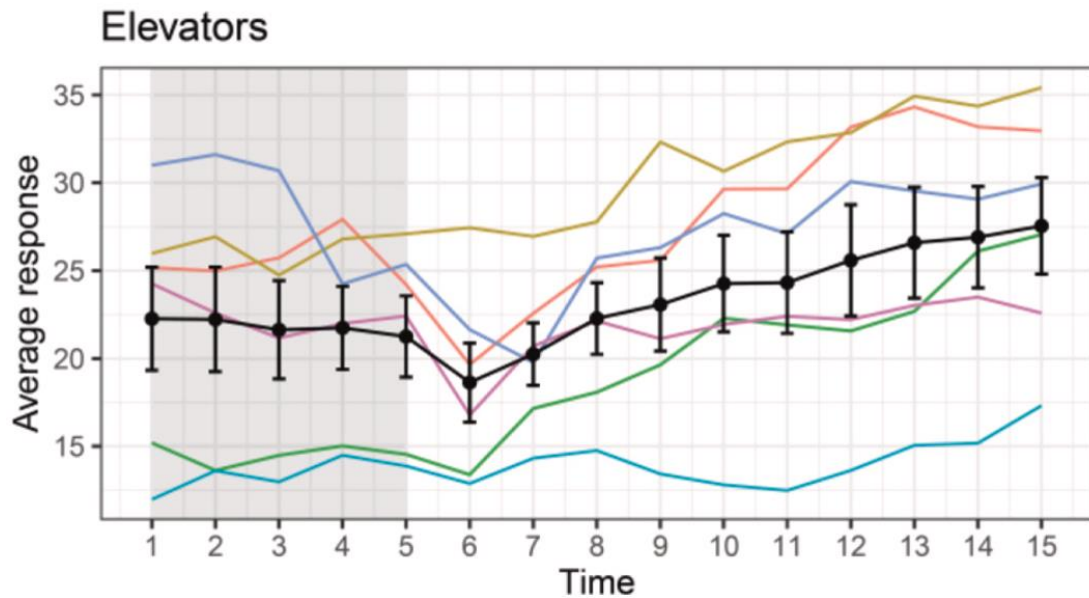
Bat & Ball: Silence phase



Time step 1

# Risultati – H1b

Per i due problemi fattuali si è osservata una riduzione della distanza dalla risposta corretta solo nella fase di discussione ( $\beta = -2.31$ ,  $SE = 0.74$ ,  $df = 6586$ ,  $t = -3.12$ ,  $p = .002$ )



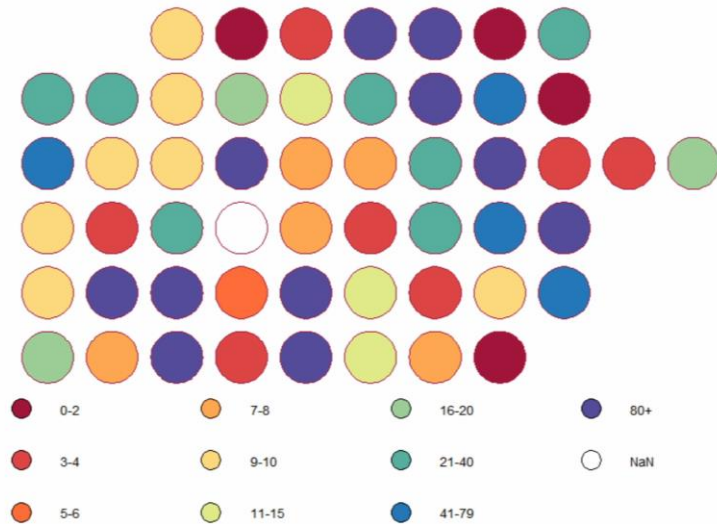
Fase silenzio in grigio. Fase discussione in bianco. Ogni linea colorata rappresenta la media delle risposte di ogni gruppo. La linea nera rappresenta la media tra i gruppi.



# Video-problemi fattuali

Risposta corretta: 73

Elevator: Silence phase



Time step 1

Risposta corretta: 145

World Cup: Silence phase



Time step 1

## Risultati – H2

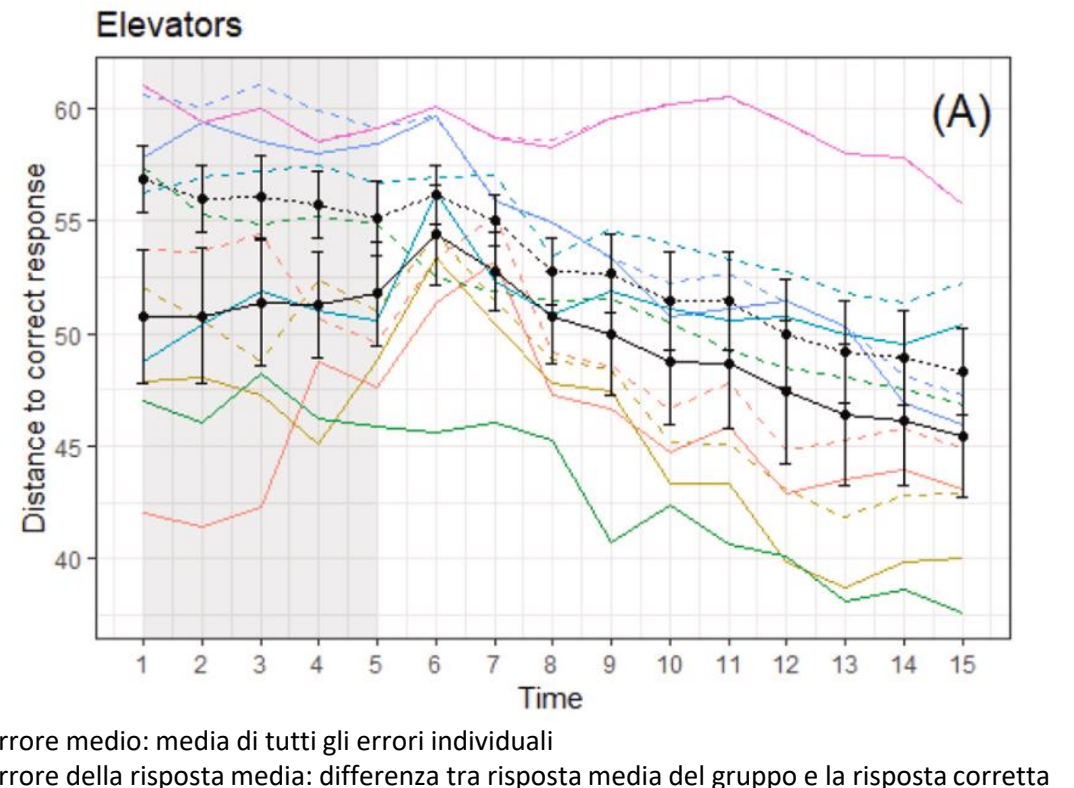
Per i problemi dimostrativi la discussione ha portato a migliori risposte aggregate (voto di maggioranza), rispetto la fase di silenzio.

Tutti i gruppi hanno avuto una maggioranza di risposte corrette alla fine della fase discussione ( $\chi^2$  McNemar = 9.10, df= 1,  $p= .003$ )

# Risultati – RQ1

Nel problema ascensori l'errore della risposta media è diminuito. Tutti e sei i gruppi avevano un errore della risposta media inferiore alla fine della fase discussione rispetto alla fase di silenzio ( $p=0.03$ ).

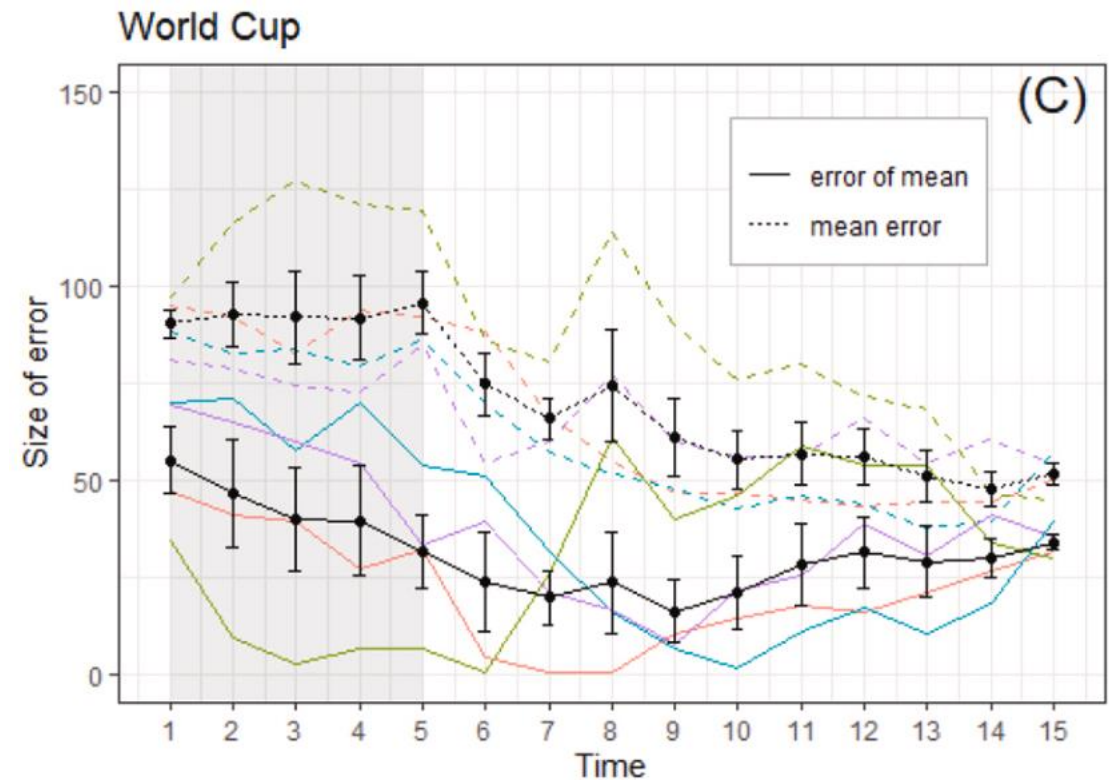
La discussione ha migliorato sia le risposte individuali (diminuzione errore medio) che quelle aggregate attraverso la media (diminuzione dell'errore della risposta media) all'interno di ciascun gruppo



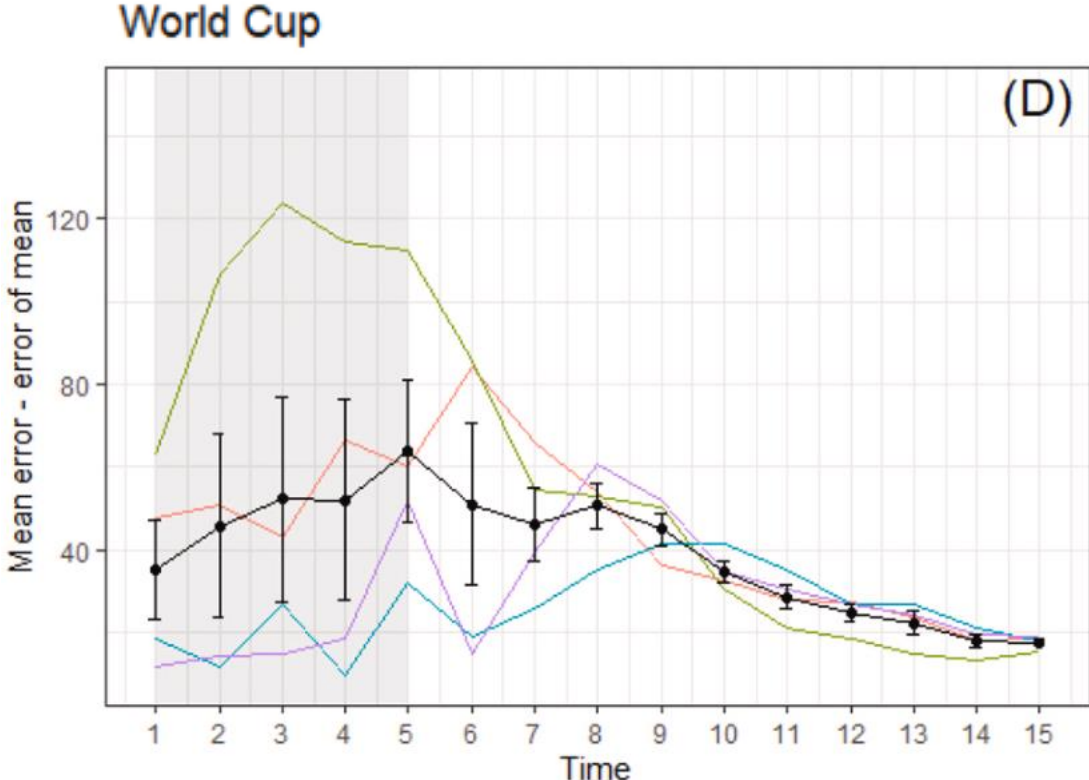
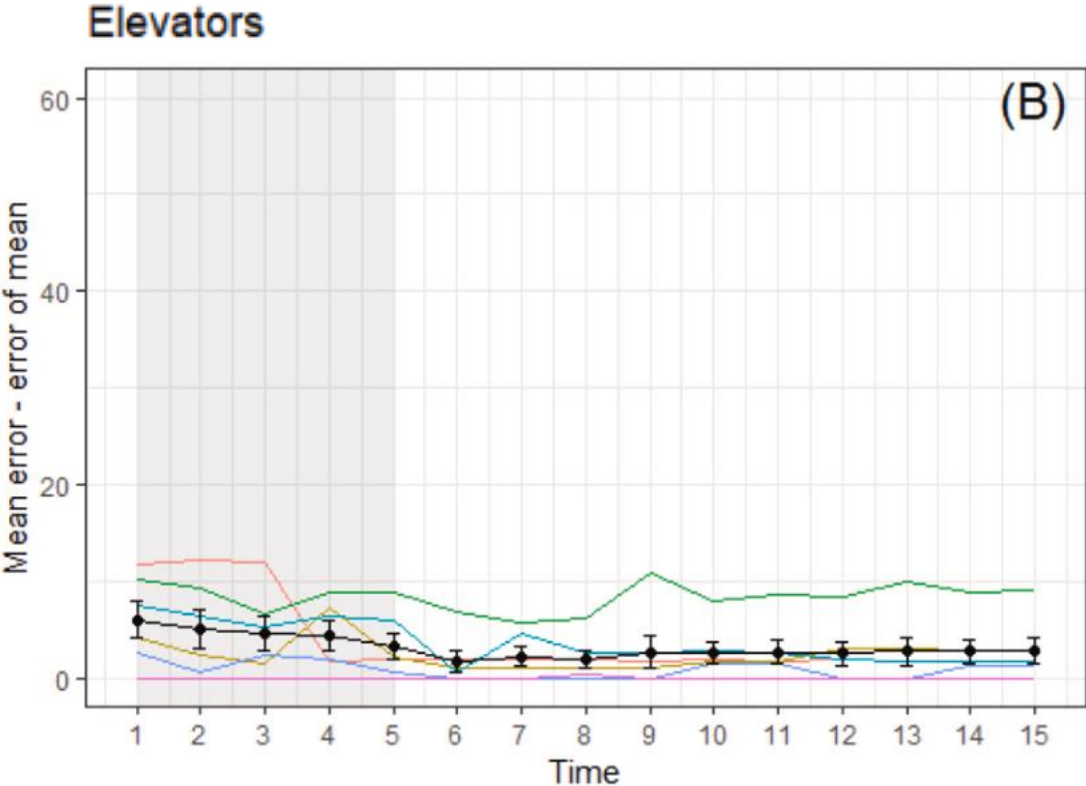
# Risultati – RQ1

Per il problema coppa del mondo, non ci sono evidenze di una diminuzione dell'errore della risposta media.

La discussione ha prodotto un miglioramento sulle risposte individuali (diminuzione dell'errore medio) ma non sulle risposte aggregate attraverso la media.



Nel problema ascensori la differenza tra l'errore medio e l'errore della risposta media rimane costante, mentre nel problema coppa del mondo è diminuita nella fase di discussione.



# Conclusioni

- ✓ La discussione ha sempre migliorato le prestazioni per i problemi dimostrativi e fattuali (vs silenzio) → H1a, b confermate
  - ✓ Per i problemi dimostrativi la discussione ha migliorato anche le risposte aggregate ottenute tramite il voto di maggioranza → H2 confermata
  - ✓ Per i problemi fattuali la discussione ha migliorato le risposte aggregate tramite la media solo nel caso del problema ascensori → RQ1 confermata parzialmente
- La discussione porta comunque ad un miglioramento delle risposte

# Conclusioni

La discussione porta ad un miglioramento delle risposte anche a livello qualitativo

- Problemi dimostrativi:

Paul e Linda: da 12% di risposte corrette a 84%

Mazza e Palla: da 41% di risposte corrette a 91%

- Problemi fattuali:

Ascensori: da 55 a 48 errore medio

Coppa del mondo: da 96 a 52 errore medio

# Limiti e prospettive future

---

- Numero limitato di persone con cui i partecipanti potevano discutere
- Solo città francesi
  
- Replicare con culture collettiviste
- Analizzare l'effetto della discussione in gruppo sui dilemmi sociali e morali
- Considerare altri mezzi di comunicazione



# Considerazioni finali

La **discussione** è uno strumento solido per migliorare non solo le **risposte individuali** ma anche **collettive**, anche in gruppi grandi e diversificati, almeno per problemi che hanno una risposta corretta.



# Bibliografia

---

Claidière, N., Trouche, E., & Mercier, H. (2017). Argumentation and the diffusion of counter-intuitive beliefs. *Journal of Experimental Psychology: General*, 146(7), 1052–1066.

---

Condorcet. (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. *L'imprimerie royale*.

---

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020–9025.

---

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., , Swift, S. A., et al. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, 25(5), 1106–1115.

---

Mercier, H., & Claidière, N. (2022). Does discussion make crowds any wiser?. *Cognition*, 222, 104912.

---

Navajas, J., Niella, T., Garbulsky, G., Bahrami, B., & Sigman, M. (2018). Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour*, 2(2), 126.

[https://it.wikipedia.org/wiki/Saggezza\\_della\\_folla](https://it.wikipedia.org/wiki/Saggezza_della_folla)

# Grazie per l'attenzione!

- Augello Chiara
- Domanti Umberto
- Leone Claudia
- Savoia Alice

