IAFPA 2022

The impact of reflection and retention intervals on earwitness accuracy: two experiments

Francis Nolan, Nikolas Pautz, Kirsty McDougall, Katrin Müller-Johnson, Harriet M.J. Smith and Alice Paver

IVIP: Improving Voice Identification Procedures

IVIP

Project Team

University of Cambridge:
- PI: Kirsty McDougall (phonetics)
- Francis Nolan (phonetics)
- Alice Paver (phonetics)

Nottingham Trent University:
- Nikolas Pautz (psychology)
- Natalie Braber (sociolinguistics)
- David Wright (sociolinguistics)
- Harriet Smith (psychology)

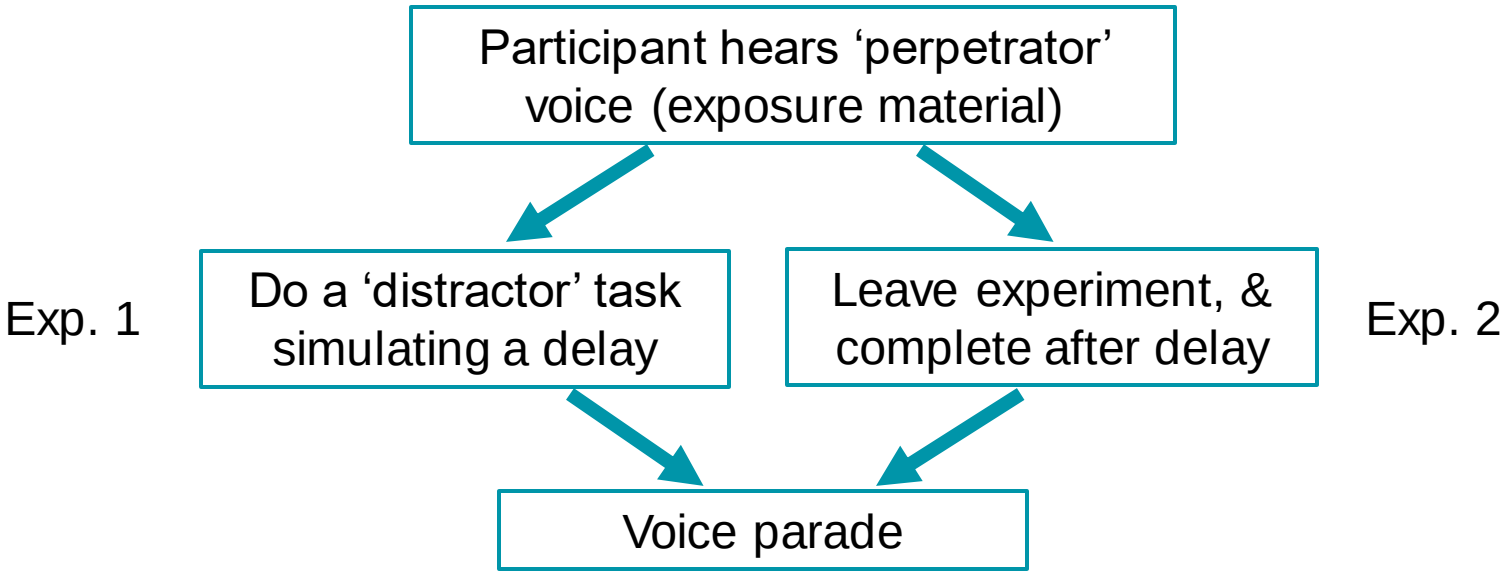
De Montfort University:
- Jeremy Robson (law)

University of Oxford:
- Katrin Müller-Johnson (criminology, psychology)

UK ESRC Grant Ref: ES/S015965/1

UKRI Economic and Social Research Council

Simulating an earwitness situation



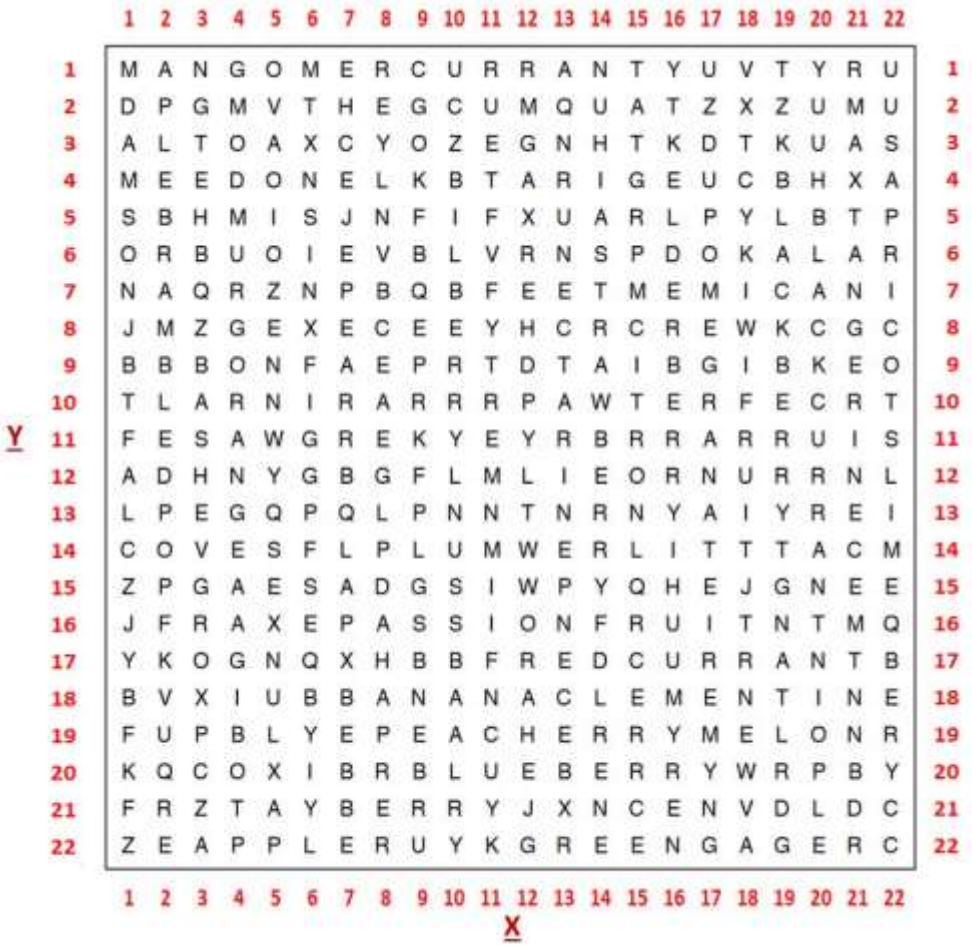
- Yet, in crime situations, if a witness realises they have heard a perpetrator, they're likely to think back over the event
- **RQ: Does explicit post-encoding (post target-exposure) reflection improve voice recognition accuracy?**

Experiment 1: distractor task, no delay

	target present	target absent
with reflection	✓	✓
no reflection	✓	✓

3 separate parades, each with its own target (if present), and 3 separate sets of foils, were used; but results will be combined here

Word-spotting distractor task



Y	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	M	A	N	G	O	M	E	R	C	U	R	R	A	N	T	Y	U	V	T	Y	R	U
2	D	P	G	M	V	T	H	E	G	C	U	M	Q	U	A	T	Z	X	Z	U	M	U
3	A	L	T	O	A	X	C	Y	O	Z	E	G	N	H	T	K	D	T	K	U	A	S
4	M	E	E	D	O	N	E	L	K	B	T	A	R	I	G	E	U	C	B	H	X	A
5	S	B	H	M	I	S	J	N	F	I	F	X	U	A	R	L	P	Y	L	B	T	P
6	O	R	B	U	O	I	E	V	B	L	V	R	N	S	P	D	O	K	A	L	A	R
7	N	A	Q	R	Z	N	P	B	Q	B	F	E	E	T	M	E	M	I	C	A	N	I
8	J	M	Z	G	E	X	E	C	E	E	Y	H	C	R	C	R	E	W	K	C	G	C
9	B	B	B	O	N	F	A	E	P	R	T	D	T	A	I	B	G	I	B	K	E	O
10	T	L	A	R	N	I	R	A	R	R	R	P	A	W	T	E	R	F	E	C	R	T
11	F	E	S	A	W	G	R	E	K	Y	E	Y	R	B	R	R	A	R	R	U	I	S
12	A	D	H	N	Y	G	B	G	F	L	M	L	I	E	O	R	N	U	R	R	N	L
13	L	P	E	G	Q	P	Q	L	P	N	N	T	N	R	N	Y	A	I	Y	R	E	I
14	C	O	V	E	S	F	L	P	L	U	M	W	E	R	L	I	T	T	T	A	C	M
15	Z	P	G	A	E	S	A	D	G	S	I	W	P	Y	Q	H	E	J	G	N	E	E
16	J	F	R	A	X	E	P	A	S	S	I	O	N	F	R	U	I	T	N	T	M	Q
17	Y	K	O	G	N	Q	X	H	B	B	F	R	E	D	C	U	R	R	A	N	T	B
18	B	V	X	I	U	B	B	A	N	A	N	A	C	L	E	M	E	N	T	I	N	E
19	F	U	P	B	L	Y	E	P	E	A	C	H	E	R	R	Y	M	E	L	O	N	R
20	K	Q	C	O	X	I	B	R	B	L	U	E	B	E	R	R	Y	W	R	P	B	Y
21	F	R	Z	T	A	Y	B	E	R	R	Y	J	X	N	C	E	N	V	D	L	D	C
22	Z	E	A	P	P	L	E	R	U	Y	K	G	R	E	E	N	G	A	G	E	R	C

X

Try and find as many words as you can in 5 minutes.

When you find a word, type the coordinates (X,Y) of the **first** letter in brackets after the word.

Please use a semi-colon (;) to separate your answers.

For example: APPLE (3,22); MANGO (1,1); .....

Speakers

- Targets & foils from *DyViS* database (Nolan *et al.* 2009)
  - male, Standard Southern British English, aged 18-25 years
- 3 groups of 15 speakers, 1 per target
- Targets and foils chosen as in McDougall et al. 2021 IAFPA:
  - Listeners judged similarity of all pairings in a group of 15
  - Multidimensional scaling > pseudoperceptual space
  - 9 foils (for target-absent) or 8 (for target-present) speakers sounding most similar to the target selected from each group of 15

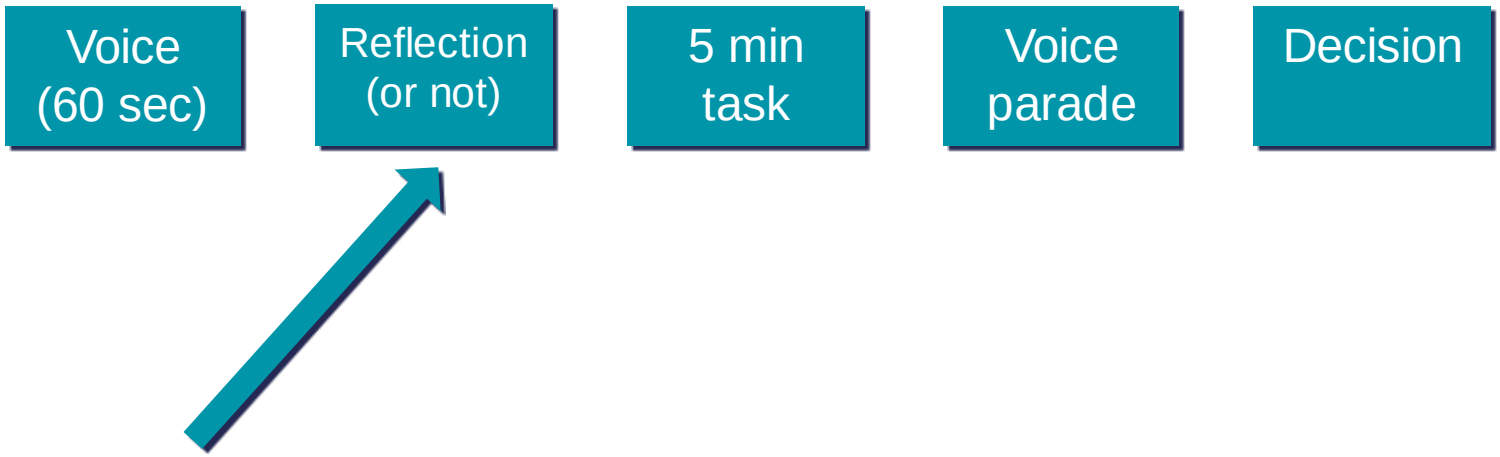
Speech material

- Exposure material for target voice: 60s sample from telephone call (‘perpetrator’ side; studio quality)
- Parade samples: 15s samples from simulated police interview task, using collage technique of Home Office guidelines
- Experiment conducted online using Gorilla

Participants

- N = 180 participants recruited via Prolific (randomly assigned to 1 of 3 targets)
  - born in and lived most of their pre-18 lives in England
  - 1st language English
  - No hearing loss or hearing difficulties
  - 88 male, 92 female, aged 18-40 years (M= 27.72, SD = 6.4)
  - Minimum approval rating of 90% on Prolific

Procedure

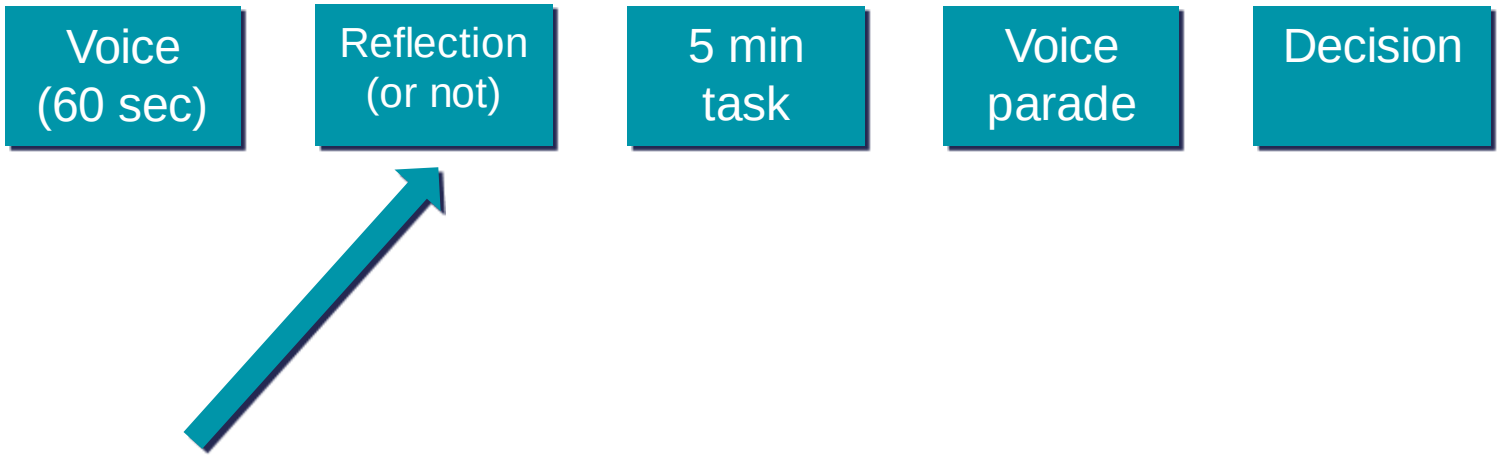


Reflection condition:

"Imagine that the voice you have just heard is that of a criminal. You may be asked by the police to make an identification some time in the future. Take a few moments now to reflect on the voice."

(20 seconds)

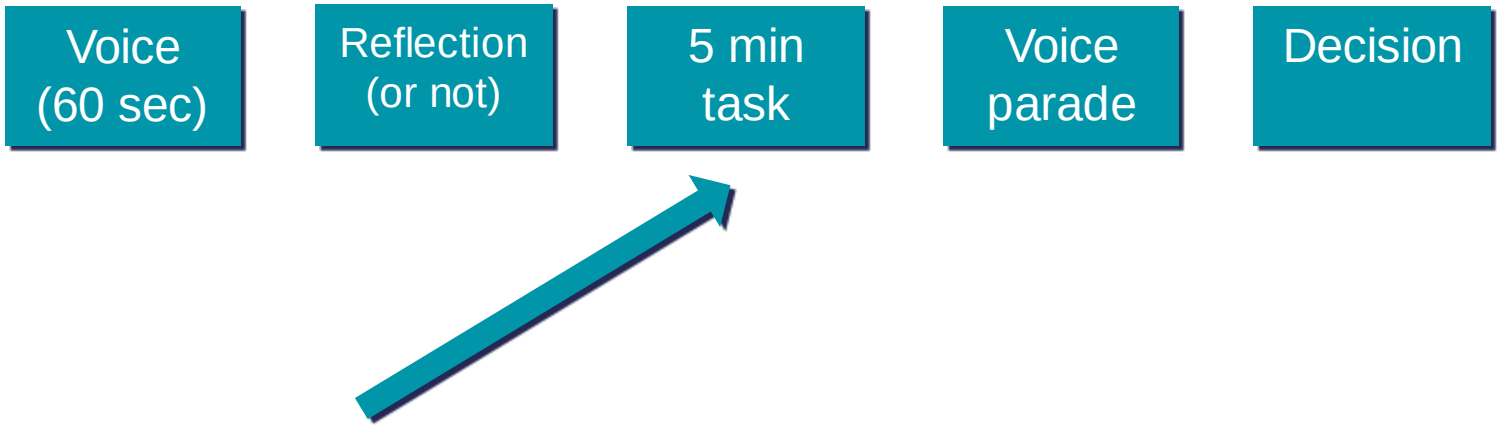
Procedure



No reflection condition:

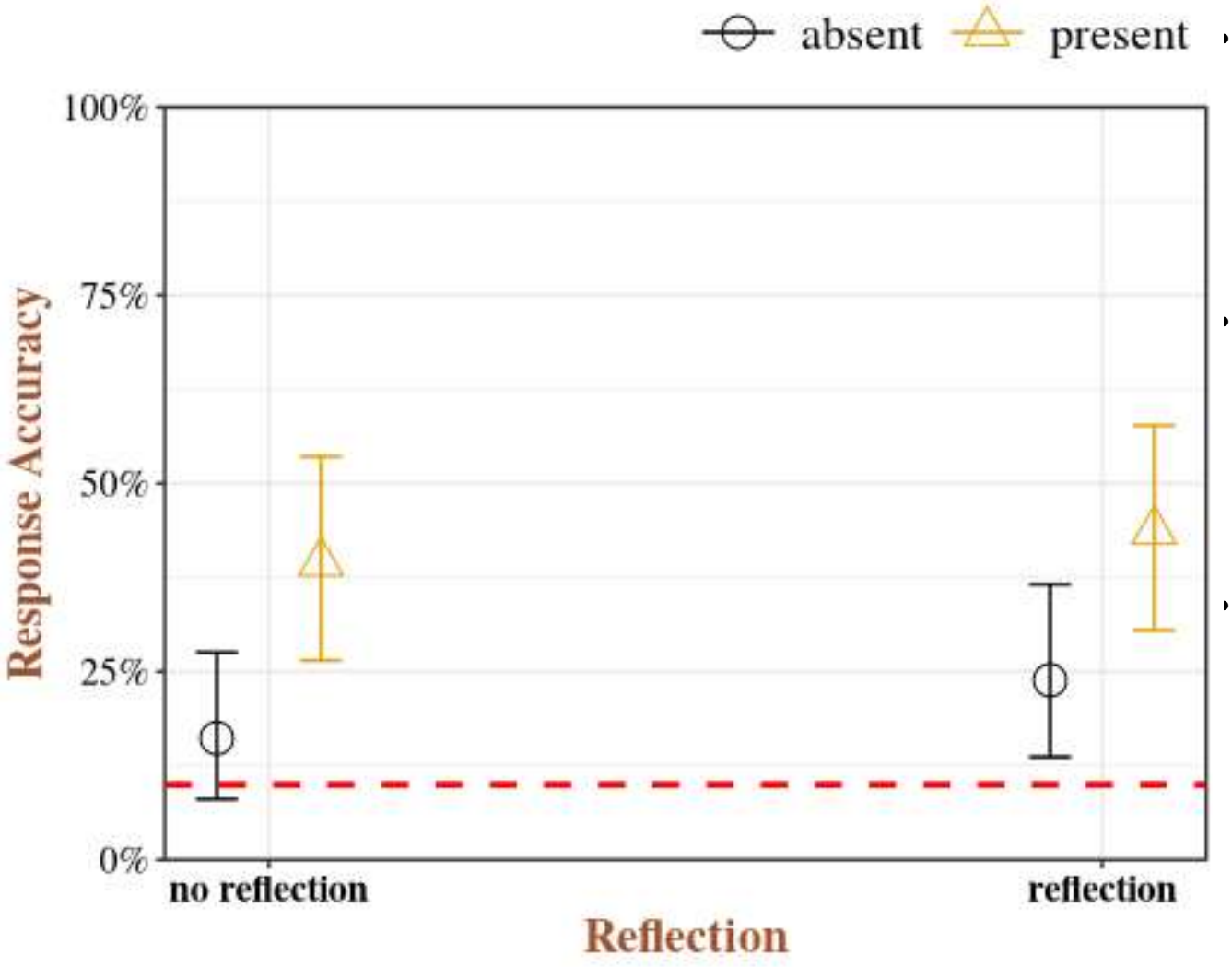
Simple attention task – participants push space bar when + appears on screen, several times
(20 seconds)

Procedure



Retention interval:
- 5 min distractor task
(word search, accompanied by lobby noise)
- exceeds short-term memory capacity; relies on long-term-memory

Results: effect of reflection (before distraction task)



- Poor accuracy percentages overall, but above chance (except in target-absent/no reflection)
- Target-present parades give best performance (consistent with previous findings, e.g. Smith et al. 2020)
- **No meaningful differences in performance between with- and without-reflection** (Bayesian mixed models)

Experiment 2: overnight delay

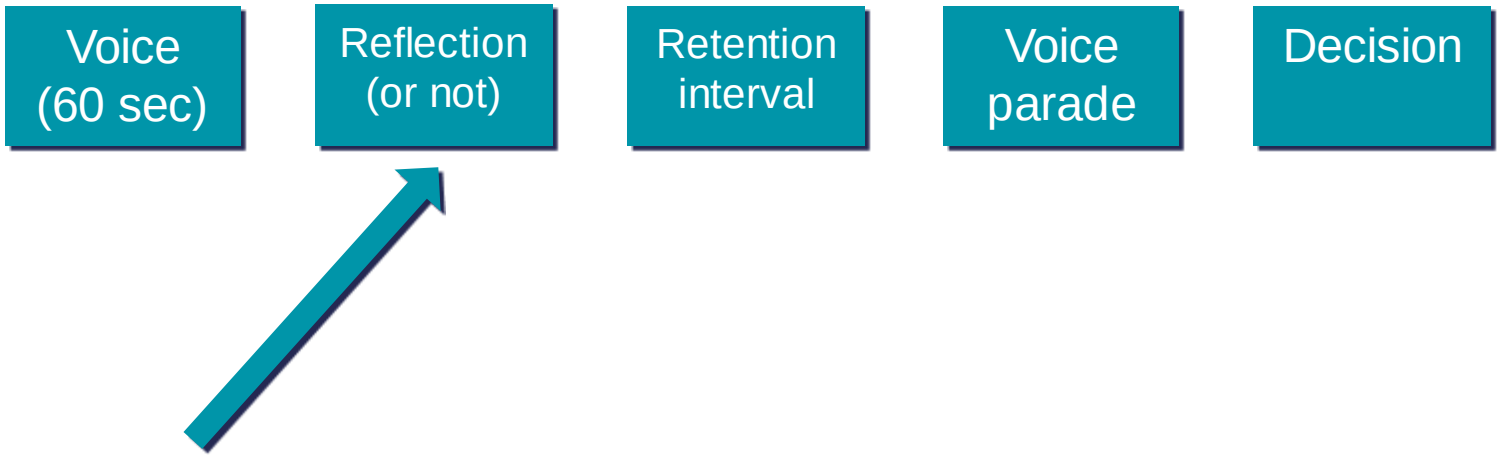
- Same design as Experiment 1, but 20-28 hour retention interval between exposure and parade instead of distractor task
- Participants (same recruitment and requirements):
  - N = 181
  - 87 male, 93 female
  - aged 18-40 years (M= 27.97, SD = 6.01)

Experiment 2: overnight delay, no distraction task

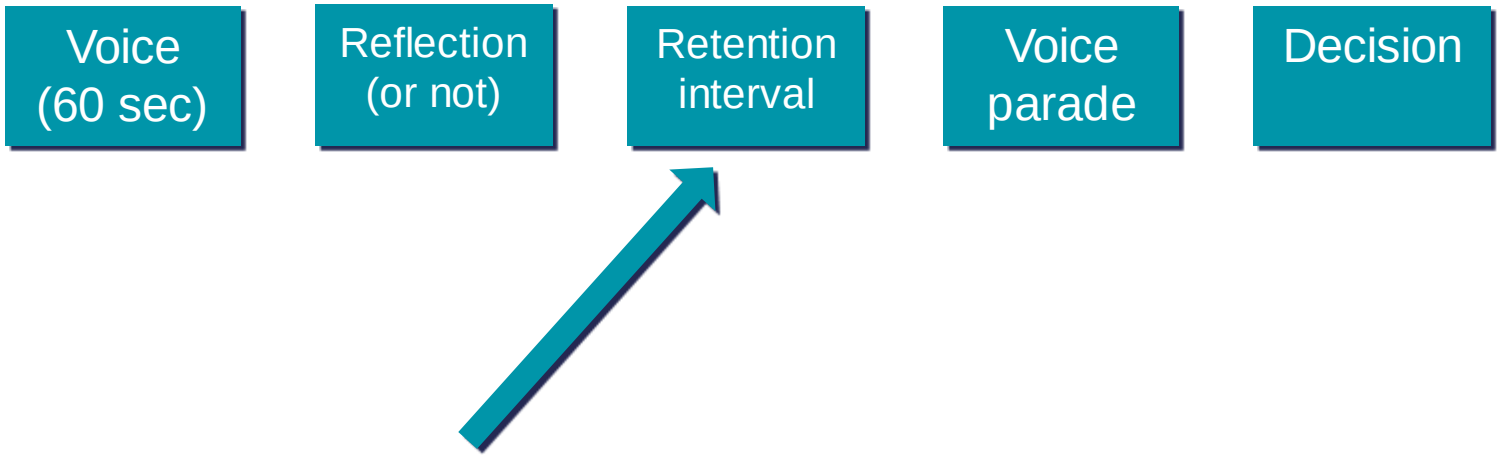
	target present	target absent
with reflection	✓	✓
no reflection	✓	✓

3 separate parades, each with its own target (if present), and 3 separate sets of foils, were used; but results will be combined here

Procedure

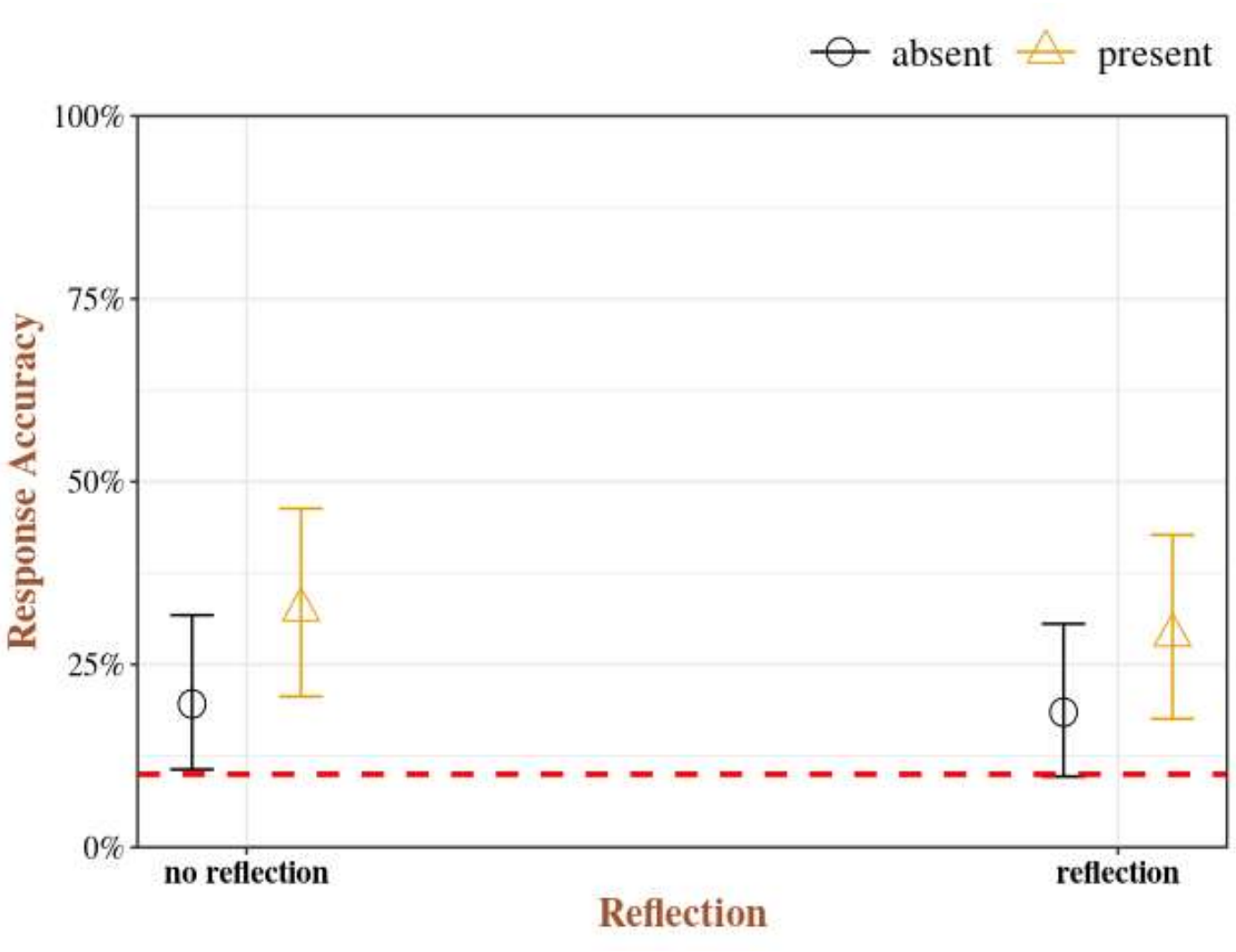

Voice
(60 sec)
Reflection
(or not)
Retention
interval
Voice
parade
Decision
Same reflection/no reflection manipulation

Procedure



20-28 hour delay between exposure and parade instead of distractor task

Results: effect of reflection (before overnight delay)



Again, poor accuracy percentages overall,

Again, target-present parades give best performance

No meaningful differences in performance between with and without reflection (Bayesian mixed models)

No interaction btw target presence and reflection

Discussion of ‘reflection’ (1)

- The motivation for ‘reflection’ was that the immediate cognitive load of the word-finding task might hinder memory consolidation
- Disappointingly, neither with word-finding simulating a delay, nor with an actual overnight delay, did a period of reflection improve scores

Discussion of ‘reflection’ (2)

- However, we don’t think this closes the case on ‘reflection’:
  - our period of reflection was very short (20 seconds)
  - it did not allow for repeated ‘rehearsal’ of the auditory memory, as might happen in a real event
  - we had no check on whether online participants actually reflected on the target voice, rather than (e.g.) their shopping list
- Longer reflection, at least, will be worth exploring

Why is accuracy low? (1)

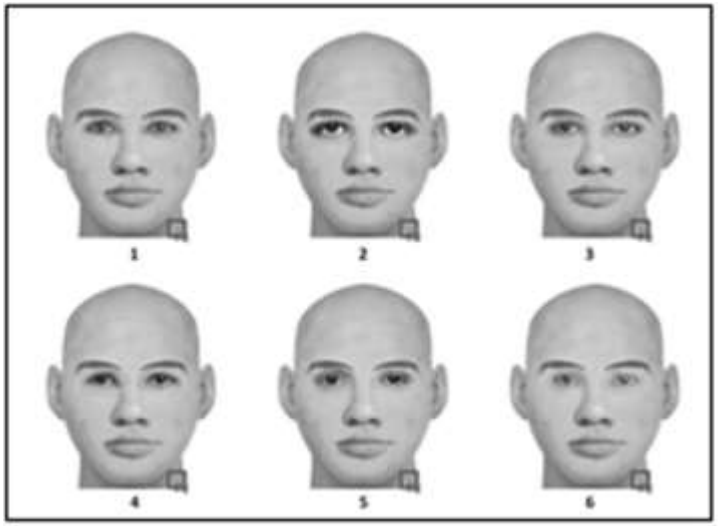
- Our design minimises ‘propitious heterogeneity’ (Wells 1993), in order that we can potentially see improvements when factors are varied. By design we use
  - tightly accent-controlled speaker population
  - rigorous selection of perceptually close foils and targets
- i.e., we make the participants’ task as earwitnesses as hard as we can
- Carlson et al. 2019 on visual parades:
  - ”empirical discriminability decreases as fillers [foils] become too similar to each other and the suspect”

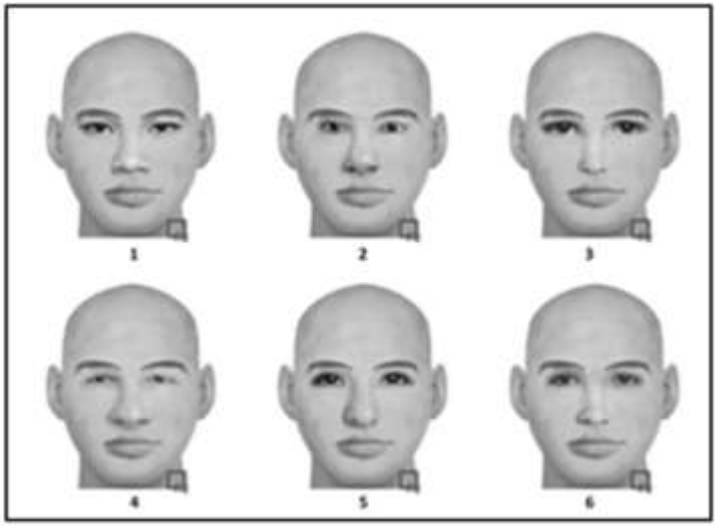
Wells, G. L. (1993) What do we know about eyewitness identification? *American Psychologist* 48(5), 553–571.

Carlson, C.A., Jones, A.R., Whittington, J.E., Lockameyer, M.A.C. & Wooton, A.R. (2019) Lineup fairness: propitious heterogeneity and the diagnostic feature-detection hypothesis. *Cognitive Research: Principles and Implications.* https://doi.org/10.1186/s41235-019-0172-5

Carlson et al.'s computer-generated faces

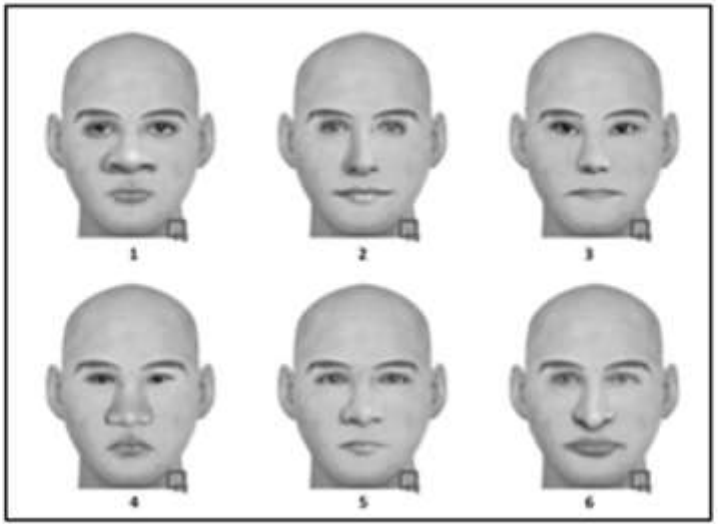
eyes only





eyes + nose

eyes + nose + mouth



Why is accuracy low? (2)

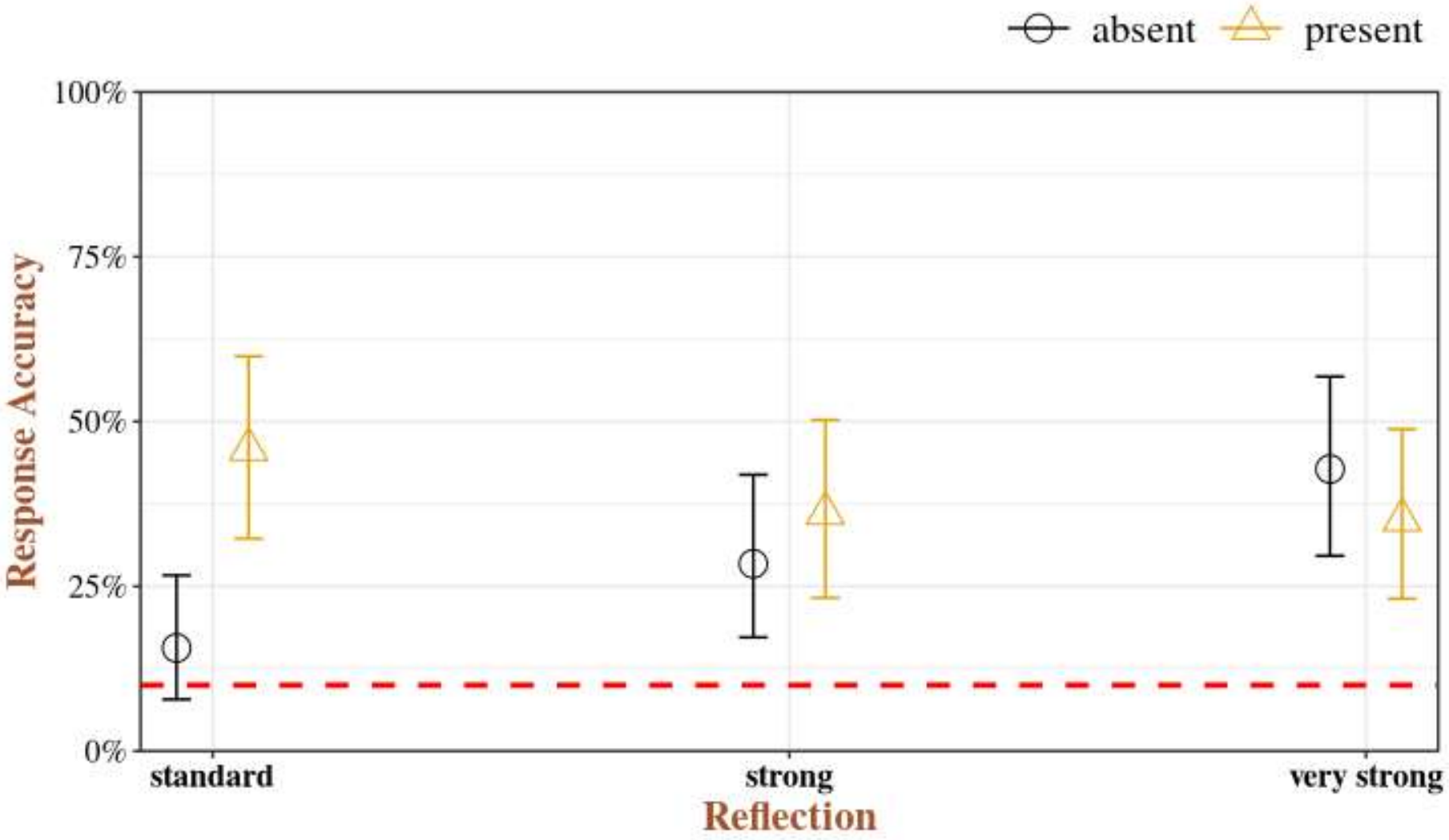
- We also suspect the (unavoidable) online presentation reduces participants' engagement and motivation
  - McDougall, Nolan & Hudson (2015), in an in-person simulated parade, report **76%** correct for target-present — and that after a week's retention interval — compared to **30-40%** here
- In the case of target-absent parades, results just emerging from another IVIP experiment suggest the strength of warning before the parade is crucial:

K. McDougall, F. Nolan and T. Hudson (2015) 'Telephone Transmission and Earwitnesses: Performance on Voice Parades Controlled for Voice Similarity.' *Phonetica* 72: 257-272.

Three strengths of warning

- WEAK: “Remember that the voice you heard at the beginning of the experiment may or may not be present.”
- STRONG: “Remember that the perpetrator may or may not be present. Please consider your response carefully. In a real case, selecting someone from the lineup when the perpetrator is not present could lead to a wrongful conviction.”
- VERY STRONG: “Remember that the perpetrator may or may not be present. Please consider your response carefully. In a real case, selecting someone from the lineup when the perpetrator is not present could lead to a wrongful conviction. Voice recognition can be very difficult. Only make a positive identification if you are very sure.”

Effect of pre-parade warning



Conclusions and further work

- Our two experiments failed to show an effect of a period of ‘reflection’ — either with simulated (word-task) or real (overnight) delay
  - the reflection allowed may have been too short
  - we had no check that participants really engaged
- Future work might test longer reflection, and check ‘engagement’
- What does improve ‘target-absent’ accuracy significantly, emerging results suggest, is the strength of warning

See IVIP website for updates

https://www.phonetics.mmll.cam.ac.uk/ivip/

IVIP

UKRI Economic and Social Research Council