HUSO 2025 March 9 - 13, 2025 Lisbon, Portugal

An Investigation of Inconsistent Expectations of Horse Racing Experts

Yasuhiko Watanabe, Hideaki Nakanishi,
and Yoshihiro Okada
watanabe@rins.ryukoku.ac.jp
Ryukoku University

presenter information

- Doctor of Informatics
- Lecturer, Ryukoku University, Kyoto, Japan

Yasuhiko Watanabe

our research interest

- communications in SNS
- user behavior analysis
- trust and security in SNS

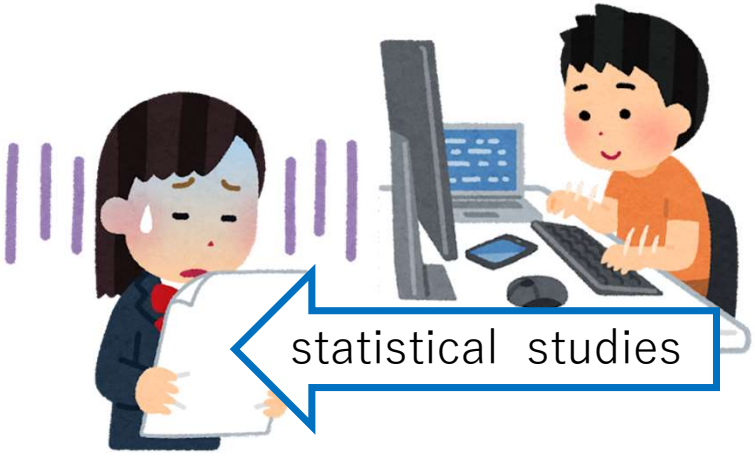
background

statistical studies showed that
even experts can make mistakes

We experts have

- knowledge,
- experience, and
- supporting staff

but, sometimes …



“fly-ball revolution” [Sawchik 2017]

fly-balls

grounders

◎ fly-balls
△ grounders
(from 2017)

statistical research

△ fly-balls
◎ grounders
(until 2017)

baseball experts

Which is better:
fly-balls, or grounders?

What we learned

even experts can make mistakes

A new question

whether experts have
inconsistent expectations
on their professional issues ?

A new question

whether experts have inconsistent expectations ?

Research target

horse racing experts
(horse owners and trainers)

horse racing experts' purposes

win races and get the prize money

prize money in Japan Racing Association (JRA)

Horse owners get prize money
when their horses finish within fifth place.

horse racing experts' tactics (1/2)

race selection

- select races that are likely to have good outcomes
- not select races that are likely to have poor outcomes

horse racing experts' tactics (2/2)

horse selection

- consider carefully
  which horses seem likely to win
  even in unfavorable races.

- enter good performance horses into
  races of a similar distance repeatedly

Our reseach purpose

detect cases of
experts' inconsistent expectations

detect races of a certain distance that horse racing experts thought

e.g., horses of a certain sire line were favorable and unfavorable to win

inconsistent expectation

horse racing experts' belief in selecting races

Many horse racing experts often say

a sire line can indicate
the potential abilities of a horse
(e.g. which distance races the horse is good at)

sire line: paternal lineage or ancestry of a horse

Our approach

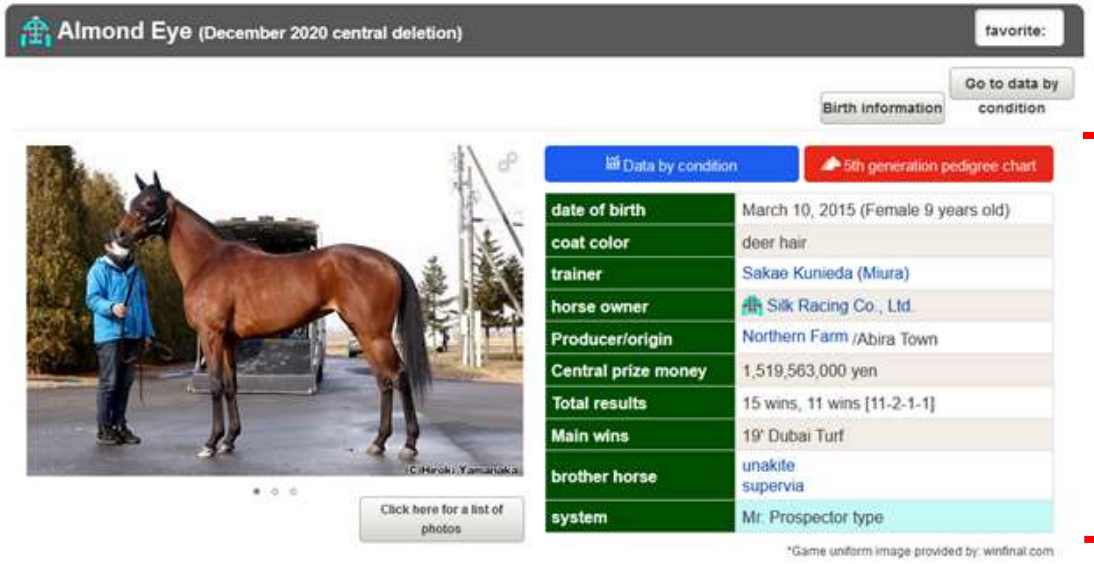
we focus on

- sire line,
- race distance, and
- order of finish

and discuss

- experts' race selections,
- race results, and
- experts' judgements on horse performance

by using statistical analysis



gathering horse info from Keiba Lab. (https://www.keibalab.jp)

Personal Information:
name, date of birth, sex, …, owner, trainer, …, **sire line**, …

ancestors up to three generations ago

Almond Eye race results

year month day	place	course	weather	Baba	race	Popularity	With	jockey	Weight	Number of animals	Frame number	horse number	time	difference in delivery	pace	Inbound	B	horse weight	Passing order	Winning horse (2nd place)
2020/11/29	5th 9	Tokyo	Shiba 2400	cloudy	good	Japan C (GI)	1	1	Lemaire	55.0	15	2	2	2:23.0	0.2	35.3 - 37.8(H)	34.7	490(0)	Four Five Four Four	(contrail)
2020/11/01	4th 8	Tokyo	Shiba 2000	cloudy	good	Emperor's Award (Autumn) (GI)	1	1	Lemaire	56.0	12	7	9	1:57.8	0.1	36.5 - 33.6(S)	33.1	490(+2)	3 3 Four	(Fiermann)
2020/06/07	3rd 2	Tokyo	Shiba 1600	Sunny	a little	Yasuda Memorial (GI)	1	2	Lemaire	56.0	14	Four	Five	1:32.0	0.4	34.2 - 34.3(M)	33.9	488(+2)	11 11	Gran Alegria
2020/05/17	2nd 8	Tokyo	Shiba 1600	Sunny	good	Victoria (GI)	1	1	Lemaire	55.0	18	6	12	1:30.6	0.7	34.2 - 33.9(M)	32.9	486(0)	Four Four	(Sound Chiara)

Race results:
Venue, date, race name, …, distance, …, order of finish, …

our obtained data of racehorses

36922 horses registered with JRA from 2010 to 2017

Year	# of horses
2010	4470
2011	4524
2012	4505
2013	4595
2014	4672
2015	4663
2016	4738
2017	4755
Total	36922

# of horses classified into three famous sire lines

sire line	# of horses
Native Dancer Line	8799
Nearctic Line	6383
Royal Charger Line	18104
others	3636
Total	36922

(Note) We grouped many kinds of branched sire lines into four kinds above

# of times the 36922 horses had competed in

race distance	# of races
1000 -- 1399m	98122
1400 -- 1799m	133635
1800 -- 2199m	131178
2200 -- 2799m	22009
2800m --	10882
Total	395826

# of times the 36922 horses of four sire lines had competed in races of various distances

	race distance					
sire line	1000 -- 1399m	1400 -- 1799m	1800 -- 2199m	2200 -- 2799m	2800m --	Total
Native Dancer	27008	31619	28568	4173	2511	93879
Nearctic	18710	22444	20072	2838	1647	65711
Royal Charger	42525	67514	71758	13181	5848	200826
others	9879	12058	10780	1817	876	35410
Total	98122	133635	131178	22009	10882	395826

# of times the 36922 horses of four sire lines had finished in first place in races of various distances

	race distance					
sire line	1000 -- 1399m	1400 -- 1799m	1800 -- 2199m	2200 -- 2799m	2800m --	Total
Native Dancer	1947	2261	2121	341	188	6858
Nearctic	1347	1511	1399	206	143	4606
Royal Charger	2580	4767	5496	1078	495	14416
others	677	855	671	105	52	2360
Total	6551	9394	9687	1730	878	28240

# of times the 36922 horses of four sire lines had finished within fifth place in races of various distances

	race distance					
sire line	1000 -- 1399m	1400 -- 1799m	1800 -- 2199m	2200 -- 2799m	2800m --	Total
Native Dancer	9345	10912	10552	1748	1120	33677
Nearctic	6462	7700	7112	1070	728	23072
Royal Charger	13893	23937	26949	5369	2713	72861
others	3203	4054	3564	655	317	11793
Total	32903	46603	48177	8842	4878	141403

(Note) horses within fifth place get prize money in the JRA races

statistical analysis of experts' inconsistent expectations

we focus on

sire line, race distance, and order of finish

and discuss

- experts' race selections
- race results
- experts' judgements of horse performance

by using two-sided binomial test and detect cases of experts' inconsistent expectations

statistical analysis of experts' inconsistent expectations

two-sided binomial test

- **Hypothesis *ES*** → experts' race selections
- **Hypothesis *RR*** → race results
- **Hypothesis *EJ*** → experts' judgements of horse performance

two-sided binomial test on experts' selections by using Hypothesis *ES*

Hypothesis *ES*

$$\frac{\#\text{ of times horses were entered into races of distance } d_j}{\#\text{ of times horses were entered into races}} \times \#\text{ of times horses of sire line } S_i \text{ were entered into races of distance } d_j$$

probability: an expert enters his/her horse into a race of distance d_j

two-sided binomial test on race results by using Hypothesis *RR*

Hypothesis *RR*

$$\frac{\#\text{ of times horses finished within } rank\text{-th place in races of distance } d_j}{\#\text{ of times horses were entered into races of distance } d_j} \times \#\text{ of times horses of sire line } S_i \text{ were entered into races of distance } d_j$$

probability: a horse finished within *rank*-th place into a race of distance d_j

two-sided binomial test on experts' judgements of horse performance by using Hypothesis *EJ*

Hypothesis *EJ*

$$\frac{\#\text{ of times horses of sire line } S_i \text{ were entered in races of distance } d_j}{\#\text{ of times horses were entered into races of distance } d_j} \times \#\text{ of times horse of sire line } S_i \text{ were entered into races of distance } d_j$$

probability: an expert enters a horse of sire line S_i into a race of distance d_j

How to detect cases of experts' inconsistent expectations (1/3)

1. We calculate the p-value of
   - experts' race selections
   - race results

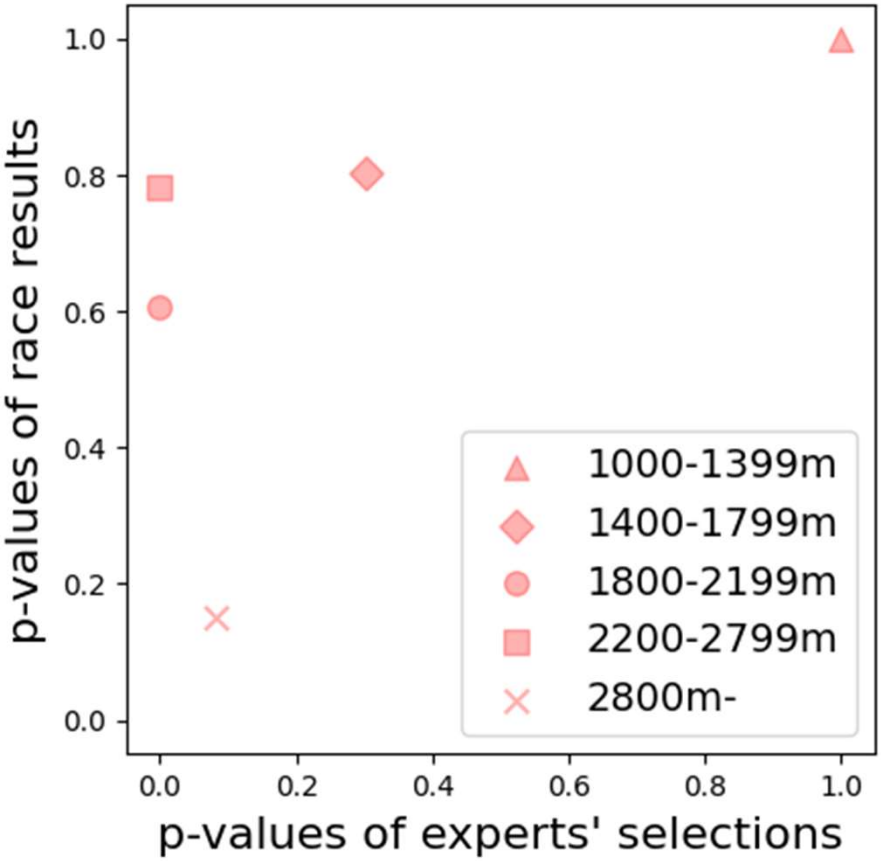
   by applying Hypothesis *ES* and *RR* of

two-sided binomial test

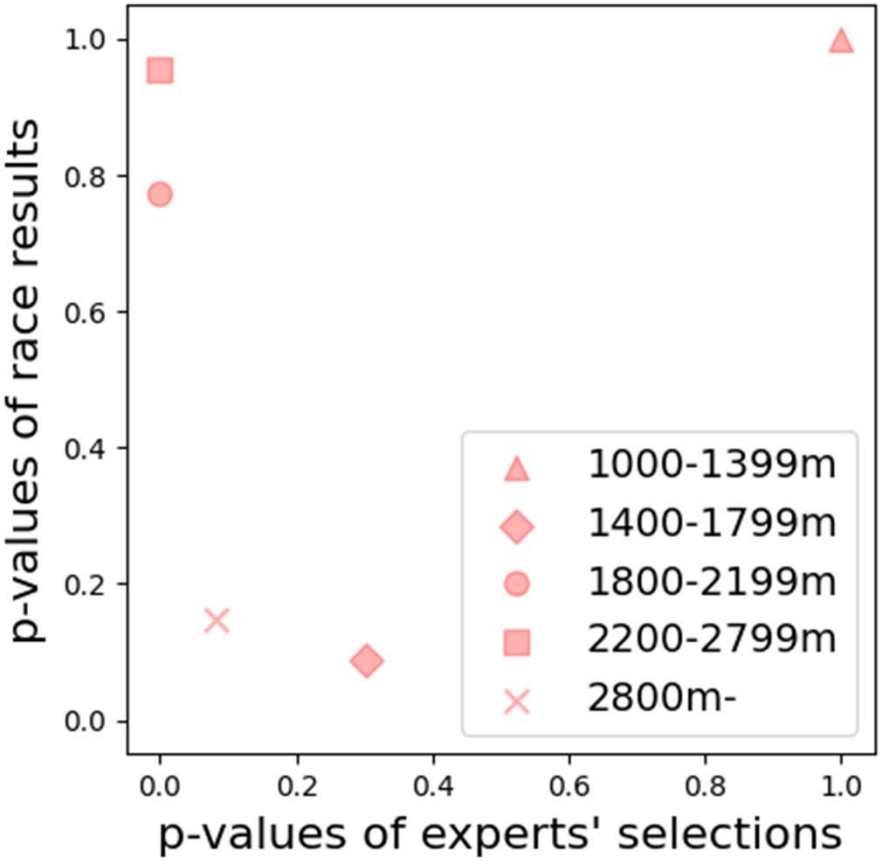
sire lines		race distances	
Native Dancer Line	X	1000 – 1399m	2200 – 2799m
		1400 – 1799m	2800m –
		1800 – 2199m	

experts' selections vs race results

Native Dancer Line

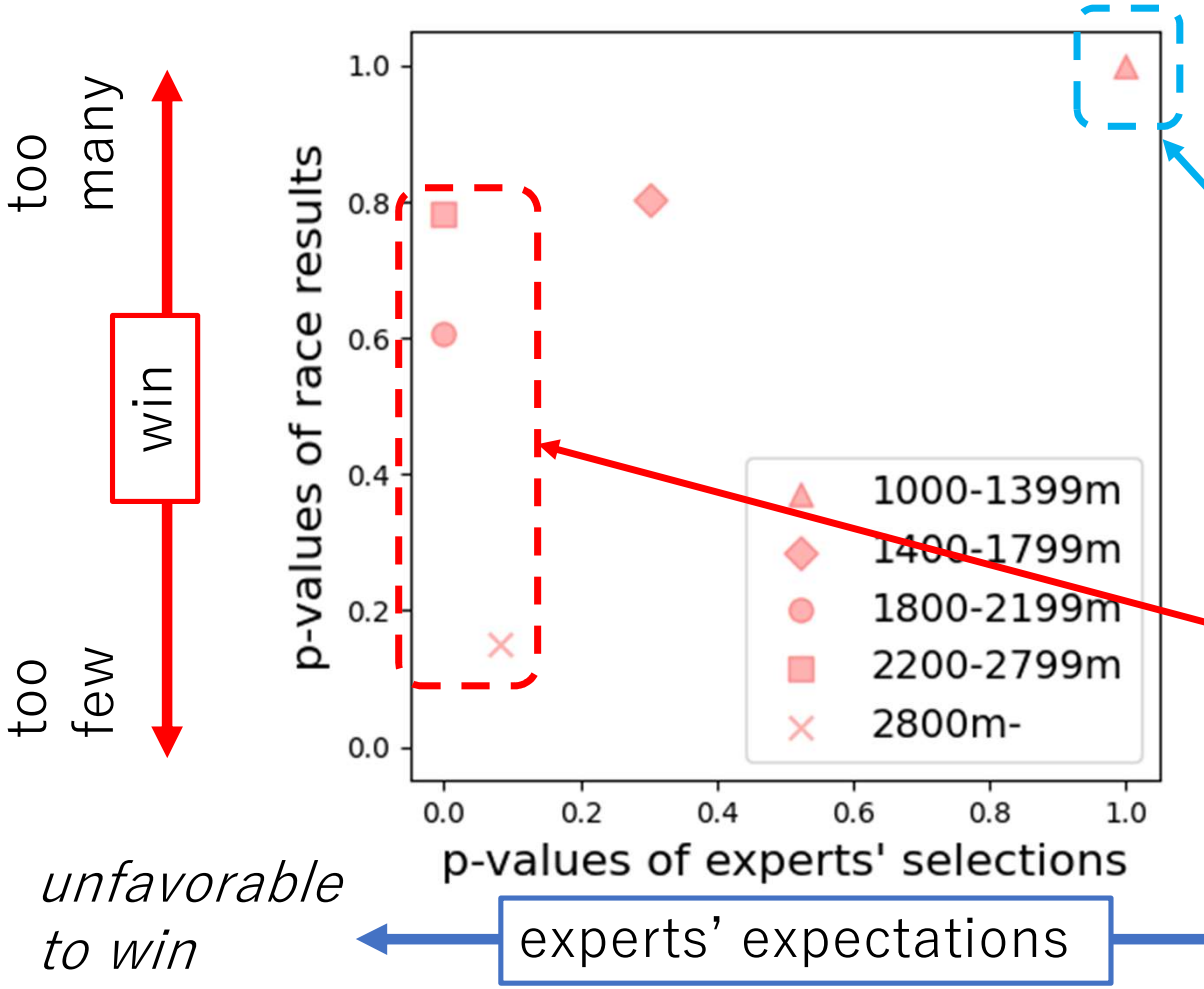


first place



within fifth place

how to read graphs of experts' selections vs race results (first place)



Native Dancer Line (first place)

favorable to win

✓ 1000 – 1399 m

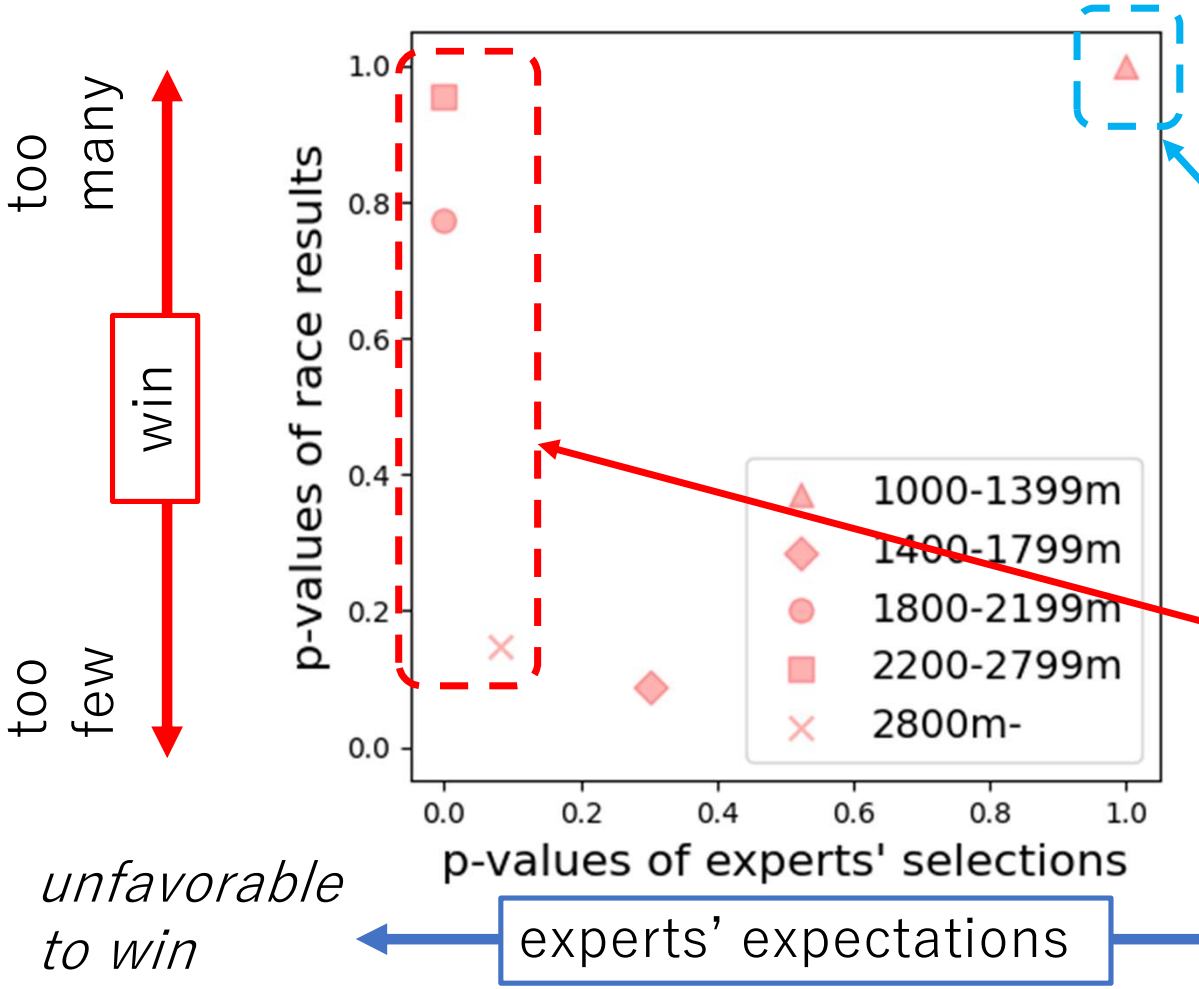
unfavorable to win

✓ 1800 – 2199 m

✓ 2200 – 2799 m

✓ 2800m -

how to read graphs of experts' selections vs race results (within fifth)



Native Dancer Line (within fifth)

favorable to win

- ✓ 1000 – 1399 m

unfavorable to win

- ✓ 1800 – 2199 m
- ✓ 2200 – 2799 m
- ✓ 2800m -

How to detect cases of experts' inconsistent expectations (2/3)

two-sided binomial test

2. We calculate the p-value of

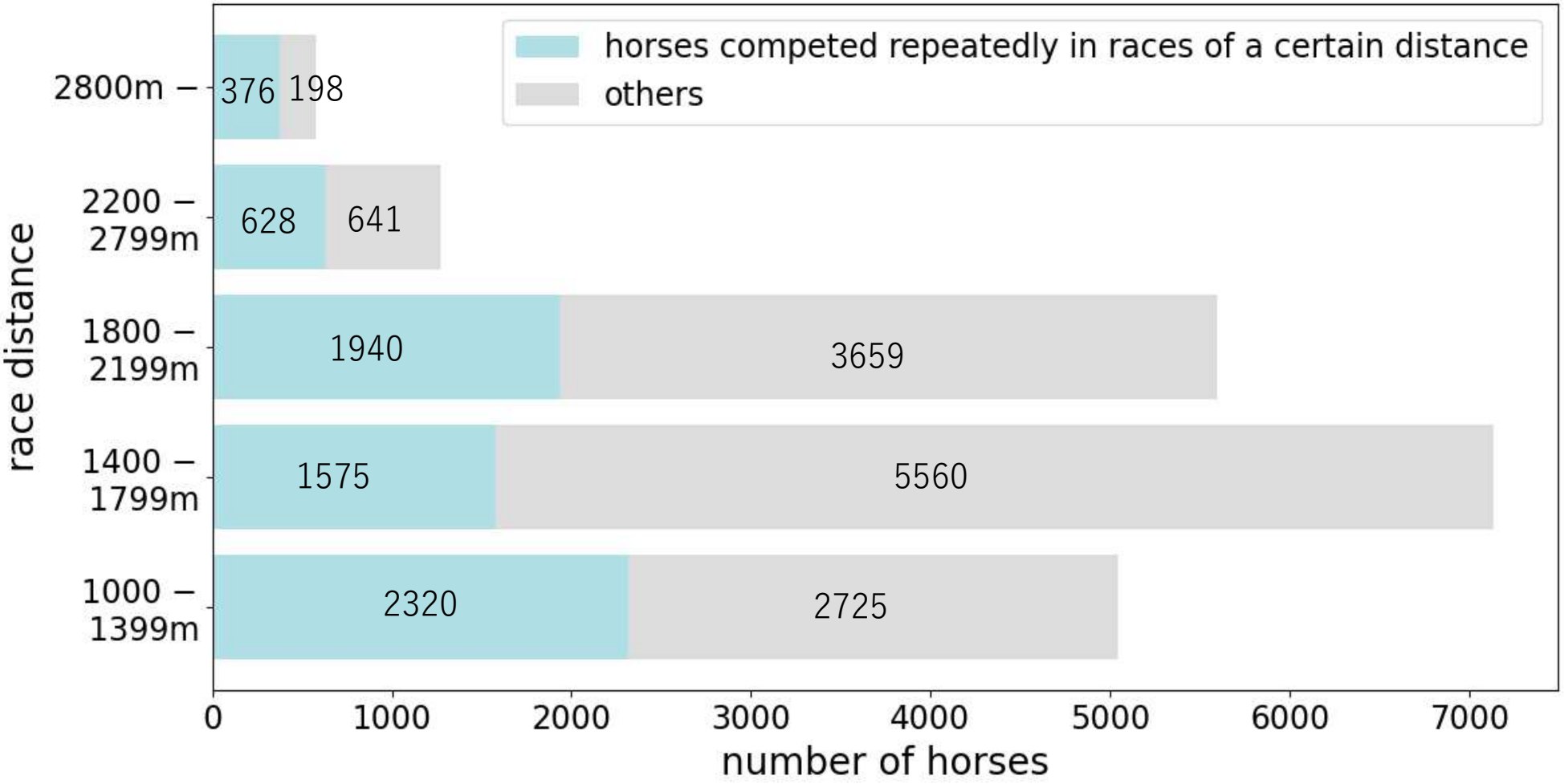
- experts' judgements of horse performance

by applying Hypothesis EJ and detect horses competed repeatedly in races of a certain distance

sire lines		race distances	
Native Dancer Line	X	1000 – 1399m	2200 – 2799m
		1400 – 1799m	2800m –
		1800 – 2199m	

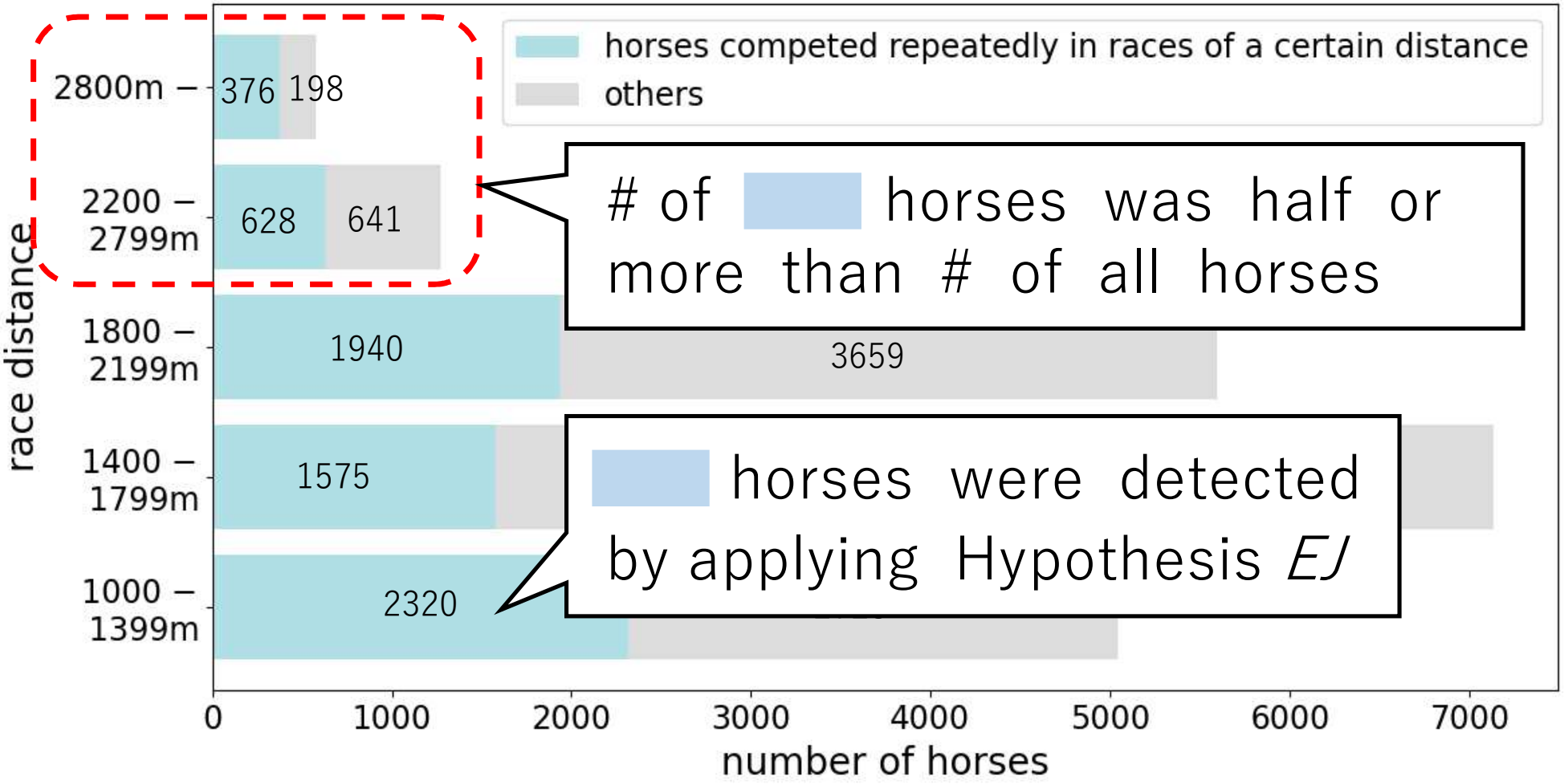
experts' judgements of horse performance (1/3)

Native Dancer Line



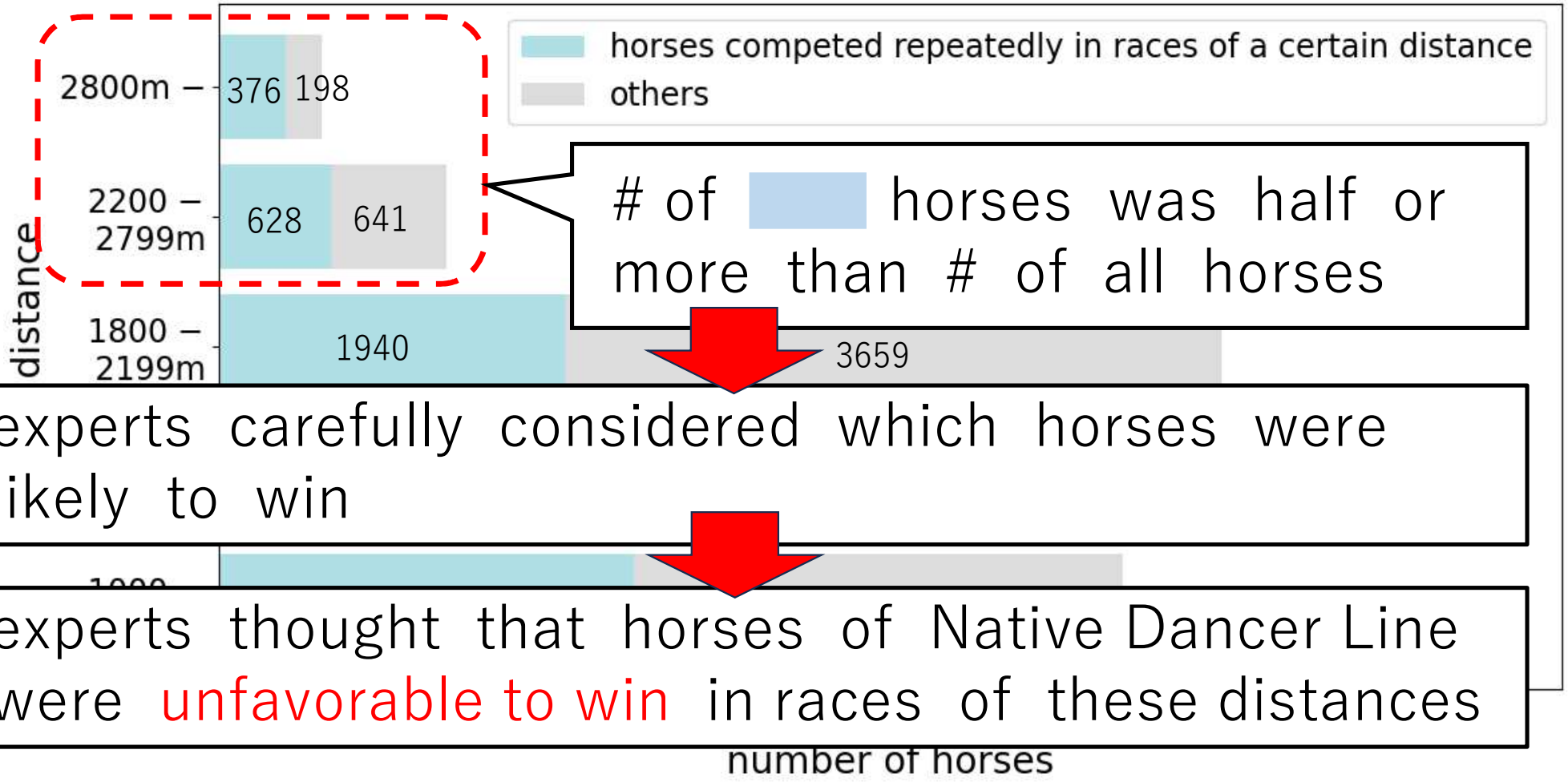
experts' judgements of horse performance (2A/3)

Native Dancer Line



experts' judgements of horse performance (2B/3)

Native Dancer Line



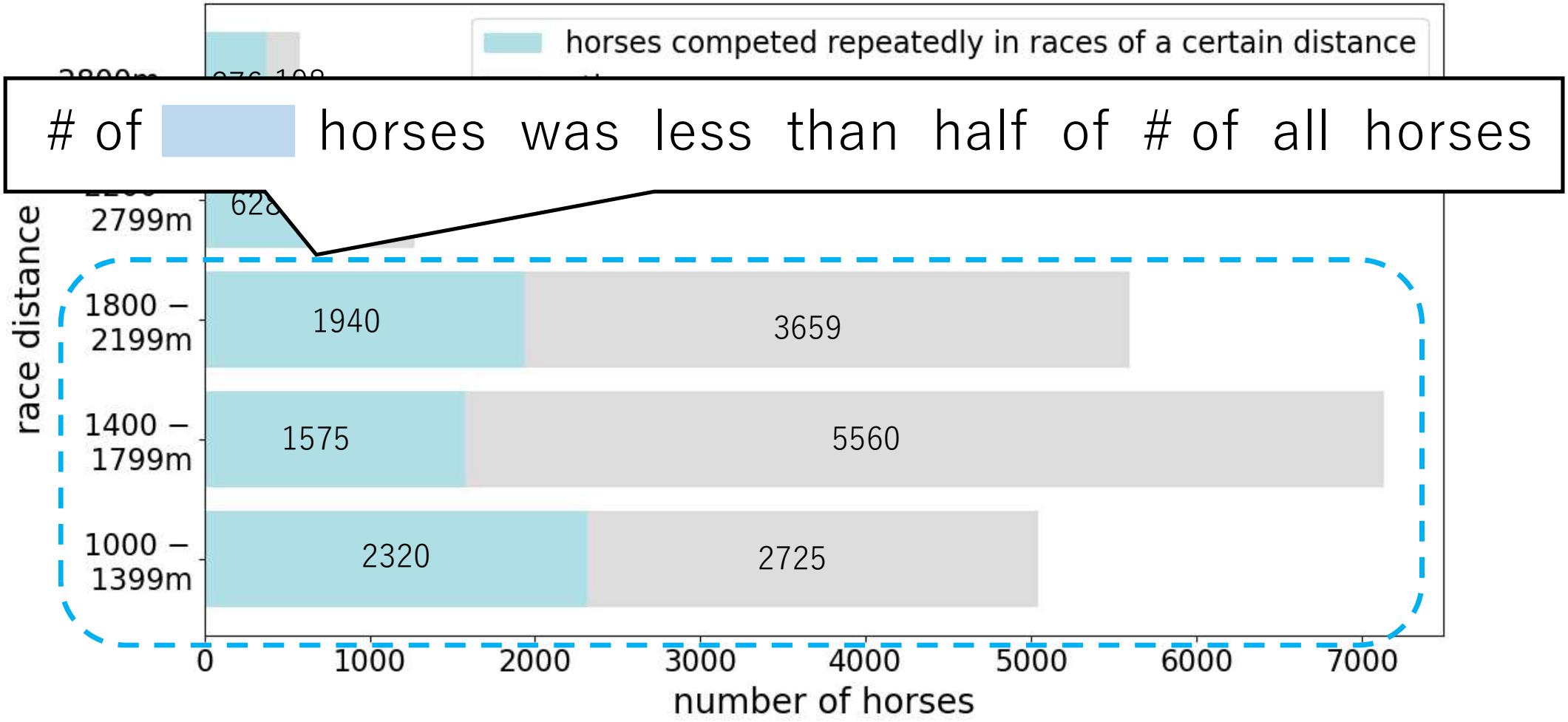
# of horses was half or more than # of all horses

experts carefully considered which horses were likely to win

experts thought that horses of Native Dancer Line were unfavorable to win in races of these distances

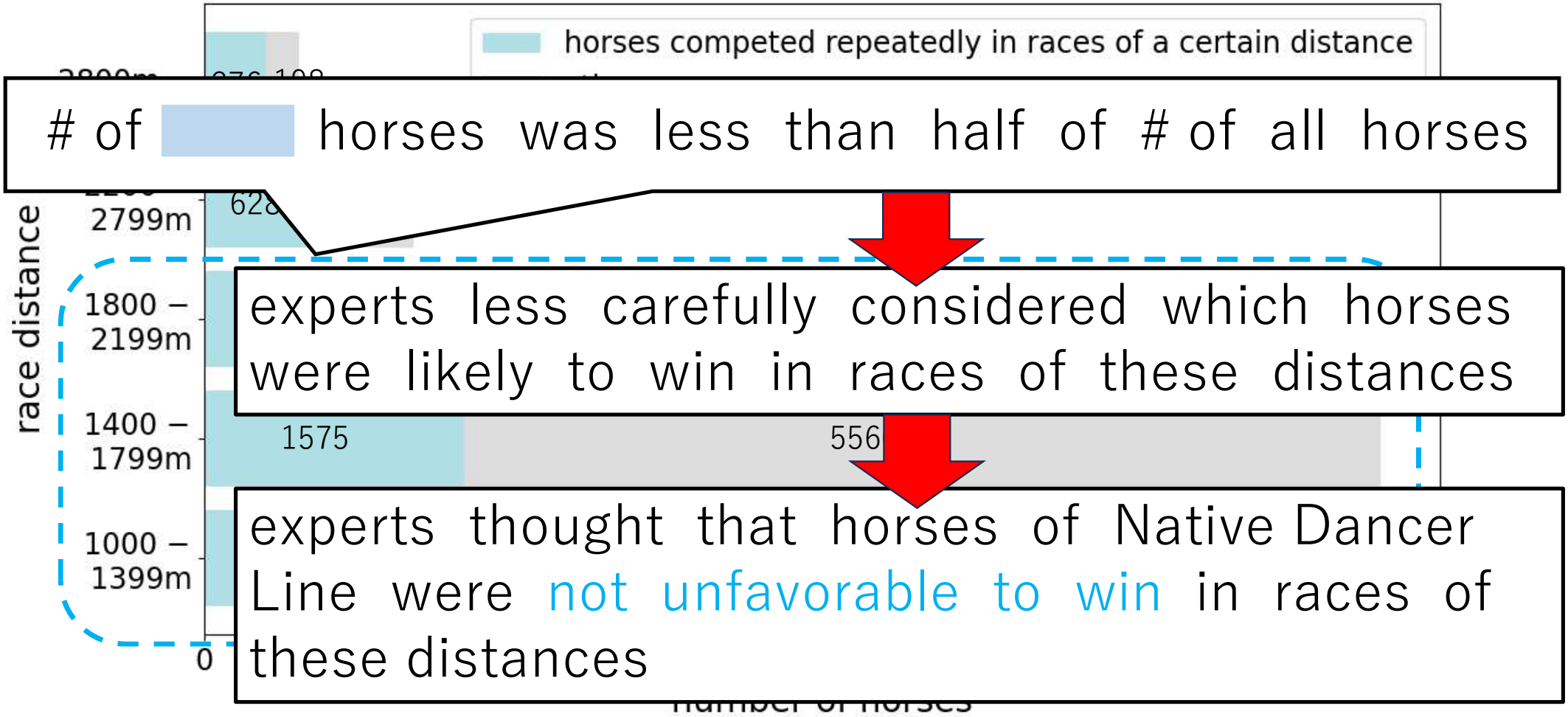
experts' judgements of horse performance (3A/3)

Native Dancer Line



experts' judgements of horse performance (3B/3)

Native Dancer Line



# of horses was less than half of # of all horses

experts less carefully considered which horses were likely to win in races of these distances

experts thought that horses of Native Dancer Line were not unfavorable to win in races of these distances

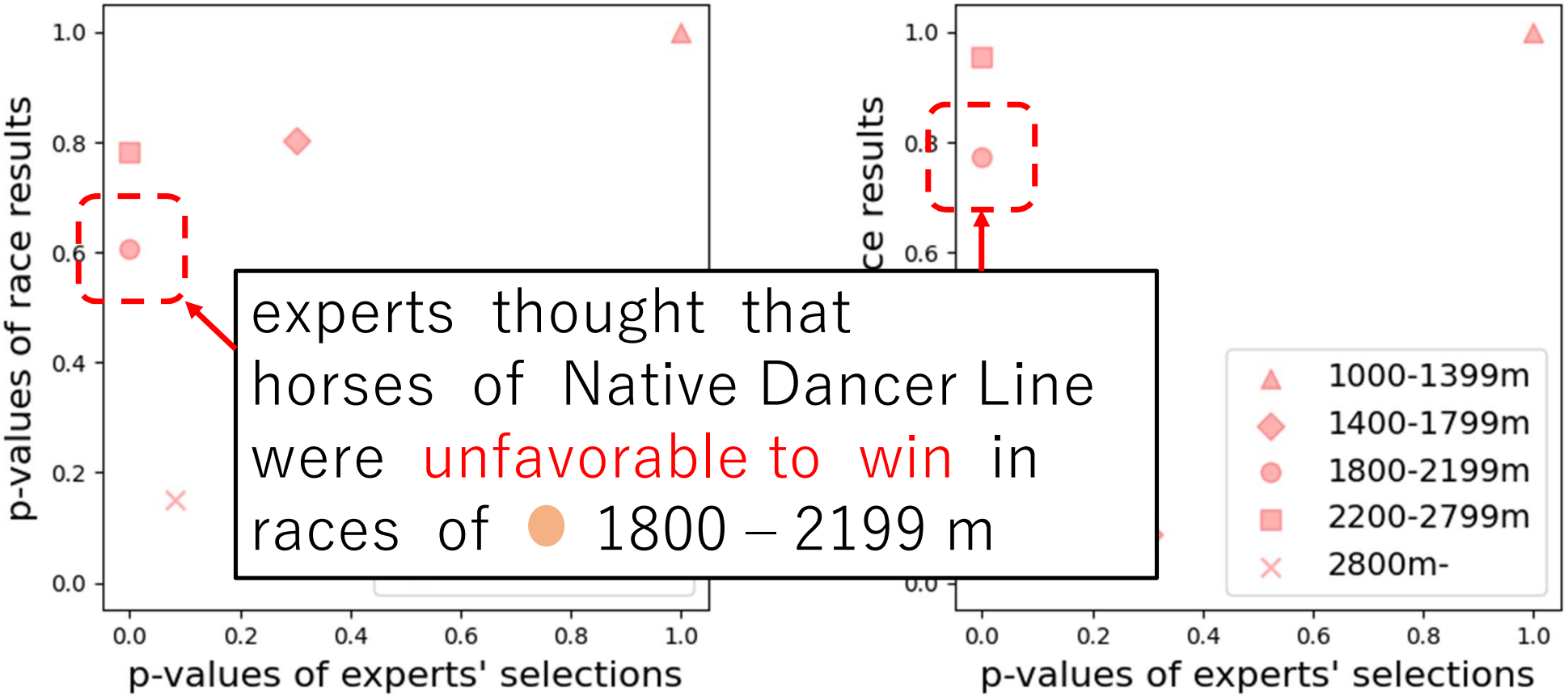
How to detect cases of experts' inconsistent expectations (3/3)

3. We detect experts' inconsistent expectations by analyzing the results of
   - experts' race selections
   - experts' judgements of horse performance

sire lines		race distances	
Native Dancer Line	X	1000 – 1399m	2200 – 2799m
		1400 – 1799m	2800m –
		1800 – 2199m	

experts' expectations (1/2)

Native Dancer Line



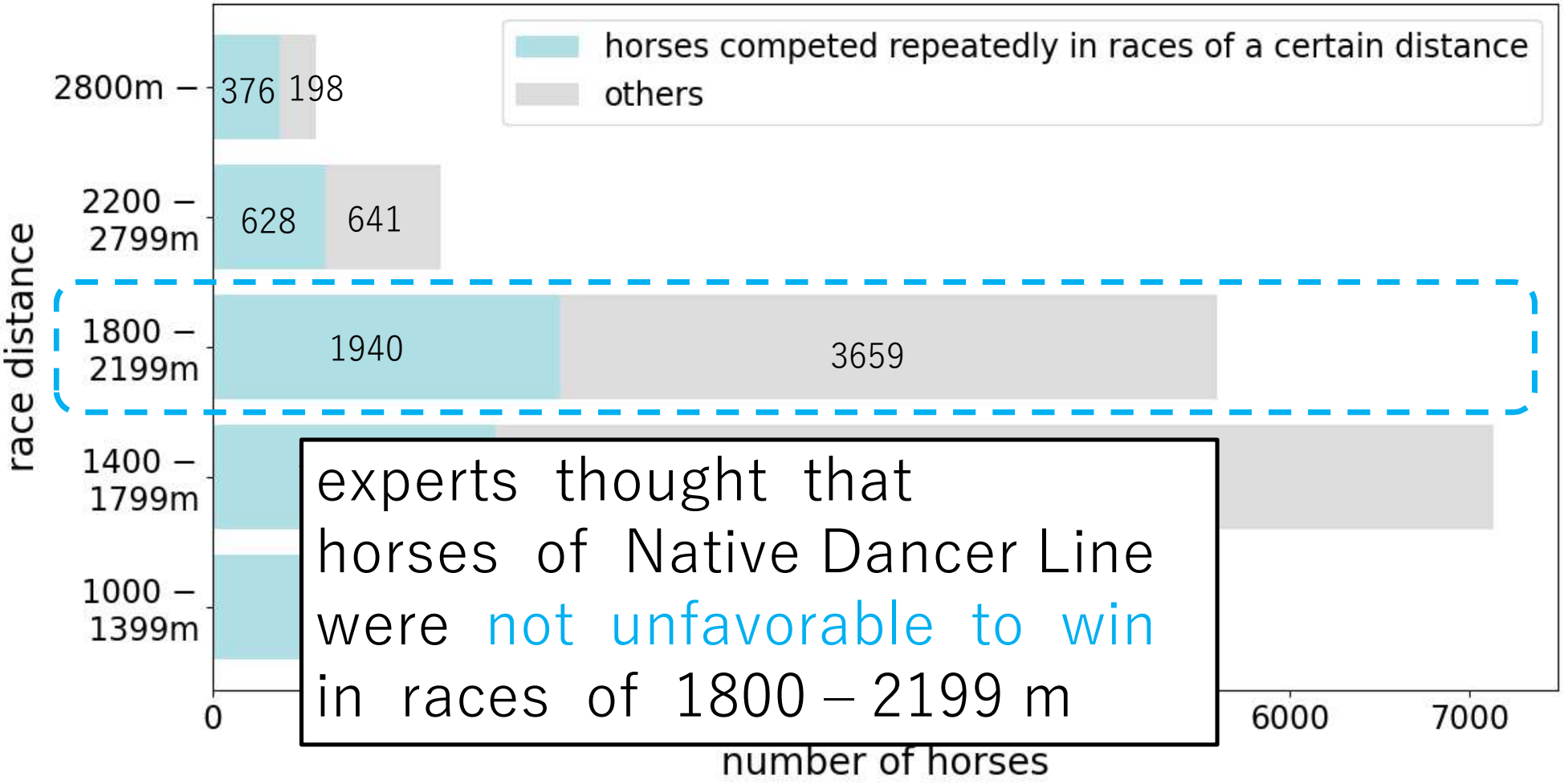
experts thought that horses of Native Dancer Line were unfavorable to win in races of ● 1800 – 2199 m

first place

within fifth place

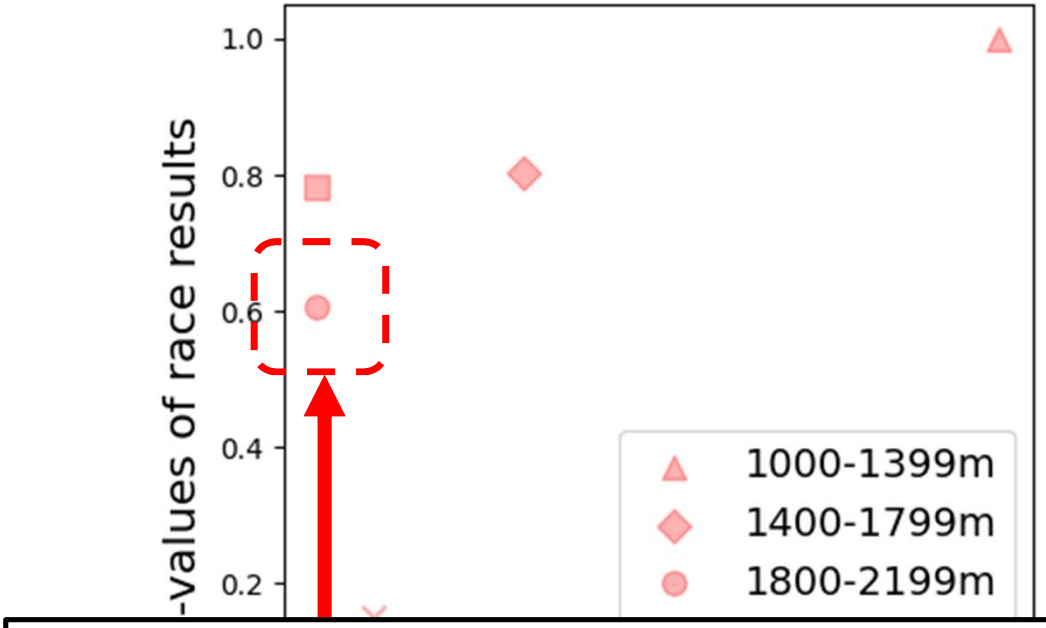
experts' expectations (2/2)

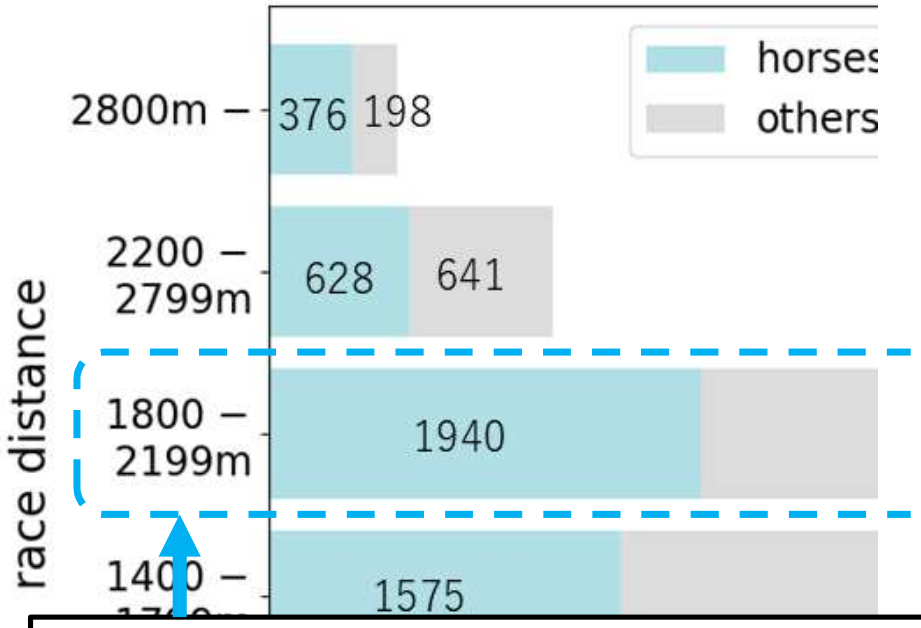
Native Dancer Line



experts thought that horses of Native Dancer Line were not unfavorable to win in races of 1800 – 2199 m

[conclusion]　experts' inconsistent expectations





unfavorable to win
in races of ● 1800 – 2199 m

⇔

not unfavorable to win
in races of 1800 – 2199 m

experts' expectations were inconsistent

Why horse racing experts had inconsistent expectations? (1/2)

We think it is because

many experts unconsciously knew that horses of Native Dancer Line were unfavorable to win in races of 1800 – 2199 m

They may have unconsciously avoided selecting races of 1800 – 2199 m

Why horse racing experts had inconsistent expectations? (2/2)

On the other hand,

their conscious minds may not have known that the horses were unfavorable to win in the races of 1800 – 2199 m

→ many experts may not have carefully considered which horses were likely to win in the races

Future works

To generalize this finding, we intend to

- analyze race performance data in other countries
- compare the results with those obtained in this study