Tactical positioning behaviours in short-track speed skating: A static and dynamic sequence analysis

HEXT, Andrew <http://orcid.org/0000-0003-2070-6601>, HETTINGA, Florentina Johanna and MCINERNEY, Ciarán

Citation:

Tactical positioning behaviours in short-track speed skating: A static and dynamic sequence analysis

Andrew Hext[a]*, Florentina Johanna Hettinga[b], Ciarán McInerney[c]

[a] *Sports Engineering Research Group, Sport & Physical Activity Research Centre, Sheffield Hallam University, Sheffield, United Kingdom.*

[b] *Department of Sport, Exercise & Rehabilitation, Faculty of Health and Life Sciences, Northumbria University, Newcastle Upon Tyne, United Kingdom.*

[c] *Academic Unit of Primary Medical Care, University of Sheffield, Sheffield, United Kingdom.*

*Corresponding author: Andrew Hext, Sports Engineering Research Group, Sheffield Hallam University, AWRC, Olympic Legacy Park, 2 Old Hall Road, Sheffield, S9 3TU, United Kingdom, a.hext@shu.ac.uk

ORCID:

Andrew Hext, http://orcid.org/0000-0003-2070-6601

Florentina Johanna Hettinga, https://orcid.org/0000-0002-7027-8126

Ciarán McInerney, https://orcid.org/0000-0001-7620-7110

Word count:

4,232

1   **Abstract**

2   Tactical positioning is essential for success in short-track speed skating as the race format

3   (direct, head-to-head competition over multiple laps) prioritises finishing position over

4   finishing time. Despite this, current research into tactical positioning treats the race's laps as

5   discrete, independent events. Accordingly, the aggregate metrics used to summarise each

6   lap's tactical positioning behaviour do not allow us to explore the sequential nature of the

7   data, e.g., Lap 2 occurs after Lap 1 and before Lap 3. Here, we capture the sequential

8   relationships between laps to investigate tactical positioning behaviours in short-track speed

9   skating. Using intermediate and final rankings from 500 m, 1,000 m, and 1,500 m elite short-

10   track races, we analyse whole-race and sub-race race sequences of group and winner tactical

11   positioning behaviours. This approach, combined with a large dataset of races collected over

12   eight seasons of competition ($n = 4{,}135$), provides the most rigorous and comprehensive

13   description of tactical positioning behaviours in short-track speed skating to date. Our results

14   quantify the time-evolving complexity of tactical positioning, offer new thoughts on race

15   strategy, and can help practitioners design more representative learning tasks to enhance skill

16   transfer.

17   **Keywords**

20   **Introduction**

21   Short-track speed skating is a form of competitive ice speed skating that consists of

22   individual events (500 m, 1,000 m, 1,500 m) and relay events (2,000 m, 3,000 m, 5,000 m)

23   performed anticlockwise on a 111.12 m oval (International Skating Union, 2021). In all

24   events, athletes and teams must qualify through several rounds of competition to reach the

25   medal contest (e.g., heats, quarterfinals, and semi-finals), with each qualifying race

26   characterised by multiple skaters or teams (typically four to six) racing head-to-head at

27   speeds exceeding 11 m/s (Bullock et al., 2008; ISU, 2021). Critically, advancement through

28   the competition and medal colour depends on an athlete's or team's finishing rank and not

29   their finishing time. For example, an athlete could win/ qualify from *semi-final 1* with a

30   slower finishing time than an athlete who failed to qualify from *semi-final 2* (Hext et al.,

31   2022). For this reason, an athlete's decisions regarding how and when to invest their limited

32   energy resources – both before (strategic) and during (tactical) the race – are considered

33   essential for success (Hext et al., 2017, 2022; Muehlbauer & Schindler, 2011). This goal-

34   directed regulation of exercise intensity is known as 'pacing' (Abbiss & Laursen, 2008).

35   In recent years, researchers have highlighted the importance of athlete-environment

36   interactions for understanding pacing behaviour, i.e., the outcome of the strategic and tactical

37   decision-making process (Hettinga et al., 2017; Konings & Hettinga, 2018c; Renfree et al.,

38   2014; Renfree & Casado, 2018; Smits et al., 2014). For example, factors that characterise the

39   environment in short-track speed skating, such as the competition stage, the competition

40   importance, and preceding race efforts, all alter pacing behaviour (Konings & Hettinga,

41   2018b, 2018d). Arguably the most crucial athlete-environment interaction for understanding

42   pacing behaviour in short-track speed skating is those between athlete and opponent

43   (Hettinga et al., 2017; Hext et al., 2022; Konings & Hettinga, 2018c). Konings & Hettinga

44   (2018a) showed that the high variability observed in between-race finishing times is

45   primarily due to athletes altering their pacing behaviour to that of other opponents,

46   particularly during the race's early stages. Furthermore, drafting possibilities, competing for

47   the optimum line, avoiding collisions, minimising fall risk, and overtaking all represent other

48   athlete-opponent interactions that may cause an athlete to alter their pace (Konings et al.,

3

49   2016; Noorbergen et al., 2016). For these reasons, previous research has investigated tactical

50   positioning – i.e., athletes ranking within the race (1st, 2nd, 3rd, 4th, etc.) – to help contextualise

51   pacing behaviour (Konings et al., 2016; Noorbergen et al., 2016), explore how it can be

52   learned (Menting et al., 2019), and as a subject in its own right, based on position being the

53   most important performance outcome (Haug et al., 2015; Hext et al., 2022; Maw et al., 2006;

54   Muehlbauer & Schindler, 2011; Sun et al., 2021).

55   Our understanding of tactical positioning in short-track speed skating is based on two

56   levels of analysis: group and individual behaviours. The group level of analysis focuses on

57   the collective behaviour of all athletes in the race, with researchers quantifying the tactical

58   importance of athlete ranking at the race start and end of each lap by using Kendall's Tau-b,

59   τ_b, to measure the similarity between athletes' intermediate and final rankings (Haug et al.,

60   2015; Konings et al., 2016; Maw et al., 2006; Menting et al., 2019; Muehlbauer & Schindler,

61   2011; Noorbergen et al., 2016; Sun et al., 2021). In contrast, the individual level of analysis

62   focuses on the tactical positioning behaviours of individuals and usually those of the winner,

63   assuming that they are the most successful at the decision-making process and, therefore,

64   their actions are of interest (Konings et al., 2016). In this scenario, researchers quantify how

65   winners position themselves at the race start and end of each lap by calculating the proportion

66   of races where they skated at each ranking or their mean rank (Bullock et al., 2008; Konings

67   et al., 2016; Maw et al., 2006; Muehlbauer & Schindler, 2011; Noorbergen et al., 2016; Sun

68   et al., 2021). Typically, researchers only use the group level of analysis to infer race strategy.

69   Although, on occasions, both group and winner behaviours are used. For example,

70   Noorbergen et al. (2016) concluded that tactical positioning is crucial from the race start in

71   the 500 m as there was a positive association between athletes' start position and final

72   rankings ($\tau_b = 0.38$), and 51% of winners started in first position.

73   Without underestimating these approaches' insights, the methods used to investigate

74 tactical positioning behaviours treat the race start and each lap as discrete, independent

75 events. Accordingly, the aggregate metrics used to summarise each lap do not allow us to

76 explore the sequential nature of the data, e.g., Lap 2 occurs after Lap 1 and before Lap 3. In

77 Table 1, we demonstrate how this limits our capacity to understand tactical positioning by

78 analysing the tactical positioning behaviour of winners for the 'Start' and 'Lap 1' in 10 races.

79 Using traditional discrete lap analyses, Table 1 shows that 50% of winners started in 1st

80 position (mean rank: 1.8 ± 0.9), and 60% were ranked 1st at the end of Lap 1 (mean rank: 1.6

81 ± 0.8). The standard interpretation of such results would infer that: (1) tactical positioning is

82 crucial from the race start, and (2) acknowledge that some variation in tactics exists based on

83 the relatively high standard deviation in winner rank. Now let us consider the data's

84 sequential structure and produce ten race sequences of the form: (Start position, Lap 1 rank).

85 First, Table 1 shows that only one race had a sequence where the winner started and remained

86 in 1st, i.e., (1, 1). Such a finding would question the discrete lap analysis's interpretation that

87 tactical positioning is crucial from the race start. Second, analysis of the ten sequences allows

88 us to surmise the different winning tactical positioning behaviours, unlike the discrete lap

89 analysis, which could only propose their presence. Table 1 reveals five unique winning

90 sequences: (1, 3), (1, 2), (1, 1), (3, 1), and (2, 1), and identifies the modal sequence as (3,1),

91 appearing in 3 out of 10 sequences. Note that this measure of central tendency also respects

92 the semantic definition of rank, unlike the mean, as an athlete cannot occupy a ranking of 1.8

93 or 1.6. Both examples highlight how techniques that capture the relationship between discrete

94 events offer a deeper understanding of performance in sport (Borrie et al., 2002).

95   State sequence analysis is a statistical framework for identifying patterns in

96 temporally ordered lists of objects, states, or events (Lowe et al., 2020). Originally used for

97 sequence matching in bioinformatics and later developed for investigating longitudinal

98   patterns in the social sciences (Abbott & Tsay, 2000; Ritschard & Studer, 2018), researchers

99   have since applied sequence analysis to a variety of domains (Conway et al., 2019; Lowe et

100   al., 2020; Roux et al., 2019; Vanasse et al., 2020). We recently demonstrated the utility of

101   state sequence analysis for investigating tactical positioning in short-track speed skating

102   (Hext et al., 2022). We showed that the higher level of measurement granularity afforded by

103   the state sequence analysis better captured the complexity of tactical positioning. In the 1,000

104   m event, we detected 1,269 unique sequences of group behaviour compared to the single

105   sequence produced by the traditional discrete lap approach, which combined the aggregate

106   metrics used to summarise each lap. We concluded that capturing this complexity offers a

107   more detailed understanding of tactical positioning that could enhance the strategic and

108   tactical decisions essential for success in short-track speed skating (Hext et al., 2022).

109   For these reasons, this study investigates tactical positioning in short-track speed

110   skating using state sequence analysis. Specifically, we use static and dynamic sequence

111   analysis to investigate group and winner tactical positioning behaviours in the 500 m, 1,000

112   m, and 1,500 m events. The static sequence analysis provides an overall view of tactical

113   positioning behaviours as it treats the complete race sequence (start to end) as a single unit of

114   analysis. The dynamic sequence analysis provides a more nuanced view of how tactical

115   positioning evolves throughout the race as it evaluates nested race sequences with a constant

116   endpoint (the final lap) but varying start points (e.g., Lap 1, Lap 2, Lap 3). To comment on

117   the utility of accounting for the dataset's sequential structure, we compare our results with the

118   most up-to-date, complete race, discrete lap analyses of group and winner behaviours

119   (Konings et al., 2016; Noorbergen et al., 2016). As such, we do not stratify our analysis by

120   different environmental factors in this study.

121   **Method**

122   This study was approved by the Research Ethics Committee at Sheffield Hallam University,

123   UK.

124   *Dataset*

125   Our dataset consisted of 10,804 races (500 m = 4,308; 1,000 m = 4,056; 1,500 m = 2,440),

126   from 62 competitions (44 World Cups, 8 European Championships, 8 World Championships,

127   and 2 Winter Olympic Games), over an 8-season period (2010/11 to 2017/18). All data was

128   retrieved from the International Skating Union's results website

129   (https://shorttrack.sportresult.com/). For each race, the dataset contained all competitors'

130   starting position, intermediate rankings, and final rankings. The dataset coded starting

131   positions from 1 (innermost track position) to n (outermost track position) and intermediate/

132   final rankings from 1 (leading athlete) to n (last athlete). Note the deliberate distinction in

133   terminology between start position and intermediate/ final rankings: at the race start, all

134   athletes have the same ranking but different spatial positions as they are distributed across a

135   start line perpendicular to the direction of the track (Hext et al., 2022).

136   Before analysing the dataset, we excluded races with falls, disqualifications, missing

137   values, tied intermediate rankings, and races where the number of athletes competing was not

138   equal to the event's modal value, i.e., 4 athletes for the 500 m and 1,000 m, and 6 athletes for

139   the 1,500 m. These strict inclusion criteria were in line with previous short-track speed

140   skating research (Hext et al., 2022; Konings et al., 2016; Noorbergen et al., 2016). Our final

141   dataset included 4,135 of the 10,804 races (500 m = 2,020, 46.9%; 1,000 m = 1,549, 38.2%;

142   1,500 m = 566, 23.2%). We provide a complete breakdown of our cleaning procedure in

143   Supplemental Table 1.

Data analysis

Sequence definition

146   We consider a sequence, x_i, as an ordered, discrete-time series of elements, a, of length, l,

147   that can be represented as $(a_1, a_2, ..., a_n)$, where $l_{(x_i)} = a_n$. The discrete-time series

148   represents the points in the race where we measure athlete rank: the race start and end of each

149   lap. Each element in the series has a state that belongs to a finite set of states that characterise

150   tactical positioning behaviour, i.e., the state-space. We use two different state spaces to

151   characterise tactical positioning behaviour. The first state-space characterised the group's

152   tactical positioning behaviours. As described in Hext et al. (2022), we quantified the group's

153   tactical positioning behaviour by measuring the similarity between start/intermediate and

154   final rankings in each race using Kendall's Tau-b, τ_b. A Kendall's $\tau_b = 1$ represents a perfect

155   agreement between start position/ intermediate and final rankings, and a Kendall's $\tau_b = -1$

156   represents a perfect disagreement. The second state-space characterised the winner's tactical

157   positioning behaviours and included all possible athlete rankings. We summarise each

158   event's group and winner behaviour state-space in Table 2 of the Supplemental material.

159   *Static and dynamic sequence formation*

160   We formed static and dynamic sequences for each race in the dataset. Figure 1 illustrates this

161   process. Our static analysis generated a complete race sequence of the group's and winner's

162   tactical positioning behaviours. For example, the winner's tactical positioning behaviour, (4,

163   4, 1, 1, 1, 1), indicates that the winner started and remained in 4th until Lap 2. From this point

164   onwards, the winner was ranked first at the end of each lap. Note that in Figure 1, we

165   represent all sequences using the state-permanence-sequence format (Aassve et al., 2007). In

166   this format, each successive distinct state in a sequence is given together with its duration, t,

167   so that $x_i = (a_1, t_1) - (a_2, t_2) - ... - (a_n, t_n)$. Accordingly, we represent the winning

168   sequence, (4, 4, 1, 1, 1, 1), as (4, 2) – (1, 4). Given the length of our sequences, particularly in

169   the 1,500 m event ($l_{(x)} = 15$), we will use this format for the remainder of the manuscript.

170   Our dynamic analysis generated nested sequences of the group's and winner's tactical

171   positioning behaviours. The nested sequences had a constant endpoint (the race's final lap)

172   but varying lengths. The number of nested sequences for each race was equal to $1 - l_{(x_i)}$,

173   where $l_{(x_i)}$ is the number of elements in the static race sequence. Starting from Lap 1, we

174   incremented the start point of each nested sequence by one lap until the start point equalled

175   the race's final lap. As demonstrated in Figure 1, we create five nested sequences (Lap 1–5,

176   Lap 2–5, …, Lap 5) from the six elements (Start, Lap 1, …, Lap 5) in the 500 m. Overall, the

177   static and dynamic sequences formed 6, 10, and 15 sequence periods in the 500 m (4.5 laps),

178   1,000 m (9 laps), and 1,500 m (13.5 laps), respectively. Here, note that our analysis uses Lap

179   1 to represent the tactical positioning behaviours at the end of the initial half-lap in the 500 m

180   and 1,500 m events.

181   *Sequence metrics*

182   For each sequence period, we calculated the number of unique sequences, n_x, and the

183   sequence duplication rate, $sdr = \left(1 - \left(\frac{n_x}{n}\right)\right) \cdot 100$, where n is the number of races in the

184   dataset. A sequence duplication rate of 0% indicates that no sequences are the same, and a

185   sequence duplication rate of 100% indicates that all sequences are the same. In addition, we

186   calculated each sequence's absolute and relative support. A sequence's absolute support,

187      $sup(x_i)$, denotes the number of times the sequence occurs in the sequence period, with its

188   relative support, $relSup(x_i) = \left(\frac{sup(x_i)}{n}\right) \cdot 100$ (Fournier-Viger et al., 2017; Hext et al.,

189   2022). For example, in a dataset of 1,000 races, an absolute support of 500 would indicate

190   that 500 races had the same sequence of tactical positioning behaviours, representing a

191   relative support of 50%. We performed all sequence analyses in MATLAB 2021b and used

192   the R statistical programming language (version 4.0.0) to interrogate the data.

193   **Results**

194   Figure 2 quantifies the time-evolving complexity of tactical positioning in short-track speed

195   skating. First, note that regardless of the level of analysis (group or winner behaviour), the

196   complexity of tactical positioning increases with race distance, i.e., the static sequence

197   duplication rates decrease (Start– Lap n). Second, tactical positioning becomes less complex

198   as the race progresses, i.e., the dynamic sequence duplication rates increase as the length of

199   the nested sequences decreases. Here, the only exception is the start of the 1,500 m. Until the

200   sequence period Lap 5–15, the group sequence duplication rates remain at 0%. In other

201   words, all observed sequences are unique. Finally, the group tactical positioning behaviour is

202   more complex than the winner and, as a result, converges to 100% slower. That is, the group

203   sequence duplication rate is always lower than the winner sequence duplication rate, for each

204   sequence period, in all events.

205   Figure 3 illustrates the time-evolving distribution of all detected sequences' relative

206   support, with Table 2 reporting the most frequent sequence for each period. We provide the

207   complete list of unique group and winner sequences and their associated support in the

208   supplemental material. Note that for all events and levels of analysis, the median support is

209   always close to 0%. In other words, exact group and winner behaviours typically do not

210   reoccur on multiple occasions. Nevertheless, we did identify behaviours that frequently

211   recurred, i.e., sequences with a support greater or equal to the inner fence: $Q3 +$

212     $(1.5 * Interquartile\ range)$ (Tukey, 1977). Generally, the most frequent sequence

213   represents behaviours where the winner is ranked first, and the group order mimics the final

214   rankings for the entirety of the race. The only exception is in the 1,500 m event, where we

215 only saw this group behaviour from the sequence periods Lap 11–14 onwards. When

216 considering the complete static race sequence, the support for these group and winner

217 behaviours is greatest in the 500 m and decreases with race distance. When considering the

218 dynamic race sequences, the support increases as the race progresses and is always greater for

219 the winner, regardless of the event.

220 **Discussion**

221 We have used static and dynamic sequence analysis to investigate tactical positioning

222 behaviours in short-track speed skating. To our knowledge, we are the first to use this

223 statistical framework in short-track speed skating performance analysis. Whereas existing

224 research treats laps as discrete events, we captured the sequential relationship between these

225 events for the entire race sequence (static) and nested race sequences with a constant

226 endpoint but varying lengths (dynamic). By combining this approach with a large dataset of

227 races collected over eight seasons ($n = 4,135$), our results provide the most rigorous and

228 comprehensive description of tactical positioning behaviours in short-track speed skating to

229 date.

230 A key feature of our sequence analysis is that we do not use aggregate metrics to

231 summarise each lap. Instead, we form sequences that capture the athlete-opponent

232 interactions throughout – and specific to – each race. In doing so, we provide stronger

233 evidence that reaffirms several beliefs about tactical positioning in short-track speed skating.

234 For example, current discrete lap analyses suggest that tactical positioning is crucial from the

235 race start in the 500 m because the start/ intermediate rankings (end of each lap) positively

236 correlate with the final rankings (Haug et al., 2015; Maw et al., 2006; Muehlbauer &

237 Schindler, 2011; Noorbergen et al., 2016). However, as this evidence evaluates each lap

238 independently from all other laps, it can only infer – rather than show – that winners adopt a

239   skate-from-the-front strategy during races. In contrast, our analysis considers how each

240   winner positioned themselves from one lap to the next for the entirety of the race.

241   Accordingly, our finding that nearly one in every two races was won by the athlete starting

242   and remaining in first position ($relSup = 47.5\%$) is stronger empirical evidence that

243   controlling the race from the front is a key determinant of success in the 500 m. Similarly,

244   current discrete lap analyses propose that athletes reduce their effort to skate-from-the-front

245   as the race distance increases (Muehlbauer & Schindler, 2011; Noorbergen et al., 2016; Sun

246   et al., 2021) and that the number of ways to win increases with the race distance (Sun et al.,

247   2021). The former is inferred from positive correlations between start/ intermediate rankings

248   and final rankings decreasing, and the latter is inferred from the standard deviation of the

249   winner's rank increasing. Our analysis provides stronger empirical evidence as we can show

250   that: (1) the support for the skate-from-the-front strategy decreases ($relSup_{500\,m} = 47.5\%$,

251    $relSup_{1,000\,m} = 8.1\%$, $relSup_{1,500\,m} = 0.5\%$); and (2) the winner's sequence duplication rate

252   decreases ($sdr_{500\,m} = 94.1\%$, $sdr_{1,000\,m} = 57.7\%$, and $sdr_{1,500\,m} = 0.7\%$), as the race

253   distance increases.

254   While confirming established ideas on tactical positioning behaviours in short-track

255   speed skating, our analysis also offers new perspectives. For example, we found that the most

256   recurring winning behaviour in the 1,000 m and 1,500 m events was to skate-from-the-front.

257   With seven laps to go, this sequence represented at least one in every four races

258   ($relSup_{1,000\,m} = 30.5\%$, Lap 3–9; $relSup_{1,500\,m} = 24.7\%$, Lap 8–14). This strategy differs

259   from current discrete lap analyses, which advocate that athletes should conserve energy by

260   occupying a ranking other than first until Lap 6 (1,000 m) and Lap 10 (1,500 m), because

261   there is not a strong relationship between intermediate and final rankings before this

262   (Konings et al., 2016; Noorbergen et al., 2016). Note that we ensured this finding was due to

263   our sequence analysis rather than due to analysing different datasets (8 seasons from 2010/11

264   to 2017/18 compared to 1 season from 2012/13) by replicating the traditional discrete lap

265   analysis on our dataset. As illustrated in the Supplemental Material (Supplemental Figures 1–

266   4), there were no discernible differences between the two datasets – and the inferences drawn

267   – when treating laps as discrete events. Our observation of this most-recurring sequence

268   suggests that some winners choose to forgo the physiological benefit of drafting and lead for

269   the majority. This decision, in part, may be due to: (1) the athlete deciding that the difficulty

270   in overtaking is more costly than having other competitors benefit from drafting them

271   (Hoffman et al., 1998); and (2) attempting to mitigate the risk of falls associated with

272   collisions (Hext et al., 2022). Such a strategy, therefore, may be more suited to an athlete

273   with a higher perception of risk (Micklewright et al., 2015). While we do not endorse one

274   strategy over another, it is clear that more than one winning strategy exists. Importantly,

275   sequence analysis allows us to capture and broaden our understanding of the different race

276   strategies adopted in short-track speed skating.

277   Our analysis also has several more direct implications for race strategy and

278   performance research in short-track speed skating. First, we provide an empirical list of group

279   and winner behaviours, and their associated support, at any stage of the race. We hope that

280   practitioners can use this list to: (1) support the formulation of race strategies and tactics; and

281   (2) inform the design of practice constraints and learning tasks that represent the performance

282   environment and thus enhance the transfer of skill from training to competition (Pinder et al.,

283   2011). Second, we believe both researchers and practitioners should use individual levels of

284   analysis to inform race strategy rather than the norm of using group behaviour, as the latter is

285   more complex and therefore appears to underestimate the importance of tactical positioning

286   for an individual, particularly during the race's earlier stages. For example, consider the

287   1,000 m event where the most recurring group and winner sequence supported a skate-from-

288   the-front strategy. For the sequence period Lap 2–9 (89% of the total race distance), the

289   winner support for this strategy was at least 1 in 5 races ($relSup = 22.2\%$) compared to at

290   least 1 in 20 races for the group behaviour ($relSup = 5.2\%$). Third, future work should

291   explore detecting commonalities between unique sequences to create a taxonomy of tactical

292   positioning behaviours. Such a taxonomy would enhance our understanding of the different

293   race strategies and tactics utilised in short-track speed skating by capturing the latent

294   structures of the many behaviours observed. State sequence analysis is well suited for this

295   work because it offers a suite of metrics and methods for estimating sequence dissimilarity

296   and building sequence typologies (Lowe et al., 2020; Ritschard & Studer, 2018). Finally,

297   future work should replicate our sequence analysis for different race scenarios because

298   previous research has shown that environmental factors, such as the season, competition

299   round, whether athletes are male or female, and the competition importance, can evoke

300   modifications in tactical positioning or pacing behaviour (Konings & Hettinga, 2018b; Maw

301   et al., 2006; Muehlbauer & Schindler, 2011; Sun et al., 2021). Such analyses would help

302   coaches and athletes tailor their race preparation for the relevant performance environment.

303   While our work represents an advance in tactical positioning analysis in short-track

304   speed skating, we should note two limitations. First, our dataset only represented race

305   scenarios with each event's modal number of athletes, no falls, and no disqualifications.

306   While these strict inclusion criteria resulted in our analysis excluding over half of each

307   event's data, we could compare our sequential analysis directly with results from traditional

308   analyses. Second, our sequences only characterised group and winner behaviour at the race

309   start and end of each lap. Accordingly, we could not characterise within-lap position changes

310   as this exceeds the dataset's resolution. For example, an athlete may start and finish a lap

311   ranked 2nd but be ranked 1st halfway through the lap. Note, however, that this issue is present

312   in all studies that use competition results to investigate tactical positioning behaviours in

313   short-track speed skating.

314 **Conclusion**

315 Tactical positioning behaviour is a complex process that emerges from multiple athletes

316 interacting continuously over many laps. By accounting for the sequential structure of these

317 interactions, we can begin to quantify and decode this complexity. Here, we have taken the

318 first step by providing the most rigorous and comprehensive description of tactical

319 positioning behaviours in short-track speed skating to date. This empirical aid quantifies the

320 time-evolving complexity of tactical positioning, offers new thoughts on race strategy based

321 on the prevalence of winners choosing to skate-from-the-front, and can help practitioners

322 design more representative learning tasks to enhance skill transfer.

323 **Funding details**

326 **Disclosure statement**

327 We report no conflicts of interest.

328 **References**

329 Aassve, A., Billari, F. C., & Piccarreta, R. (2007). Strings of adulthood: A sequence analysis

330   of Young British Women's work-family trajectories. *European Journal of Population*,

331   *23*(3–4), 369–388. https://doi.org/10.1007/s10680-007-9134-6

332 Abbiss, C. R., & Laursen, P. B. (2008). Describing and understanding pacing strategies

333   during athletic competition. *Sports Medicine*, *38*(3), 239–252.

334   https://doi.org/10.2165/00007256-200838030-00004

335 Abbott, A., & Tsay, A. (2000). Sequence Analysis and Optimal Matching Methods in

336   Sociology. *Sociological Methods & Research*, *29*(1), 3–33.

337   https://doi.org/10.1177/0049124100029001001

338   Borrie, A., Jonsson, G. K., & Magnusson, M. S. (2002). Temporal pattern analysis and its

339   applicability in sport: An explanation and exemplar data. *Journal of Sports Sciences*,

340   *20*(10), 845–852. https://doi.org/10.1080/026404102320675675

341   Bullock, N., Martin, T. D., & Zhang, A. (2008). Performance analysis of world class short-

342   track speed skating: What does it take to win? *International Journal of Performance*

343   *Analysis in Sport*, *8*(1), 9–18. https://doi.org/10.1080/24748668.2008.11868418

344   Conway, A., Collins, P., Chang, K., Mafeld, S., Sutherland, J., & Fingleton, J. (2019).

345   Sequence analysis of capnography waveform abnormalities during nurse-administered

346   procedural sedation and analgesia in the cardiac catheterization laboratory. *Scientific*

347   *Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-46751-2

348   Fournier-Viger, P., Lin, J. C.-W., Kiran, R. U., Koh, Y. S., & Thomas, R. (2017). A survey of

349   sequential pattern mining. *Data Science and Pattern Recognition*, *1*(1), 54–77.

350   https://doi.org/10.1007/978-3-030-04921-8_4

351   Haug, W. B., Drinkwater, E. J., Mitchell, L. J., & Chapman, D. W. (2015). The relationship

352   between start performance and race outcome in elite 500 m short-track speed skating.

353   *International Journal of Sports Physiology and Performance*, *10*(7), 902–906.

354   https://doi.org/10.1123/ijspp.2014-0504

355   Hettinga, F. J., Konings, M. J., & Pepping, G.-J. (2017). The science of racing against

356   opponents: Affordance competition and the regulation of exercise intensity in head-to-

357   head competition. *Frontiers in Physiology*, *8*(February), 1–7.

358   https://doi.org/10.3389/fphys.2017.00118

359   Hext, A., Heller, B., Kelley, J. W., & Goodwill, S. R. (2017). Relay exchanges in elite short-

360   track speed skating. *European Journal of Sport Science*, *17*(5), 503–510.

361   Hext, A., Hettinga, F. J., & McInernery, C. (2022). Tactical positioning in short-track speed

362   skating: The utility of race-specific athlete-opponent interactions. *European Journal of*

363   *Sport Science*, 1–10. https://doi.org/10.1080/17461391.2022.2069513

364   Hoffman, E., Listemann, E., McManaman, C., & Rundell, K. W. (1998). Short track speed

365   skating: Analysis of drafting during world championship competition. *Medicine and*

366   *Science in Sports and Exercise*, *30*((Suppplement 5)), 310.

367   ISU. (2021). *Special regulations and rules short-track speed skating*.

368   Konings, M. J., & Hettinga, F. J. (2018a). Objectifying tactics: Athlete and race variability in

369   elite short-track speed skating. *International Journal of Sports Physiology and*

370   *Performance*, *13*(2), 170–175. https://doi.org/10.1123/ijspp.2016-0779

371   Konings, M. J., & Hettinga, F. J. (2018b). The impact of different competitive environments

372   on pacing and performance. *International Journal of Sports Physiology and*

373   *Performance*, *13*(6), 701–708. https://doi.org/10.1123/ijspp.2017-0407

374   Konings, M. J., & Hettinga, F. J. (2018c). Pacing decision making in sport and the effects of

375   interpersonal competition: A critical review. *Sports Medicine*, *48*(8), 1829–1843.

376   https://doi.org/10.1007/s40279-018-0937-x

377   Konings, M. J., & Hettinga, F. J. (2018d). The effect of preceding race efforts on pacing and

378   short-track speed skating performance. *International Journal of Sports Physiology and*

379   *Performance*, *13*(8), 970–976. https://doi.org/10.1123/ijspp.2017-0637

380   Konings, M. J., Noorbergen, O. S., Parry, D., & Hettinga, F. J. (2016). Pacing behavior and

381   tactical positioning in 1500 m short-track speed skating. *International Journal of Sports*

382   *Physiology and Performance*, *11*(1), 122–129. https://doi.org/10.1123/ijspp.2015-0137

383   Lowe, M. R., Holbrook, C. M., & Hondorp, D. W. (2020). Detecting commonality in

384   multidimensional fish movement histories using sequence analysis. *Animal*

385   *Biotelemetry*, *8*(1), 1–14. https://doi.org/10.1186/s40317-020-00195-y

386   Maw, S., Proctor, L., Vredenburg, J., & Ehlers, P. (2006). Influence of starting position on

387   finishing position in World Cup 500 m short-track speed skating. *Journal of Sports*

388   *Sciences*, *24*(12), 1239–1246. https://doi.org/10.1080/02640410500497733

389   Menting, S. G. P., Konings, M. J., Elferink-Gemser, M. T., & Hettinga, F. J. (2019). Pacing
390   behavior of elite youth athletes: Analyzing 1500 m short-track speed skating.
391   *International Journal of Sports Physiology and Performance*, *14*(2), 222–231.
392   https://doi.org/10.1123/ijspp.2018-0285

393   Micklewright, D., Parry, D., Robinson, T., Deacon, G., Renfree, A., Gibson, A. S. C., &
394   Matthews, W. J. (2015). Risk perception influences athletic pacing strategy. *Medicine*
395   *and Science in Sports and Exercise*, *47*(5), 1026–1037.
396   https://doi.org/10.1249/MSS.0000000000000500

397   Muehlbauer, T., & Schindler, C. (2011). Relationship between starting and finishing position
398   in short-track speed skating races. *European Journal of Sport Science*, *11*(4), 225–230.
399   https://doi.org/10.1080/17461391.2010.499968

400   Noorbergen, O. S., Konings, M. J., Micklewright, D., Elferink-Gemser, M. T., & Hettinga, F.
401   J. (2016). Pacing behavior and tactical positioning in 500- and 1000-m short-track speed
402   skating. *International Journal of Sports Physiology and Performance*, *11*(6), 742–748.
403   https://doi.org/10.1123/ijspp.2015-0384

404   Pinder, R. A., Davids, K., Renshaw, I., & Araújo, D. (2011). Representative learning design
405   and functionality of research and practice in sport. *Journal of Sport and Exercise*
406   *Psychology*, *33*(1), 146–155. https://doi.org/10.1123/jsep.33.1.146

407   Renfree, A., & Casado, A. (2018). Athletic races represent complex systems, and pacing
408   behavior should be viewed as an emergent phenomenon. *Frontiers in Physiology*,
409   *9*(OCT), 1–6. https://doi.org/10.3389/fphys.2018.01432

410   Renfree, A., Martin, L., Micklewright, D., & St Clair Gibson, A. (2014). Application of
411   decision-making theory to the regulation of muscular work rate during self-paced
412   competitive endurance activity. *Sports Medicine*, *44*(2), 147–158.
413   https://doi.org/10.1007/s40279-013-0107-0

414   Ritschard, G., & Studer, M. (2018). Sequence Analysis: Where Are We, Where Are We
415   Going? In G. Ritschard & M. Studer (Eds.), *Sequence Analysis and Related*

416   *Approaches: Innovative Methods and Applications* (pp. 1–11). Springer International

417   Publishing. https://doi.org/10.1007/978-3-319-95420-2_1

418   Roux, J., Grimaud, O., & Leray, E. (2019). Use of state sequence analysis for care pathway

419   analysis: The example of multiple sclerosis. *Statistical Methods in Medical Research*,

420   *28*(6), 1651–1663. https://doi.org/10.1177/0962280218772068

421   Smits, B. L. M., Pepping, G.-J., & Hettinga, F. J. (2014). Pacing and decision making in sport

422   and exercise: The roles of perception and action in the regulation of exercise intensity.

423   *Sports Medicine*, *44*(6), 763–775. https://doi.org/10.1007/s40279-014-0163-0

424   Sun, L., Guo, T., Liu, F., & Tao, K. (2021). Champion Position Analysis in Short Track

425   Speed Skating Competitions From 2007 to 2019. *Frontiers in Psychology*,

426   *12*(December), 1–7. https://doi.org/10.3389/fpsyg.2021.760900

427   Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

428   Vanasse, A., Courteau, J., Courteau, M., Benigeri, M., Chiu, Y. M., Dufour, I., Couillard, S.,

429   Larivée, P., & Hudon, C. (2020). Healthcare utilization after a first hospitalization for

430   COPD: A new approach of State Sequence Analysis based on the "6W"

431   multidimensional model of care trajectories. *BMC Health Services Research*, *20*(1), 1–

432   15. https://doi.org/10.1186/s12913-020-5030-0

433

Tables

Table 1. Example analysis of winners tactical positioning behaviours ($n = 10$) using discrete lap analysis and sequence analysis

	Discrete Lap Analysis		Sequence Analysis
Race Id	Start Position	Lap 1 Rank	Race Sequence
1	1	3	(1, 3)
2	1	2	(1, 2)
3	1	2	(1, 2)
4	1	3	(1, 3)
5	1	1	(1, 1)
6	3	1	(3, 1)
7	2	1	(2, 1)
8	3	1	(3, 1)
9	3	1	(3, 1)
10	2	1	(2, 1)
Ranked 1st	50%	60%	–
Mean Rank	1.8	1.6	–
Standard Deviation Rank	0.9	0.8	–

Starting positions are coded from 1 (innermost track position) to n (outermost track position). Lap 1 rankings are coded Lap 1 from 1 (leading athlete) to n (last athlete).

Table 2. Most recurring group and winner tactical positioning behaviours in the 500 m, 1,000 m, and 1,500 m events

Event	Race period	Group			Winner		
		Sequence	*Sup*	*relSup*	Sequence	*Sup*	*relSup*
500 m	Start–Lap 5	(1,6)	250	12.4%	(1,6)	960	47.5%
	Lap 1–5	(1,5)	613	30.3%	(1,5)	1,356	67.1%
	Lap 2–5	(1,4)	777	38.5%	(1,4)	1,452	71.9%
	Lap 3–5	(1,3)	1,056	52.3%	(1,3)	1,621	80.2%
	Lap 4–5	(1,2)	1,463	72.4%	(1,2)	1,813	89.8%
	Lap 5	(1,1)	2,020	100%	(1,1)	2,020	100%
1,000 m	Start–Lap 9	–	–	–	(1,10)	126	8.1%
	Lap 1–9	(1,9)	53	3.4%	(1,9)	261	16.8%
	Lap 2–9	(1,8)	81	5.2%	(1,8)	344	22.2%
	Lap 3–9	(1,7)	144	9.3%	(1,7)	472	30.5%
	Lap 4–9	(1,6)	222	14.3%	(1,6)	614	39.6%

	Lap 5–9	(1,5)	336	21.7%	(1,5)	772	49.8%
	Lap 6–9	(1,4)	457	29.5%	(1,4)	946	61.1%
	Lap 7–9	(1,3)	683	44.1%	(1,3)	1,160	74.9%
	Lap 8–9	(1,2)	1,070	69.1%	(1,2)	1,380	89.1%
	Lap 9	(1,1)	1,549	100%	(1,1)	1,549	100%
1,500 m	Start–Lap 14	–	–	–	–	–	–
	Lap 1–14	–	–	–	–	–	–
	Lap 2–14	–	–	–	–	–	–
	Lap 3–14	–	–	–	–	–	–
	Lap 4–14	–	–	–	(1,11)	12	2.1%
	Lap 5–14	–	–	–	(1,10)	20	3.5%
	Lap 6–14	–	–	–	(1,9)	43	7.6%
	Lap 7–14	–	–	–	(1,8)	91	16.1%
	Lap 8–14	–	–	–	(1,7)	139	24.6%
	Lap 9–14	(0.87,2) – (1,4)	18	3.2%	(1,6)	180	31.8%
	Lap 10–14	(0.87,1) – (1,4)	29	5.1%	(1,5)	235	41.5%

Lap 11–14	(1,4)	68	12.0%	(1,4)	307	54.2%
Lap 12–14	(1,3)	135	23.9%	(1,3)	395	69.8%
Lap 13–14	(1,2)	272	48.1%	(1,2)	487	86.0%
Lap 14	(1,1)	566	100%	(1,1)	566	100%

The group behaviour represents the similarity between all athletes' intermediate and final rankings measured using Kendall's Tau-b. The winner behaviour represents the rank of the winner. The state-permanence-sequence format provides each successive distinct state in the sequence with its duration. For example, a winner sequence of (1,6) in the 500 m indicates that the winner started and remained in first for the entirety of the race. Sup = Absolute support; the number of times the sequence occurs in the sequence period. $relSup$ = Relative support; the proportion of races that the sequence occurs in the sequence period. Note, we only report sequences with a relative support $\geq 2\%$, i.e., a sequence that occurs 1 in every 50 races.

Figure 1

Step 1) Extract group and winner tactical positioning behaviours
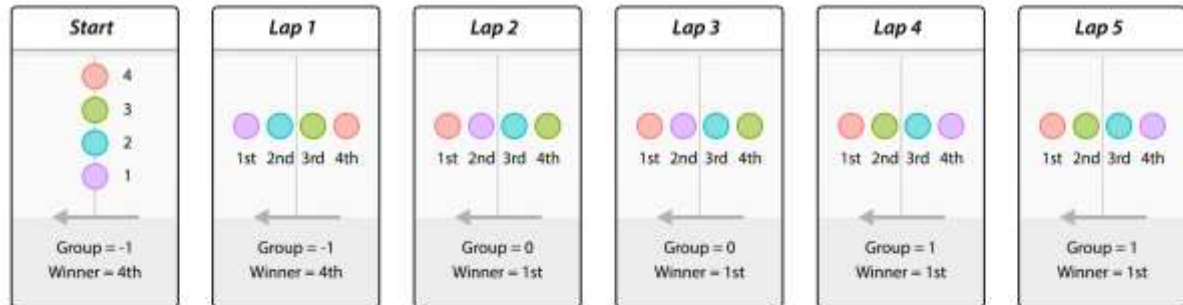
Group behaviour represents the similarity between athletes' intermediate (Start, Lap 1, ..., Lap 4) and final (Lap 5) rankings using Kendall's Tau-b, τ_b. As illustrated above, a $\tau_b = 1$ represents a perfect agreement between rankings (e.g., Lap 4), and a $\tau_b = -1$ represents a perfect disagreement (e.g., Lap 1). Winner behaviour represents the winner's intermediate ranking (e.g., the winner was in 4th position at the end of Lap 1).
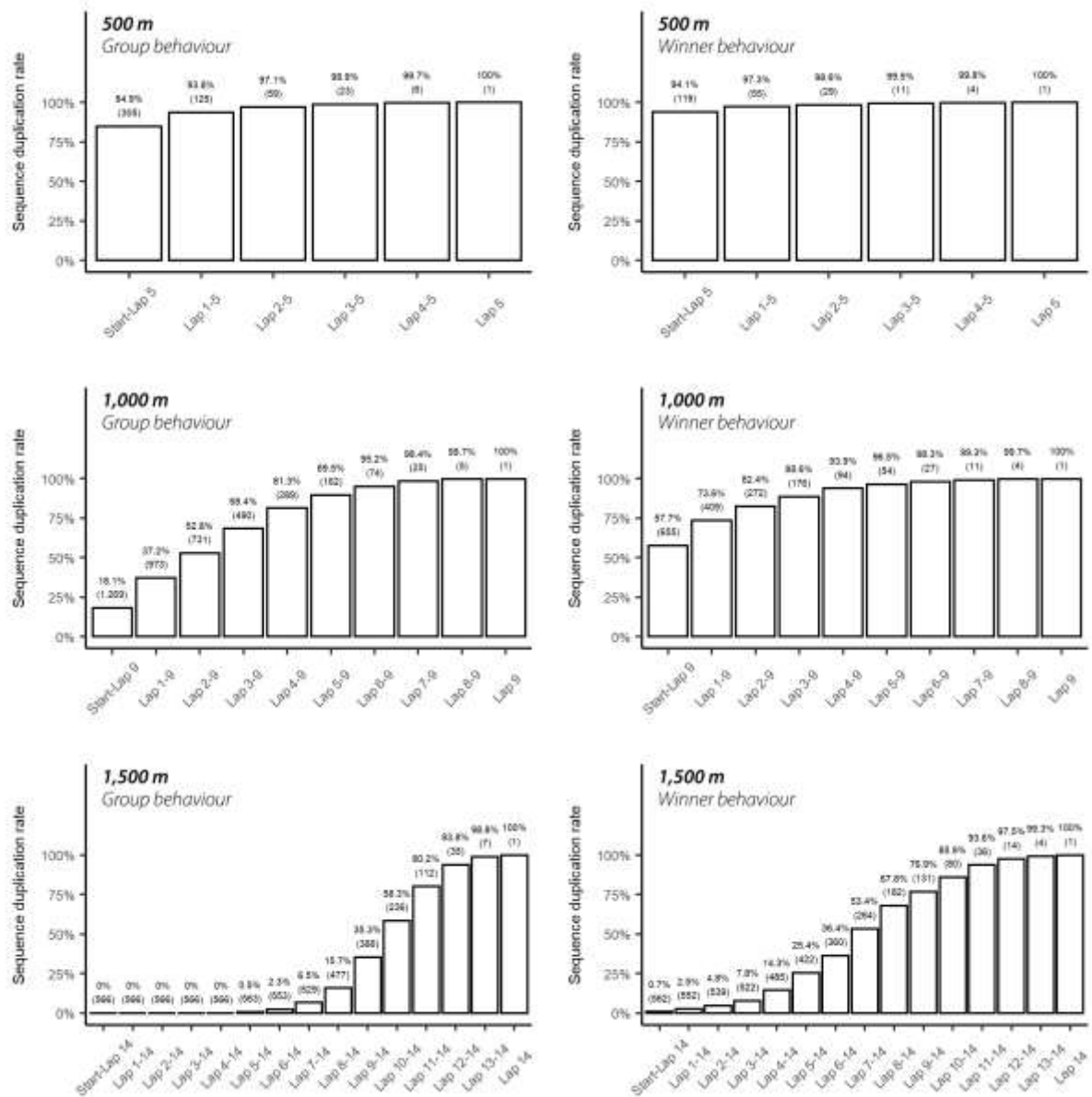
Step 2) Form static and dynamic race sequences
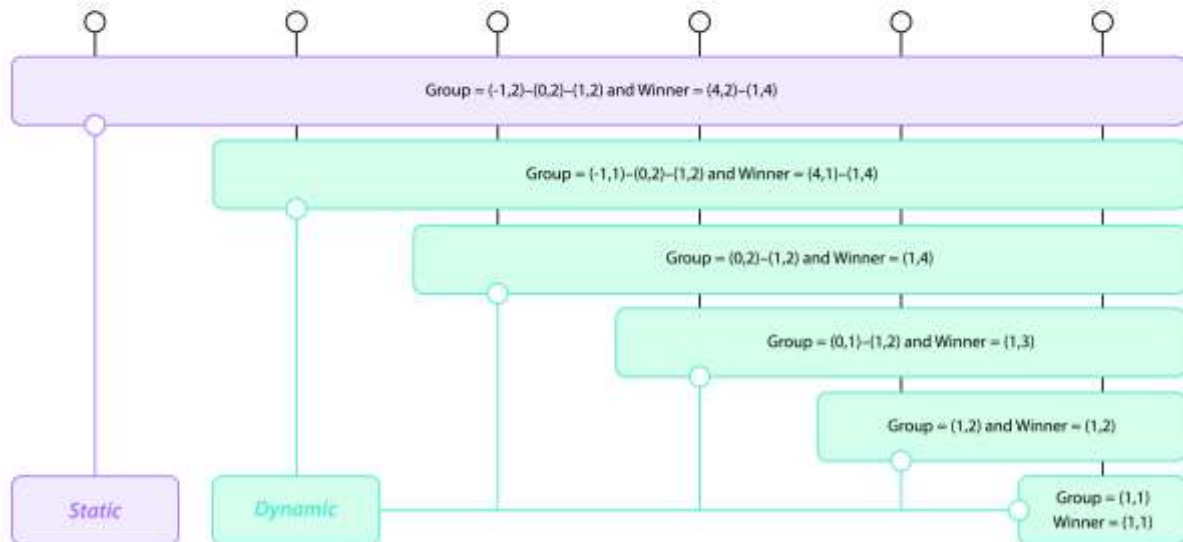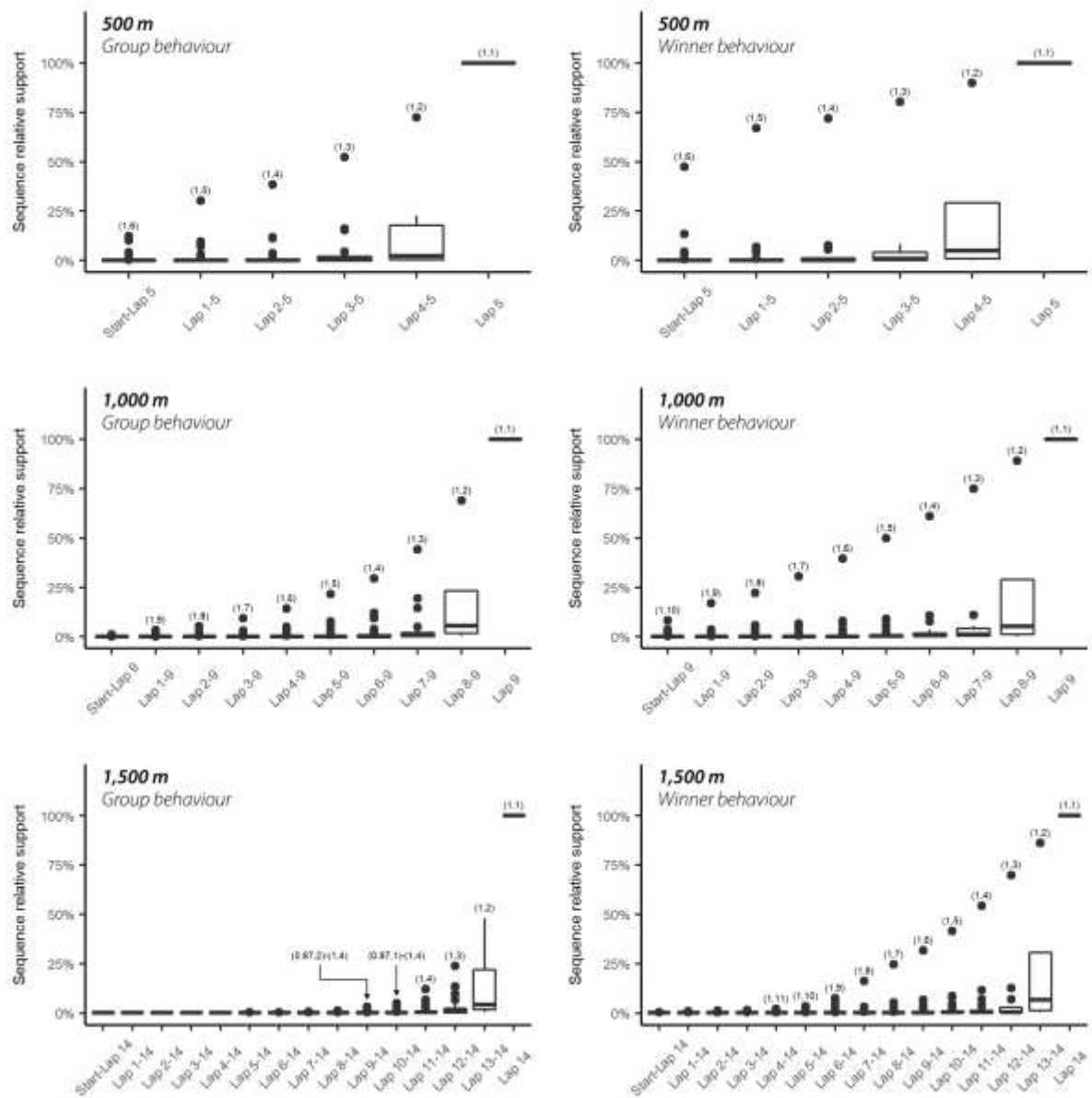
Figure 2

Figure 3

Figure captions

Figure 1. Static and dynamic race sequence formation. First, we extract the winner (denoted by the red circle) and group tactical positioning behaviours at the race start and end of each lap. Second, we form our static and dynamic sequences. The static sequence treats the whole race as a single unit of analysis. The dynamic sequences are nested race sequences with a constant endpoint (the final lap) but varying start points (e.g., Lap 1–5, Lap 2–5, Lap 3–5). Note that we represent all sequences using the state-permanence-sequence format, i.e., each successive distinct state in a sequence is given together with its duration. For example, we represent the static winning sequence: (4, 4, 1, 1, 1, 1), as (4, 2) – (1, 4).

Figure 2. The sequence duplication rate for group and winner tactical positioning behaviours in the 500 m, 1,000 m, and 1,500 m. A sequence duplication rate of 0% indicates that no sequences are identical, and a sequence duplication rate of 100% indicates that all sequences are identical. We report the number of unique sequences detected in the brackets.

Figure 3. Boxplots of sequence relative support for group and winner tactical positioning behaviours in the 500 m, 1,000 m, and 1,500 m. Frequently recurring sequences are identified as those greater or equal to the inner fence: $Q3 + (1.5 * Interquartile\ range)$. Annotated sequences represent the most frequent sequence with a relative support $\geq 2\%$.