Reinforcement learning enhances training and performance in skilled alpine skiers compared to traditional coaching instruction

Christian Magelssen*, Matthias Gilgien[1], Simen Leithe Tajet[2], Thomas Losnegard[1], Per Haugen[1], Robert Reid[3], Romy Frömer[4]

[1*]Department for Physical Performance, Norwegian School of Sport Sciences, Sognsveien 220, Oslo, 0863, Oslo, Norway.
[2]Department of Sport and Social Science, Norwegian School of Sport Sciences, Sognsveien 220, Oslo, 0863, Oslo, Norway.
[3]Department for alpine skiing , The Norwegian Ski Federation, Sognsveien 75 B1, Oslo, 0840, Oslo, Norway.
[4]School of Psychology, University of Birmingham, Edgbaston Park Rd, Birmingham, B15 2TT, Birmingham, United Kingdom.

*Corresponding author(s). E-mail(s): cmagelssen@gmail.com;
Contributing authors: matthiasg@nih.no; simenlt@nih.no; thomasl@nih.no; perh@nih.no; robert.reid@skiforbundet.no; r.froemer@bham.ac.uk;

Abstract

Skilled athletes need powerful movement strategies to solve tasks effectively. Typically, athletes learn these strategies with instruction-based teaching methods where coaches offer athletes a correct solution. Inspired by recent evidence from decision neuroscience, we asked whether skilled athletes learn strategy choices better with an evaluation-based training strategy (reinforcement learning). To address this question, we conducted a three-day learning experiment with skilled alpine ski racers (n=98) designed to improve their performance on flat slopes on slaloms with four strategies at their disposal to achieve this goal. We compared performance and strategy choices of three groups: a reinforcement learning group, that only received feedback about their race times after every run, a supervised (free choice) learning group, that received strategy instructions from their coach, and a supervised (target skill) learning group, being coached to use the theoretically optimal strategy for skiing well on flats. We found that despite making

similar strategy choices, the skiers in the reinforcement learning group, showed greater improvements in their race times during the training sessions than their counterparts in the supervised (free choice) learning group and outperformed them during a subsequent retention test. Surprisingly, the skiers in the reinforcement learning group even showed descriptively (but not significantly) better performance than those in the supervised (target skill) learning group. Our findings show that reinforcement learning can be an effective training strategy for improving strategy choices and performance among skilled athletes, even among the best ones.

Keywords: Reinforcement learning, Supervised learning, Talent development, Expertise development, Alpine ski racing, Reward, Strategy learning

1 Introduction

Skilled performance involves precise and accurate execution of well-chosen strategies[1–10]. For many athletes (even the most skilled ones), these strategies may not always be optimal and can hold them back from unleashing their full potential. In those cases, athletes may need to try new strategies and learn their values to achieve further improvements. Unfortunately, many athletes are stuck on plateaus because they lack knowledge of better alternatives and rely on conventional teaching methods that focus on refining existing strategies [11–15]. Are there better training methods that can improve elite performance, by teaching athletes new strategies' values in more efficient ways than conventional teaching? If so, these training methods could greatly impact the training of present an d future learners.

In contemporary teaching practice, athletes commonly learn which movement strategy to execute with instruction-based teaching. That is, a coach or teacher advises the athlete 'what to do' (e.g., suggesting taking a shorter line around the gate) followed by corrective feedback (e.g., encouraging the learner to further shorten the line) [16–18]. This teaching strategy can be likened to supervised learning in motor learning, where the difference between the desired strategy and the athlete's movement outcome represents a teaching signal for skill improvement [2, 19, 20]. Through practice, this teaching signal can bring the athlete closer to executing what is assumed to be the correct choice. But is this teaching strategy truly the most effective approach for enabling learners to select and employ effective strategies?

Although supervised learning serves as a reasonable approach to improve athletes' strategies, the teaching strategy has limitations that may prevent athletes from converging on an optimal strategy. First, the coach's advice may not be the optimal solution, as what coaches judge as a good strategy does not always align with reality, even for the best-trained eye [21, 22]. Athletes might therefore miss opportunities to discover the best strategy when coaches opt for suboptimal strategies [12, 13]. Worse, these learned suboptimal strategies might turn into habits that can be difficult to break [23]. Second, athletes trained through supervised learning might also be constrained to adopting a single ('universal') strategy for all situations rather than acquiring a repertoire of strategies and discerning the most effective strategies for

each specific scenario. Finally, it remains uncertain whether the prescriptive approach is the most effective teaching strategy for enabling athletes to achieve long-lasting learning effects [16, 17, 24, 25].

Athletes can also learn to choose effective strategies without depending on direct advice from a coach. The cornerstone of reinforcement learning [26] is that athletes can learn by exploring strategies and evaluating their outcomes, using the successes and failures of outcomes as teaching signals. That is, the athlete learns the value of different strategies which allows them to finally pick the best solution, rather than being told the putatively correct solution to the problem. Specifically, these values are learned by comparing a given choice's outcomes with the choice's currently expected outcome. Outcomes that exceed or fall short of expectations result in errors in reward prediction, signaling that the learner must update their predictions to better anticipate future rewards following that action [27]. These reward prediction errors are then incorporated to form a new and better estimate of expected reward by updating expectations through a weighted running average. The computations underlying this form of value-learning of choices have been tremendously powerful in explaining human and animal learning [28–33] as well as training AI to perform complex tasks such as computer games starting from pixel inputs, only[34]. In addition, reinforcement learning has been shown to improve motor skill learning on simple tasks performed in the laboratory [35–38]. Based on this evidence, we asked whether reinforcement learning offers a better alternative for training learners to make better decisions about strategies than traditional supervised learning with a coach.

To address this question, we conducted a three-day learning experiment with ninety-eight skilled alpine ski racers. To achieve performance improvement among this skilled cohort of athletes, we focused on improving flat sections in slalom, an area with considerable potential for enhancement, even among the best skiers[39], and delineated four strategies, each carefully selected to enhance performance in this section by taking advantage of principles from mechanics (Fig. 2a). To study how different instructions and feedback influenced the learning and choices of these strategies, we assigned the skiers to three different learning groups (Fig. 2b). In the supervised (target skill) learning group, we recruited ski coaches to instruct skiers to select the strategy that we defined as the theoretically best strategy based on computational modeling [40–42] and field observations of elite skiers [43, 44]. This group serves as a benchmark for the upper limit of performance achievable through optimal strategy choices and thus offers an upward comparison for the performance yielded through reinforcement learning of strategy selection. In the supervised (free choice) learning group, we recruited ski coaches from the tested ski teams to train them on the strategy they believed to be the best or most appropriate for the skier. Coaches in the two supervised learning groups saw the skiers' times but were instructed to not disclose these to the skiers. In contrast, in the reinforcement learning group, the skiers chose a strategy on every trial and saw their race times immediately after each trial to inform them about which strategy to select. We hypothesized that the skiers in the reinforcement learning group would learn to choose better strategies and thus achieve better performance than skiers subject to traditional supervised learning with

a coach (supervised learning: free choice), possibly reaching similar performance as the supervised (target skill) learning group that uses the best strategy throughout.

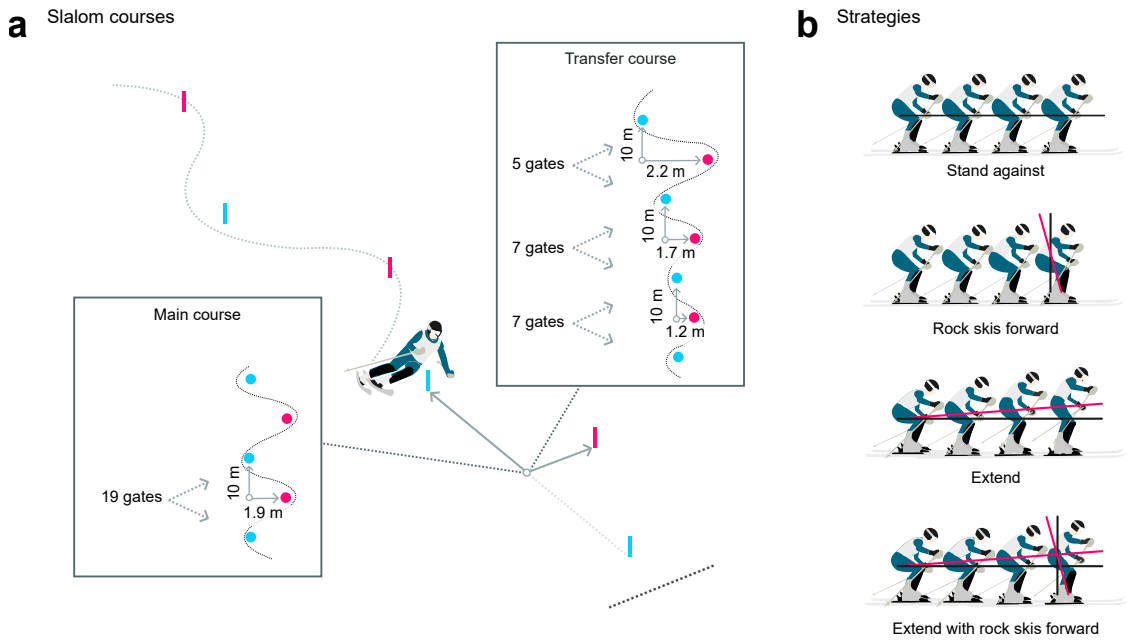

Fig. 1 **a.** Illustrations of the two slalom courses used in the study. The main slalom course was a rhythmic course deployed in all sessions except for the transfer session. The course setting for the transfer session involved a progression in gate offset, starting with the largest offset and ending with the smallest offset. **b.** Illustration of the strategies defined to enhance racing performance on flat terrain in slalom: The "stand against" strategy emphasized maintaining a stable stance against external forces without body extension along the body's longitudinal axis or rocking skis forward; "Rock skis forward" involved rocking skis forward from gate passage to completion of the turn; The "extend" strategy involves extending the body from a laterally tilted position during the turn, closer to the turn's center of rotation; The "extend with rocking skis forward" was expected to be the best strategy combining the two effects from extending and rocking skis forward, and we therefore defined this as the theoretical best strategy

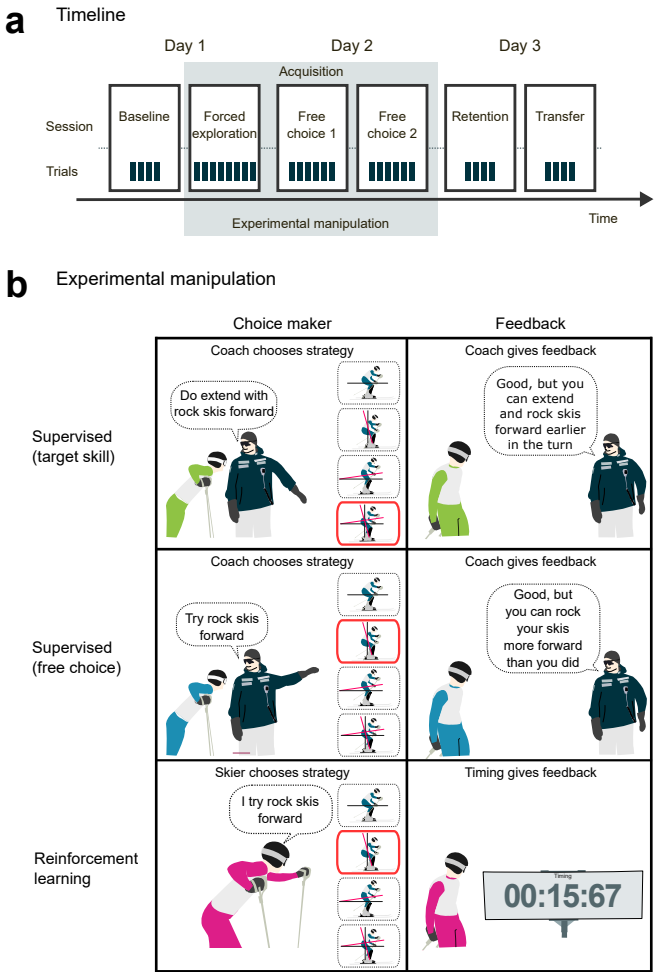

Fig. 2 Illustration of the experimental design and procedure. **a.** Timeline of the three-day learning experiment. During the baseline session, the skiers raced a slalom course in the shortest time possible without receiving race time feedback. The skiers were then assigned to one of three learning groups (see b). In their assigned group, the skiers underwent three acquisition sessions, comprising one forced exploration session (skiers performed all strategies) and two free choice sessions (skiers or coaches could choose strategies themselves). On the last day, skiers completed a retention and transfer test where they could pick strategies themselves, again without receiving race time feedback. **b.** Illustration of the learning groups in the study. The supervised (target skill) learning group involved coaches consistently choosing the theoretically best strategy (except during forced exploration), while in the supervised (free choice) learning group coaches freely selected strategies. The skiers in both of these learning groups received feedback on strategy execution from their assigned coach, while the skiers in the reinforcement learning group independently selected strategies and received feedback from the timing system to facilitate value learning of each strategy.

2 Method

2.1 Participants

Conducting studies on alpine ski racing poses challenges related to environmental control and resource constraints. Our sample size approach was justified by the 'resource constraints' and the 'whole population' criteria [45] and involved recruiting as many skiers as possible during June 2023, when we had a short time window to test skiers in the indoor ski hall. We set the minimum sample size to 80 skiers, which we deemed appropriate for this context. Prior to data collection, data and power simulations for sample sizes of 80, 100, and 120 skiers were conducted, revealing simulated powers of 0.60, 0.75, and 0.80, respectively, for the smallest effect size of interest (0.3 second difference between groups) (https://osf.io/c4t28). The smallest effect size of interest was based on our knowledge of alpine skiing and discussions with coaches, but it was intended for a 50-meter longer course and more training sessions than we ultimately ended up using due to practical considerations. We deliberately opted to recruit skiers with diverse skill levels for the study to augment the generalizability of our findings. However, to ensure a sufficient skill level to handle the specific icy snow conditions prepared in the ski hall, we recruited only skiers aged 15 and older. The sample size justification, task design, and analysis plan were preregistered before data collection (https://osf.io/tfb2w).

We managed to recruit ten alpine ski teams comprising 98 alpine ski racers from Norway and Sweden (age M = 18.1 years, SD= 2; 40 females, 58 males). Two skiers were excluded from the analysis due to an injury prior to the study (n=1) or sickness during the study (n=1); thus, a total of 96 skiers completed the entire study and were included in the analysis. Among the ski groups tested were five ski academies, three senior development teams, and two national ski teams. These skiers were generally highly skilled, with a median world rank of 605, but there was also considerable variability, as indicated by a substantial interquartile range (Q1 = 248, Q3 = 1390.5). A smaller subset of the participants (n = 13) were not world-ranked, as they had yet to compete in internationally sanctioned races that form the basis for calculating athlete points and rankings. Table 2.1 provides demographic information for each learning group.

In addition, we recruited coaches for the two supervised learning groups. This sampling process was carried out pragmatically due to space and time constraints in the ski hall, which made it impossible to test all ski teams simultaneously. The 10 ski teams with alpine skiers that were recruited were therefore divided into 4 groups for which the study was conducted at different times. For each of the four groups, we recruited two coaches from the ski teams to serve as coaches in the supervised (free choice) learning group, totaling 8 coaches (2 women; 6 men). These coaches had extensive coaching experience in coaching alpine ski racers. In addition, for each group of ski teams that completed the experiment together, we recruited a third coach to instruct the skiers in the supervised (target skill) learning group. To ensure that these coaches had sufficient credibility to make the skiers buy into our theoretical best strategy, we selectively recruited three highly experienced coaches from the Norwegian alpine ski team (one coach served twice). Importantly, all the coaches remained

unaware of the experimental manipulation. Table 2.1 provides demographic information about the coaches. All the skiers and coaches provided informed consent before the study. The study was approved by the Human Research Ethics Committee of The Norwegian School of Sport Sciences (ref. 279-040523) and the Norwegian Agency for Shared Services in Education and Research (ref. 871468).

2.2 The setup

The experiment was conducted in the indoor ski hall SNØ in Oslo, Norway (https://snooslo.no/). In this hall, we used a 210-meter-long flat slope section of the race hill, which we water-injected before testing each group of skiers to ensure uniform and fair snow conditions for all skiers (Supplementary B for detailed description). With our chosen course setup, this 210-meter-long flat section allowed space for 19 slalom gates.

We used two types of slalom courses in the experiment (Fig. 1a). The main slalom course was used in all sessions, except during the transfer test, and featured a 10 m distance and a 1.9 m offset. The course distance aligned with our previous study [44], but we opted for a slightly larger offset to better suit the skill level of our skiers. The transfer test evaluated how well the skiers transferred their learning to a new slalom course more realistic to a typical alpine ski race course. To assess this, we set a course with a progression in gate offset, starting with five gates at a 2.2-meter offset, followed by seven gates at a 1.7-meter offset, and concluding with seven gates at a 1.2-meter offset. Although we did not expect radical differences in strategy effects, we anticipated a greater emphasis on rocking the skis forward in gates with a 2.2-meter offset than in those with a 1.2-meter offset to enhance turn exit release. Both courses were set with stubbies (short gates) instead of long gates to minimize energy dissipation upon hitting the gate [46]. Using long gates can also be a distracting element in that skiers' attention is allocated to clearing the gate instead of focusing on executing the skill. Finally, this approach helped us avoid creating holes in the course, which can occur when the long gate is forcefully slammed into the ground. To minimize wear and tear on the course, we set two parallel and identical courses and routinely shifted between them.

The start gate was positioned 20 meters before the first gate. Skiers were required to start in a static position to ensure consistency in the starts, with the toe piece of the binding placed behind the starting gate. The skiers started by putting their skis in parallel and lifting the poles without using poling or skating for propulsion (See https://osf.io/9numq for a Supplement Video illustrating the starting procedure and setup). We recorded the times using a wireless photocell timing system (HC Timing wiNode and wiTimer; Oslo, Norway). Timing started when the skier crossed the first photocell pair situated 10 meters below the starting gate.

2.3 Experimental design

We employed a between-subjects design and posed the learning question of discovering effective strategies as an n-armed bandit problem [26]. The essence of this problem is that a learner repeatedly tries different options and observes their outcomes to learn which strategy is the best and, therefore, which one to choose. Finding the best

Table 1 Skier characteristics for each learning group

	Reinforcement learning		Supervised (free choice)		Supervised (target skill)	
Sex	**F**, N = 13	**M**, N = 19	**F**, N = 14	**M**, N = 19	**F**, N = 13	**M**, N = 20
Age	18.4 (2.3)	17.7 (1.8)	18.1 (2.3)	18.3 (1.8)	17.8 (2.4)	18.2 (2.1)
Training group						
National team	1 (7.7%)	1 (5.3%)	2 (14%)	1 (5.3%)	2 (15%)	4 (20%)
Senior team	3 (23%)	5 (26%)	1 (7.1%)	5 (26%)	0 (0%)	5 (25%)
Ski academy	9 (69%)	13 (68%)	11 (79%)	13 (68%)	11 (85%)	11 (55%)
FIS points						
Median (IQR)	54 (42, 80)	46 (34, 90)	58 (26, 80)	44 (35, 61)	49 (28, 66)	31 (28, 63)
Not recorded	3	2	1	2	2	3
World ranking						
Median (IQR)	630 (394, 1,217)	707 (364, 2,317)	709 (133, 1,224)	662 (387, 1,274)	527 (145, 882)	314 (220, 1,360)
Not recorded	3	2	1	2	2	3

[1] Mean (SD); n (%)

Table 2 Coach characteristics for each learning group

Characteristic	Supervised (free choice)		Supervised (target skill)
	F, N = 2	**M**, N = 6	**M**, N = 3
Age	38.5 (3.5)	44.3 (8.8)	48.0 (7.0)
Ski education (highest achieved)			
Level 2	0 (0%)	1 (17%)	
Level 3/4	2 (100%)	5 (83%)	3 (100%)
Sport science degree (highest achieved)			
MSc	1 (50%)	0 (0%)	1 (33%)
BSc	1 (50%)	1 (17%)	1 (33%)
No	0 (0%)	4 (67%)	
One-year program	0 (0%)	1 (17%)	1 (33%)
Coaching experience (years)			
National team (WC/EC)/Senior teams	5.00 (1.41)	5.00 (4.24)	15.67 (1.15)
Ski academy	7.5 (3.5)	7.5 (6.5)	2.00 (3.46)
Ski club	2.5 (3.5)	6.5 (8.7)	6.0 (6.6)

[1] Mean (SD); n (%)

strategy requires a delicate balance between exploiting the strategy known to yield the best payoff and exploring alternative strategies that may offer superior benefits. In our study, the options consisted of four strategies that skiers could employ to improve their race times on flat slopes in slalom, grounded in physics-based coaching manuals for alpine ski racing [40, 41, 47, 48], biomechanical research on elite skiers [43, 44, 49] and common strategies used by coaches. The four strategies were named "stand against", "rock skis forward", "extend", and "extend with rock skis forward (see Fig. 1b for a strategy illustration and Supplementary A for an extended explanation). To study how instruction and feedback drove learning to select effective strategies, we designed and allocated skiers to three learning groups, which allowed us to compare reinforcement learning with traditional supervised learning with a coach:

For the supervised (target skill) learning group, we provided the best possible training program by engaging highly experienced coaches who explained to the skiers that the 'extend with rock skis forward' strategy was the most effective for skiing fast on flat terrain in slalom, citing evidence from research literature in alpine skiing mechanics [40, 41, 49]. The coach then instructed the skiers to adopt this strategy and provided feedback on its execution after each trial. Note that the coach had access to the skiers after each trial but was prohibited from communicating the race time with the skiers.

In the supervised (free choice) learning group, the skiers were assigned to two coaches recruited from the tested group of ski teams. To balance the skiers' skill levels between the two coaches, we created new blocks from the ranked list from baseline testing and randomly assigned them to the coaches. We instructed the coaches to improve the skiers' race times as much as possible with our four defined strategies to their disposal to achieve this goal. For each trial, the coach selected a strategy for the skier, observed the skier during the trial and provided feedback on its execution afterward. Similar to the supervised (target skill) learning group, the coaches had access to the skiers' data after each trial but could not share this information with the skiers.

In contrast, the reinforcement learning group was not assigned to any coach. Instead of having a coach deciding the skiing strategy for them, the skiers in this learning group were told to choose a strategy for each trial by themselves to ski the course as fast as possible. To help the skiers choose and learn from evaluation, this group could see their racing times immediately after they crossed the finish line. Although this group had no coach, we assigned a person to communicate with the skiers to record their choices and encourage them to try skiing quickly to prevent boredom effects.

2.4 Procedures

Fig. 2a illustrates the procedures employed in the study. In the baseline test, the skiers began with two warm-up runs: one in a free skiing warm-up course and one in a specific warm-up in the slalom course. During these warm-up runs, skiers were instructed, trained and verified on the start procedure by an instructor. As a first run in the baseline assessment, skiers completed a straight-gliding run, where they skied straight down from start to finish in a static, upright slalom posture. Subsequently, the skiers completed four runs in the slalom course. Skiers were encouraged to ski as

quickly as they could, but they could not see their times, nor did they receive any instructions on how to perform well.

After the initial baseline assessment, the skiers took a 60-minute break. In the meantime, we allocated the skiers to the three learning groups deploying a randomized-blocked approach to account for pre-existing differences in the skiers' performance levels [50]. Specifically, we computed each skier's average across the four trials at baseline and ranked them accordingly. We then created n blocks with block sizes corresponding to our three learning groups for the entire list of skiers and assigned these skiers to these predefined blocks. Finally, we randomly allocated the skiers to the different learning groups within each block (Fig. 2b).

The learning groups participated in sessions at different times to prevent treatment diffusion [50]. As the ski group comprised teams of skiers who knew each other well and who resided together, we explicitly emphasized the importance of keeping information about the sessions private. To stay within the time frame at the ski hall, the two supervised learning groups underwent training together. To facilitate this, we arranged stations in the finishing area with space and vision dividers and ample space to impede communication between coaches in the supervised learning groups (Supplementary **??**). In addition, we developed a Python script that fetched the race times from the timing system, filtered the times for each coach, and transmitted them to the station where the coach was located, ensuring that no information leaked between the coaches. The learning group (reinforcement learning versus supervised learning) that initiated after the baseline session was randomized and counterbalanced across the group of ski teams we tested.

The first session after learning group assignment involved a forced exploration. Here, skiers within the learning group were gathered, and the session started by introducing them to the strategies. We explained that we had identified four strategies to enhance racing times on flat slopes in slalom. Subsequently, each strategy was detailed, supported by illustrative drawings in 1b Figure and corresponding word explanations as outlined in the Supplementary A. To confirm comprehension, we conducted two short familiarization trials for each strategy or until the execution met our performance standards. After reviewing the strategies with the skiers, we gathered them in their respective learning group and asked them to rank the strategies (1=best; 4=worst) for what they believe to be the best strategies for improving race times in the flat section of a slalom course. Throughout the instruction and ranking process, skiers were explicitly instructed not to discuss the strategies with each other. It is important to note that the same instructor was used for all learning groups within the tested ski group. After this, the skiers conducted a total of eight trials on the course, with two trials for each strategy. Here, we handled the practice order effect probabilistically by randomizing the order for each skier, under the condition that the first and last four trials tested all strategies. During these trials, the reinforcement learning group received feedback on the timing, whereas the coach provided feedback in the supervised learning groups. After completing the eight rounds, the skiers re-evaluated and ranked the strategies.

On the second day, the skiers completed two free choice sessions, each comprising a total of 6 trials in the same slalom courses that were used for the baseline testing

the day before. Prior to the 6 free choice trials, the skiers performed one warm-up free skiing run and one warm-up run in the slalom course. In these sessions, the supervised (target skill) learning group consistently selected the theoretically best strategy (that is, 'extend with rock skis forward'). Conversely, in the supervised (free choice) learning group and the reinforcement learning group, the coach and the skiers, respectively, had the autonomy to choose the skiing strategy for each run. After each session, coaches (except supervised target skills) and skiers were asked to re-evaluate and rank the strategies.

On the third and last day, the skiers performed a retention test and a transfer test to assess the effect of the training approaches on learning and performance. The retention test was performed in the same course as the baseline and acquisition sessions, whereas the transfer test was performed in the transfer course and involved a progression in gate offset from start to finish. Since the transfer test was a new course, we allowed the skiers to inspect the course before the test. The retention and transfer tests were conducted with the three learning groups together. None of the learning groups received any feedback from coaches or time during these tests. After each test, the skiers were asked to rank the strategies.

2.5 Analysis

The data were cleaned with custom functions built on tidyverse [51] packages in R [52]. After this process, we validated the data by performing trial counting and visual inspection of the race time to screen for errors. An extensive report of this cleaning and validation process can be found at OSF (https://osf.io/2jxgk).

Due to the hierarchical structure of the data, our general statistical strategy relies on multilevel modeling. At the first level, each skier performed multiple trials during each session. At the second level, each skier was nested within groups of ski teams that performed the experiment together. To account for these multilevel data structures, we leveraged linear mixed-effects models. To model random effects, we adopted a design-driven approach [53, 54], where we sought to account for all nonindependence introduced by repeated sampling from the same ski group and skier. We deployed classical frequentist statistics and fitted these models with the lme4 package [55] in the R [52] programming language. We used a simple coding scheme for our predictors where the intercepts represent the estimated mean of the cell means and the contrasts represent the estimated difference with respect to the reference level, which we set for reinforcement learning. Two-tailed p values and degrees of freedom for each model were derived using the lmerTest package [56] via the Satterthwaite approximation method. Alpha was set to 0.05 for all test statistics.

2.5.1 Race time

Race times were analyzed using linear mixed-effect regression models. Initially, we planned to normalize the racing times by expressing the racing times as the difference from the straight-gliding time performed at the beginning of every session. This difference better approximates the skiers' actual skill improvement by considering the variance in snow conditions. However, practical considerations led us to deviate from

this approach. This change was necessary because we had to flip or shift the course after each day to ensure snow conditions with the least damage. Unfortunately, these adjustments made maintaining a clean, straight-gliding lane difficult since the straight gliding lane crossed many areas with damage to the snow surface (holes) from the previous course set (see Supplementary C for an image of these holes). Collisions with these holes affected the race time, adding noise to the results. Therefore, we used a more conservative approach and analyzed the raw racing times instead of analyzing the normalized racing times.

For the acquisition session, we modeled race time using Session (forced exploration, free choice 1, free choice 2) and Treatment (reinforcement learning, supervised: free choice learning, supervised: target skill learning), and their interactions, as predictors. For retention and transfer, we modeled racing times at these sessions, with treatment added as a predictor. In addition, we used the average performance for each skier on the baseline test as a predictor to improve estimate precision and adjust for any group differences at baseline testing.

To model the effect and development of the strategies we broke the analysis up into different sub-models. One analysis focused on differences in the strategies and groups regarding Forced Exploration, where all participants had completed all strategies. Another analysis examined the transition from Forced Exploration to retention for both supervised (free choice) and reinforcement learning. The final model investigated the development between groups specifically for the "extend with rock skis forward" strategy. Session was coded as a continuous variable in all models.

2.5.2 Strategy choices

Strategy choices were analyzed using generalized linear mixed-effect regression models with a binomial logit-link function. To model the selection of the theoretical best strategy, we inputted the data as logistic, where for each trial (i) per skier (j) within ski group (k), we counted $y_{ijk} = 1$ when the skier chose the theoretical best strategy (that is, 'extend with rock skis forward') or 0 when they did not. We included Treatment (reinforcement learning, supervised: free choice learning, supervised: target skill learning) and Session (free choice 1, free choice 2, retention, transfer) and their interaction as two variables. To account for the nonindependence of the data structure, we allowed the intercept to vary by including a random intercept for the skier and ski groups.

We adopted the same model formula to model the selection of the estimated best strategy. This time, however, we counted $y_{ijk} = 1$ when the skier chose their estimated best strategy and 0 when they did not select that strategy. To estimate the best strategy for each skier in the sample, we used the sample-average method [26] to average the race time for each strategy and selected the strategy with the lowest estimated (that is, best) value. The sessions that we used to form this average were Forced Exploration, Free Choice 1, and Free Choice 2. Due to the scaling issues with generalized linear models, we followed the recommendation to determine the size and significance of the effects of interest using marginal effects on the probability scale [57, 58]. Interactions were assessed using discrete difference (also second difference),

which is also in line with these recommendations. To derive these estimates, we used the emmeans package [59].

To learn how the skiers and coaches used feedback to guide their choices, we constructed a 'win-stay, lose-shift' model [60–62]. For this analysis, we z-scored the race times for each skier for Free Choices 1 and 2 and counted $y_{ijk} = 1$ when the skier repeated the previous strategy and 0 when they did not. The data were modeled using a generalized linear mixed-effect regression model with a binomial logit-link function, with Treatment and z-transformed Race Time and their interaction as the two variables. To test for differences in error sensitivity, we used the marginal effects at the mean derived from the emmeans package [59].

2.5.3 Strategy evaluations and outcomes

To analyze the strategies' rankings, we used single-level linear regression owing to the singularity of our multilevel models. In this model, we inputted Session as a continuous variable (with the intercept anchored at Familiarization) and Treatment (reinforcement learning, supervised: free choice learning, supervised: target skill learning) as the predictors. For supervised (free choice) learning, we used the coaches' rankings during the sessions where they selected strategies, and we used the skiers' rankings when they selected strategies during the retention and transfer tests.

To evaluate how the race times evolved over the next sessions, we had to break down the analysis into three sub-analyses because the supervised (target skill) learning group only performed 'extend with rock skis forward' during free choice 1 and free choice 2. First, we built at model to assess the difference between the strategies and group differences during forced exploration. To this end, we included Treatment (all learning groups) and Strategy, and their interaction, as factors. For this particular model, we used sliding contrast coding scheme. Second, we built a model to test for differences in improvement on 'stand against', 'rock skis forward', and 'extend' only for the reinforcement and supervised (free choice) learning groups. To this end, we inputted Session as a continuous variable (with the intercept anchored at Familiarization) and Treatment as the predictors. Finally, we built a model to assess the development for all groups on the "extend with rock skis forward" strategy. This model consisted of Session as a continuous variable (with the intercept anchored at Familiarization) and Treatment as the predictors.

3 Results

3.1 Race times

We assumed that the strategy choice would greatly impact race time in the slalom course and that the reinforcement learning group would learn to select better strategies and consequently perform better over the course of the experiment compared to the supervised (free choice) learning group where we used the skiers' own coaches. The supervised (target skill) learning group was instructed to choose the optimal strategy and therefore served as a benchmark for the maximum expected benefit of

improved strategy selection through reinforcement learning. As our first step, we analyzed whether the learning groups showed pure time differences across the different sessions without taking the chosen strategy into account.

3.1.1 Greater performance improvement for reinforcement learning compared to supervised (free choice) learning

If the reinforcement learning group learned to select better strategies than the supervised (free choice) learning group, we would expect them to exhibit greater improvement over the course of the three acquisition sessions. Therefore, our hypothesis was that the reinforcement learning group would improve more during the acquisition sessions than the supervised (free choice) learning group. Further, we would expect the supervised (target skill) learning group to rapidly improve race time when the skiers were instructed to only select the theoretical best strategy during the last two acquisition sessions (free choice 1 and free choice 2). We expected that differences in improvement between the learning groups would emerge once the skiers or the coach had the autonomy to pick the strategy over the course of the second (free choice 1) and third (free choice 2) acquisition sessions. Importantly, the three learning groups performed similarly during the forced exploration where they had the same number of trials on each strategy; we found no statistically significant differences between the reinforcement learning group and the supervised (free choice) learning group (β = 0.06, 95% CI [-0.2, 0.32], t(92.727) = 0.48, p = 0.631) or between the supervised (target skill) learning group (β = 0.18, 95% CI[-0.08, 0.44], t(92.663) = 1.37, p = 0.174) during the first acquisition session (forced exploration).

All learning groups significantly improved their race times over the course of the free choice sessions during acquisition (Supplementary Table E1). As expected, the rate at which they improved, differed across the three groups during these two sessions. In line with our expectations, the supervised (target skill) learning group, in which skiers were coached to solely select the theoretically best strategy, showed a statistically significantly greater improvement than the reinforcement learning group from forced exploration to free choice 1 (β = -0.12, 95% CI[-0.22, -0.03], t(91.777) = -2.58, p = 0.012). Conversely, the supervised (free choice) learning group demonstrated a descriptively poorer progression than the reinforcement learning group, albeit the difference was not statistically significant (β = 0.08, 95% CI[-0.02, 0.17], t(92.5) = 1.61, p = 0.110). Continuing this trend, comparing initial performance to performance in the final acquisition session (free choice 2), the reinforcement learning group improved significantly more than the supervised (free choice) learning group did (β = 0.14, 95% CI[0.02, 0.26], t(95.743) = 2.26, p = 0.026). For the same comparison, the supervised (target skill) learning group no longer improved their race times significantly more than the reinforcement learning group (β = 0.02, 95% CI[-0.11, 0.14], t (95.651) = 0.26, p = 0.798). This is due in part to the continued improvement of the reinforcement learning group but also to a descriptive decline from free choice 1 to free choice 2 in the supervised (target skill) learning group, attenuating their initially greater improvement rate. However, we did not find statistical evidence that the reinforcement

learning group performed better than the supervised (free choice) or supervised (target skill) learning groups at free choice 1 or free choice 2 (Supplementary Table E2). Fig. 3a presents the mean race time estimates during the three acquisition sessions.

3.1.2 Superior retention performance for reinforcement learning compared to supervised (free choice) learning

In the retention session, the skiers independently chose their strategies, irrespective of their assigned groups. We reasoned that at this point, the reinforcement learning group had learned to understand which strategies were most effective and chose them. Therefore, our hypothesis was that the reinforcement learning group would outperform the supervised (free choice) learning group in retention due to better strategy selection learned from observing the race times to evaluate the strategies. Consistent with this hypothesis, we found that the reinforcement learning group performed significantly better than the supervised (free choice) learning group did (β = 0.12, 95% CI[0.01, 0.24], t(101.422) = 2.12, p = 0.037). The difference between the reinforcement learning and supervised (target skill) learning groups also favored reinforcement learning but was smaller and not statistically significant (β = 0.07, 95% CI[-0.04, 0.19], t(101.63) = 1.27, p = 0.206). We therefore provide evidence that the reinforcement learning group performed better at retention than did the supervised (free choice) learning group. Descriptively, the reinforcement learning group performed better, even than the supervised (target skill) learning group, which served as an upper bound reference. Fig. 3b presents the mean race time estimates during retention.

We also hypothesized that reinforcement learning would improve skill transfer to a new slalom course compared to the supervised (free choice) learning group. Similar to retention, the reinforcement learning group performed better than the supervised (free choice) group, yet the difference was smaller and not statistically significant (β = 0.1, 95% CI[-0.02, 0.21], t(99.979) = 1.7, p = 0.091). The race times for reinforcement learning and supervised (target skill) learning groups were on average identical (β = 0, 95% CI[-0.12, 0.11], t(100.033) = -0.04, p = 0.967). Thus, we did not find corroborating evidence for improved transfer. Fig. 3c presents the mean race time estimates during transfer.

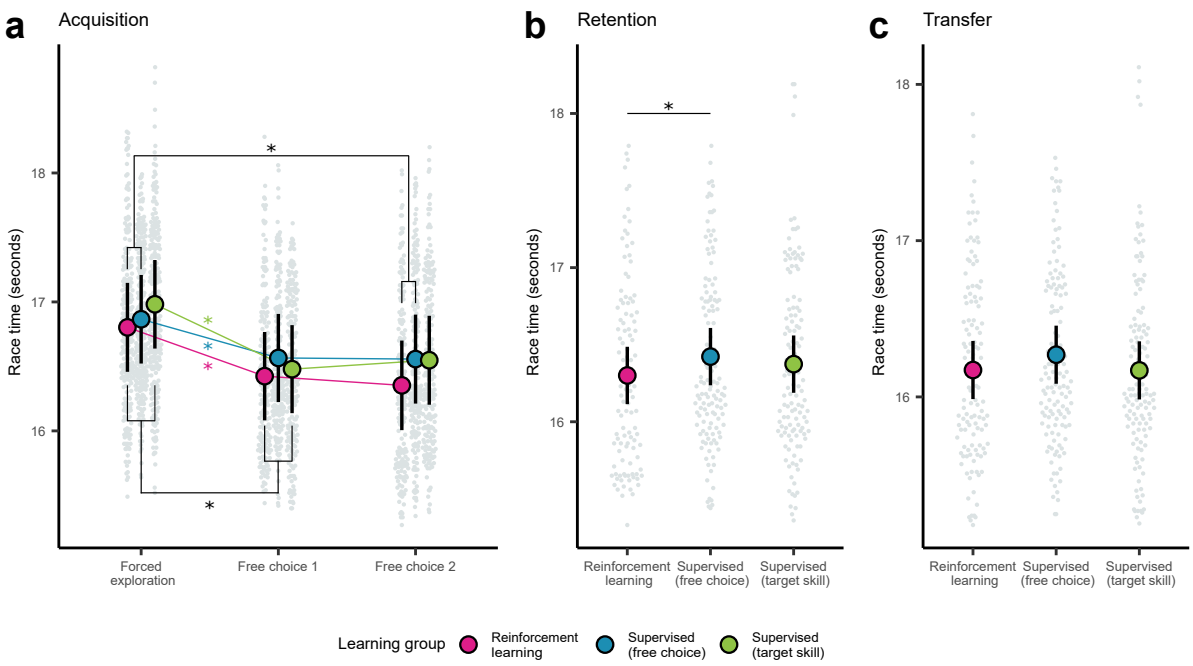

Fig. 3 Race time across the different sessions for the three learning groups **a**. Estimated race times during the three acquisition sessions. Forced exploration refers to the sessions wherein skiers tried all strategies, whereas free choice 1 and free choice 2 refer to the session wherein skiers or coaches selected strategies according to their assigned learning groups. **b.** Estimated race times for retention. **c.** Estimated race times for transfer. Intervals represent the 95% confidence intervals (CIs) derived from the models. Asterisks (*) indicate a statistically significant effect. Each light gray point represents a single trial performed by a skier.

3.2 Strategy choices

We proposed that the differences in race time between the reinforcement learning group and the supervised (free choice) learning group could be explained by the choice of strategy. Specifically, we hypothesized that the reinforcement learning group would learn to choose better strategies than the supervised (free choice) learning group by learning the strategies' values directly from observing race times. Fig. 4 displays the percentage selections of the four strategies across all sessions.

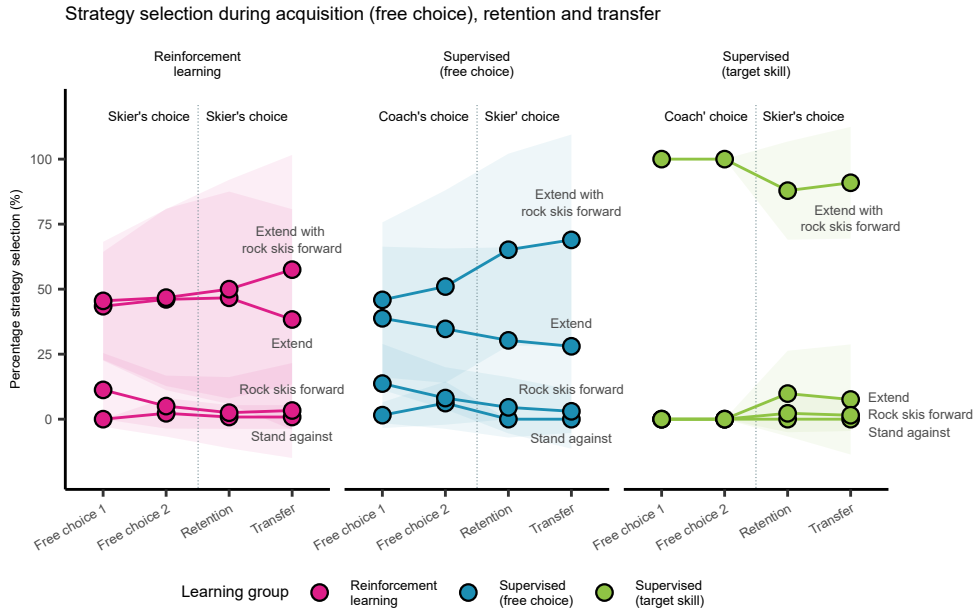

Fig. 4 Percentage of choices for each strategy and session (circles). Shaded error bars represent standard deviations.

3.2.1 No greater increase in selecting the theoretical best strategy for reinforcement compared to supervised (free choice) learning

We assumed that the skiers performed best with the strategy we defined as the theoretically best strategy (that is, with the "extend with rock skis forward") and tested whether the reinforcement learning group developed a greater probability of selecting this strategy than the supervised (free choice) learning group. We found that both learning groups showed a statistically significant increase in the probability of choosing the strategy we considered theoretically optimal across the four sessions (Supplementary Table E3); however, the timing of this increase differed between the two groups. For the supervised (free choice) learning group we found a statistically significant increase from free choice 1 to retention, when skiers in this group were given autonomy to choose strategies themselves (0.31, 95% CI[0.18, 0.44], $z = 4.59$, $p < 0.001$). This probability increase was significantly greater than the increase for the reinforcement learning group (0.24, 95% CI[0.05, 0.43], $z = 2.43$, $p = 0.015$), which did not significantly increase from free choice 1 (0.07, 95% CI[-0.07, 0.21], $z = 0.96$, $p = 0.339$). The reinforcement learning group, on the other hand, significantly increased the probability of choosing the theoretically best strategy from free choice 1 to the transfer session(0.18, 95%CI[0.04, 0.32], $z = 2.58$, $p = 0.010$). Despite the descriptively higher probability of choosing the theoretically best strategy in the supervised (free choice) learning group during the free choice 2, retention and transfer sessions, none of the differences between the groups at each session were statistically significant

(Supplementary Table E4). Fig. 5a displays the predicted probabilities for the learning groups across the four sessions where the skiers or coaches were given autonomy to select strategies themselves. Note that the supervised (target skill) learning group was excluded from this analysis because by definition only the theoretically best strategy was selected.

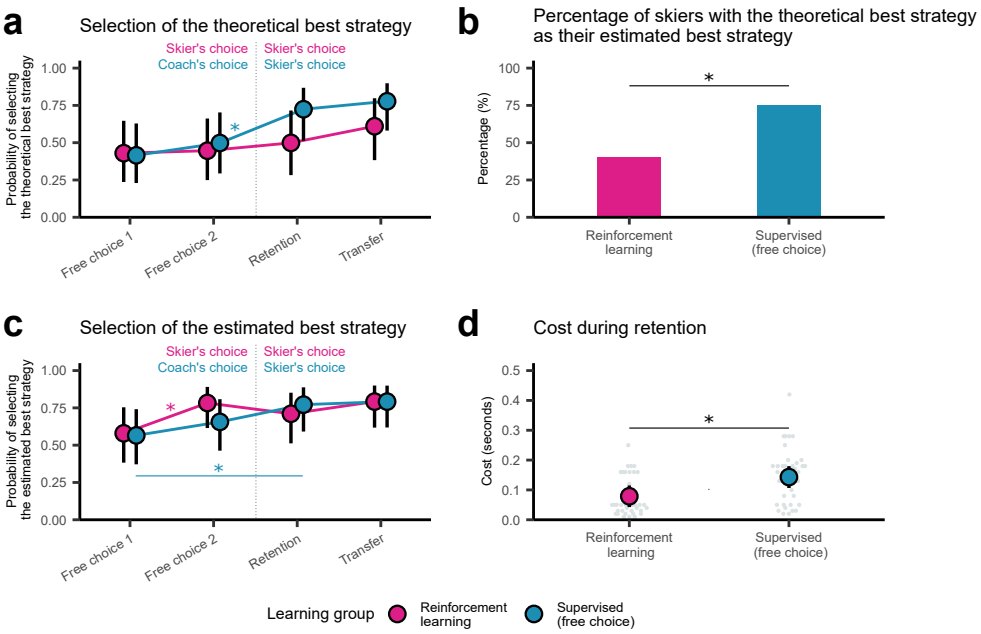

Fig. 5 Strategy selection for reinforcement learning (pink) and supervised (free choice) learning (blue) groups during acquisition. **a.** Predicted probability of choosing the theoretically best strategy (that is, 'extend with rock skis forward'). **b.** Percentage of skiers performing best using the theoretically best strategy. **c.** Predicted probability of selecting the individual skier's estimated best strategy. **d.** Estimated cost of selecting a suboptimal strategy during retention. **a., c.** Intervals represent the 95% confidence intervals (CIs) derived from the models. Asterisks (*) indicate a statistically significant effect.

3.2.2 Differences in strategy performance between reinforcement and supervised (free choice) learning groups

Based on the previous analysis, the reinforcement learning group did not develop a greater probability of selecting what we defined as the theoretically optimal strategy compared to the supervised (free choice) learning group. Instead, the descriptive trend favored supervised (free choice) learning. Therefore, we wondered whether a greater proportion of skiers in the supervised (free choice) learning group had the theoretical best strategy as their estimated best strategy, and performed a follow-up analysis to test this possibility. Overall, 75% of the skiers in the supervised (free choice) learning group performed best using the theoretically best strategy compared to only 40% in

the reinforcement learning group. A chi-square test revealed a statistically significant difference between groups $\chi^2 = 6.42$, $p = 0.01$). Fig. 5b shows the proportion of skiers for each group that performed best with the theoretical best strategy. Note that skiers in the supervised (target skill) group only used the theoretically best strategy, so they could not be included in this analysis. The observation that more skiers in the supervised (free choice) learning group had the theoretically best strategy as their estimated best strategy could explain why they had a descriptively greater probability of choosing it compared to the reinforcement learning group.

3.2.3 No greater increase in selecting individual skier's estimated best strategy for reinforcement compared to the supervised (free choice) learning

We did not find evidence that the reinforcement learning group developed a greater predicted probability of choosing the theoretically optimal strategy. However, this strategy was not the best strategy for every skier. Therefore, what if we instead base the analysis on the individual skier estimated best strategy? Would the reinforcement learning group have a higher predicted probability of choosing this strategy? During the first session, in which we allowed skiers and coaches to choose their own strategies (free choice 1), we found no statistically significant differences between the groups (0.01, 95% CI[-0.22, 0.24], $z = 0.12$, $p = 0.904$). Both groups significantly improved their choices over the course of the sessions relative to free choice 1 (Supplementary Table E5), but the significant improvement came at different time points. We found a statistically significant improvement in choice for the reinforcement learning group from free choice 1 to free choice 2 (0.2, 95% CI[0.09, 0.32], $z = 3.39$, $p < 0.001$), but not for the supervised (free choice) learning group (0.09, 95% CI[-0.03, 0.21], $z = 1.48$, $p = 0.140$). However, the reinforcement learning group did not increase their probability of choosing their individually best strategy significantly more than the supervised (free choice) group (0.11, 95% CI [-0.28, 0.05], $z = -1.33$, $p = 0.184$). For the supervised (free choice) learning group we found a statistically significant improvement in their choices when the skiers made their own strategy choices during retention (0.21, 95% CI[0.08, 0.34], $z = 3.11$, $p = 0.002$). We found no statistically significant differences between the groups in any of the sessions (Supplementary Table E6). Therefore, we did not find corroborating evidence that the reinforcement learning group learned to pick the best strategy more often than the supervised (free choice) learning group. Fig. 5c displays the predicted probabilities of choosing the estimated best strategy for the learning groups across the sessions.

3.2.4 Lower cost of choosing suboptimal strategies for reinforcement compared to supervised (free choice) learning during retention

Selecting an effective movement strategy can extend beyond the mere selection of the single best strategy, however. Sometimes, several strategies produce outcomes that are nearly identical, giving athletes the flexibility to select any option without sacrificing performance. The task then becomes one of selecting strategies that offer comparable outcomes while avoiding those that carry the risk of substantially worse performance.

We reasoned that discerning differences between fairly similar strategies could be tricky for skiers due to small time differentials that are influenced by noise, such as hitting a bump. Consequently, two or more strategies may appear nearly identical and be difficult for skiers to distinguish. We proposed that the reinforcement learning group had learned to select good strategies that all led to similarly good outcomes and learned to steer away from those that led to poor outcomes. To test this idea, we calculated the expected difference between a skier's chosen 'suboptimal' strategy and their estimated best strategy, which we termed 'cost.' Using this measure, we ran a follow-up analysis to test whether the reinforcement learning group had a lower 'cost' during retention than did the supervised (free choice) learning group. We limited the testing to retention, as it was only in this session that the skiers in the supervised (free choice) learning group were allowed to choose their own strategy and who were in the same slalom course they had previously skied. This analysis revealed that the reinforcement learning group had significantly lower costs of selecting suboptimal strategies during retention than the supervised (free choice) learning group ($\beta = 0.06$, 95% CI [0.01, 0.12], $t(30.789) = 2.55$, $p = 0.016$). This result suggests that skiers in the reinforcement learning group may have learned to select better strategies, although we did not find evidence that they had a greater probability of picking the best strategy. Fig. 5d displays the expected cost for the suboptimally chosen strategies.

3.2.5 Similar use of feedback for strategy selection in reinforcement and supervised (free choice) learning groups

Finally we examined the extent to which decision-makers (skiers in reinforcement learning and coaches in supervised learning) engaged with their feedback times to inform decisions about which strategy to select next. Our hypothesis was that, compared with coaches, skiers in the reinforcement learning group would engage more actively with these race times to improve their choices. To test this hypothesis, we conducted a 'win-stay, lose-shift' analysis, where the primary assumption is that all information used for decision-making stems from the last trial (n-1)[61, 62]. In this analysis, heightened sensitivity to the most recent feedback is reflected by a high predicted probability of repeating an action following positive feedback (a fast time, compared to the average time) and a low predicted probability following negative feedback (a slow time, compared to the average time) on the preceding trial. We found statistically significant estimated marginal effects at the mean for both the reinforcement learning group (-0.18, 95% CI[-0.26, -0.11], z = -4.8, $p < 0.001$) and the supervised (free choice) learning group (-0.11, 95% CI[-0.17, -0.04], z = -3.29, $p < 0.001$). These findings suggest that both groups were more likely to repeat a strategy if the previous trial feedback was good. Despite the large descriptive difference in the marginal effect between groups, this difference was not statistically significant (-0.08, 95% CI [-0.17, 0.02], z = -1.55, $p = 0.121$). Thus, although we found that both learning groups were sensitive to the most recent feedback (either using it themselves to determine choices, or having the coach use it to determine choices), we did not find evidence that the reinforcement learning group had greater sensitivity than the supervised (free choice) learning group. Fig. 6 shows the predicted probability of repeating the strategy on the previous trial depending on the feedback.

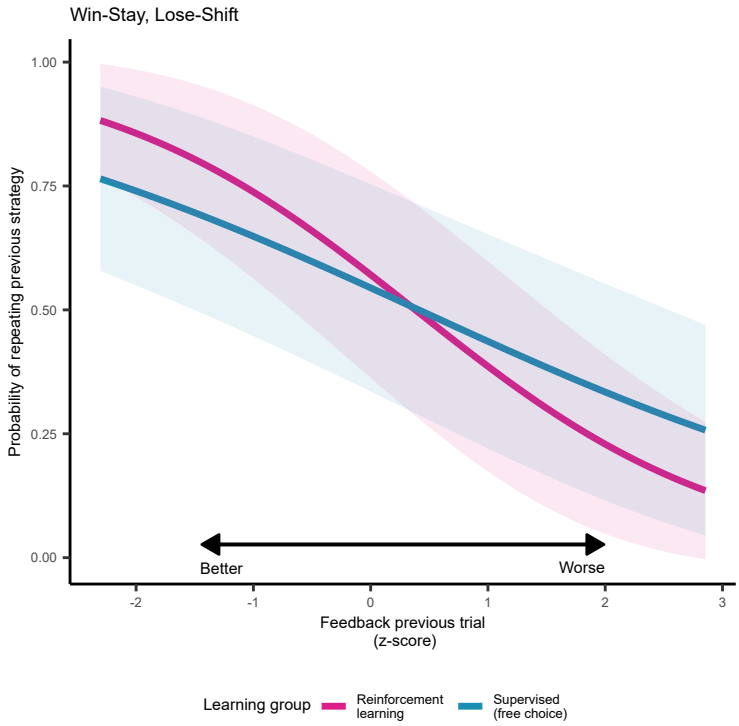

Fig. 6 Win-stay, lose-shift comparison between reinforcement learning and supervised (free choice) learning. The line shows the predicted probability of repeating the previously chosen strategy based on its trial feedback, along with a 95% CI in the shaded error bar

3.3 Strategy evaluations and outcomes

As a final step, we assessed how the learning groups' understanding of the strategies evolved throughout the experiment and how well this understanding aligned with the actual times achieved for the four strategies. The purpose of this analysis was to learn how the learning groups' understanding of the strategies governed their choices. To assess this, we divided the analysis into two parts. First, we assessed the learning group's evaluation of the strategies. Second, we assessed the effect of the learning group's race on the participants' strategies.

3.3.1 Greater separation in strategy evaluations for reinforcement compared to supervised (target skill) learning

To assess how the learning groups' knowledge of the strategies evolved and governed their choices we asked the skiers and coaches (excluding coaches involved in supervised (target skill) learning) to evaluate the strategies by ranking them from best (1) to worst (4). They first ranked the strategies upon their introduction (familiarization) to

the strategies and then after each session. We expected that the reinforcement learning group would develop more distinct evaluations of the strategies compared to the supervised learning groups by having the race times at their disposal to evaluate the strategies. To evaluate this, we gathered the coaches' rankings in the supervised (free choice) learning group for analysis during the two acquisition sessions, as they were responsible for selecting the strategy during these sessions. Conversely, we collected rankings from the skiers in the supervised (free choice) learning group for analysis during the retention and transfer sessions, as they were responsible for making decisions in these sessions. For supervised (target skill) learning, we exclusively focused on the skiers' rankings since the coach had been influenced by the information we gave them about the strategies before the experiment.

After the introduction to the strategies, but before the skiers had the chance to properly try out the strategies on the slalom course (familiarization), we found that all groups ranked 'extend with rock skis forward' as the best, followed by 'extend,' 'rock skis forward,' and 'stand against' (Supplementary Table E7 and Fig. 7a). We did not find any statistically significant difference between groups during this familiarization (Supplementary Table E8), with two exceptions: the supervised (target skill) learning group ranked 'extend' worse (β = 0.45, 95% CI[0.24, 0.67], $t(1700)$ = 4.17, $p < 0.001$) and 'extend with rock skis forward' better (β = -0.42, 95% CI[-0.63, -0.2], $t(1700)$ = -3.83, $p < 0.001$) than the reinforcement learning group.

Over the course of the sessions, we observed notable shifts in the average rankings of the strategies (Supplementary Table E9). The change was relatively flat and unchanged in terms of position for the worst ('stand against') and best ('extend with rock skis forward') strategies. However, we found that 'stand against' was ranked significantly worse (β = 0.09, 95% CI[0.04, 0.13], $t(1700)$ = 3.4, $p < 0.001$) and that 'extend with rock skis forward' was ranked significantly better (β = -0.06 , 95% CI[-0.11, -0.01], $t(1700)$ = -2.51, p = 0.012) over time in the supervised (target skill) learning group. Although the reinforcement learning and supervised (free choice) learning groups followed the same trend, these trends were not statistically significant. More marked shifts were observed for the two middle-ranked strategies: 'extend' and 'rock skis forward'. Specifically, the reinforcement learning group ranked 'extend' significantly better (β = -0.1, 95% CI[-0.15, -0.05], $t(1700)$ = -3.73, $p < 0.001$) and 'rock skis forward' significantly worse (β = 0.09, 95% CI[0.04, 0.15], $t(1700)$ = 3.57, $p < 0.001$) over the course of the sessions. The supervised (target skill) and supervised (free choice) groups also had the same trend, but the magnitude was smaller and did not reach statistical significance (Supplementary Table E9). Interestingly, we found that there was a greater change in the ranking of 'rock skis forward' in the reinforcement learning group than in supervised (target skill) learning group (β = -0.07, 95% CI[-0.14 to 0], $t(1700)$ = -1.97, p = 0.049), suggesting a larger shift in strategy ranking for this strategy in the reinforcement learning group. Overall, these results show that the groups updated their evaluations of the different strategies based on their experience and that the reinforcement learning group did so more than the supervised (free choice) learning group.

3.3.2 Strategy performance mirrors strategy evaluation

We would expect that, at least toward the later sessions, these evaluations reflect the race times for each strategy. To determine whether the pattern in strategy ranking aligned with race time for the strategies, we tested how performance changed across sessions for each strategy. Overall, race times for each strategy mirrored each group's strategy evaluations (Fig. 7b). Note that skiers' evaluations are also reflected in the fact that there are no race times for the worst-rated strategy during retention and transfer in any of the groups, because skiers did not select this strategy. During the first acquisition session when skiers tested all strategies (forced exploration), we found that all groups on average performed descriptively better with 'extend with rock skis forward', followed by 'extend', 'rock skis forward', and finally 'stand against' (Supplementary Table E10). This finding largely aligns with the rankings given for each strategy during this session (see section 3.3.1).

We found no statistically significant differences in performance between the reinforcement learning and supervised learning groups on any of the strategies (Supplementary Table E11). To evaluate how the race times evolved over the next sessions, we had to break down the analysis into two subanalyses because the supervised (target skill) learning group only performed 'extend with rock skis forward' during free choice 1 and free choice 2. First, we assessed the development for all groups on the "extend with rock skis forward" strategy. This analysis revealed that all groups improved on the "extend with rock skis forward" strategy over the course of the sessions (Supplementary Table E14). However, we found no statistically significant differences in skill improvement for this strategy between the reinforcement learning group and either supervised (free choice) learning group (β = 0, 95% CI[-0.03, 0.02], t(1064.204) = -0.3, p = 0.765) or supervised (target skill) learning group (β = 0.02, 95% CI [-0.01, 0.04], t(1062.577) = 1.07, p = 0.283). This result shows that performing all trials with the optimal strategy did not further improve the performance of the supervised (target skill) learning group. Second, we tested for differences in improvement on 'stand against', 'rock skis forward', and 'extend' only for the reinforcement and supervised (free choice) learning groups. We found a statistically significant improvement in all strategies over the course of the sessions to retention, except for 'stand against' in the reinforcement learning group (Supplementary Table E12). This deviation, however, may be attributed to the limited number of observations for this strategy within this group.

Interestingly, we found that the reinforcement learning group improved more on the "extend" strategy than did the supervised (free choice) learning group (β = 0.04 , 95% CI[0.01, 0.07], t(1378.879) = 2.34, p = 0.020). This result aligns well with the reinforcement learning group's evaluation of this strategy, which they ranked significantly better over the course of the sessions, whereas the supervised (free choice) learning group did not(see above); and this result might explain the reduced cost of selecting a suboptimal strategy. Notably, the reinforcement learning group was not assigned a coach to help them improve on that strategy, but could only rely on their race times. We did not find evidence for any interaction effect for the other two strategies (Supplementary Table E13). Taking the evaluations and outcomes collectively, it seems that these evaluations reflected the race times well for each strategy.

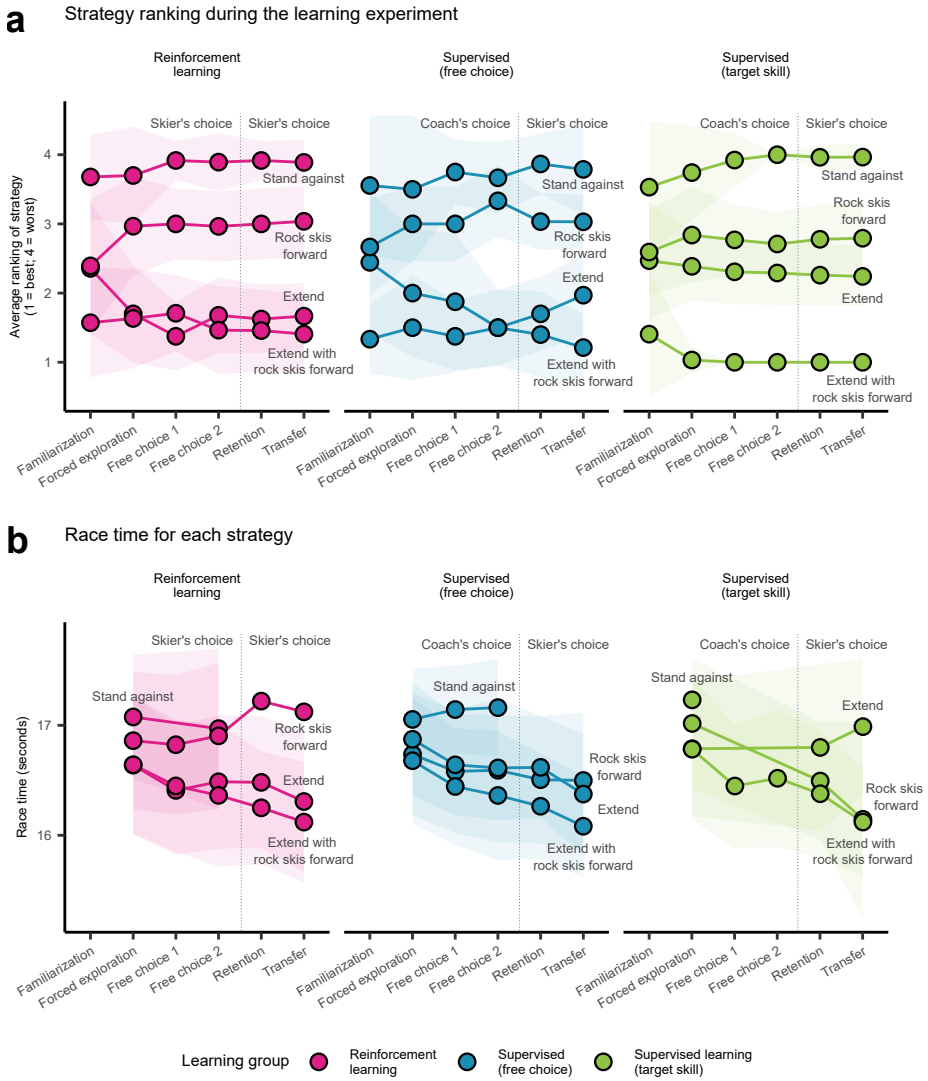

Fig. 7 Strategy evaluations and performance. **a.** Average descriptive ranking of the four strategies per treatment group. Rankings range from 1 (best) to 4 (worst). For supervised (free choice) learning, the coach's ranking during the acquisition phase and the skier's ranking during the retention and transfer phases are plotted, reflecting the decision-maker for strategy selection. The circle represents the mean, and the shaded area indicates the standard deviation (SD). **b.** Average race time of the four strategies across the three learning groups. The circle represents the mean, and the shaded area represents the SD. Note that all skiers tested the strategies during the forced exploration phase, but as the study progressed, there may have been fewer observations for some strategies. Consequently, the calculation of the mean might be heavily influenced by these observations.

4 Discussion

Skilled athletes need to skillfully and adaptively select movement strategies to effectively solve tasks. Typically, athletes learn these strategies with instruction-based teaching methods where coaches offer athletes a correct solution. Inspired by insights from decision neuroscience, we asked whether skilled athletes can instead improve strategy decisions and consequently enhance performance by shifting from an instruction (supervised learning) to an evaluation-oriented (reinforcement learning) training strategy. In a three-day skill learning experiment, we found that the reinforcement learning group achieved greater improvement over the course of the acquisition sessions and performed better during the retention session than did the supervised (free choice) learning group, where the coach selected a strategy for the skier. Interestingly, the reinforcement learning group also performed descriptively better than the supervised (target skill) learning group, which was meant to serve as a benchmark for the upper limit of performance achievable through optimal strategy choices. However, we found less convincing evidence corroborating our hypothesis that reinforcement learning improved transfer.

At a broad level, our findings dovetail with two research streams that have shown that withholding instruction or feedback can be an effective way to enhance skill learning. First, our finding that the reinforcement learning group showed better skill performance during retention than the supervised (free choice) learning group echoes earlier results from 'discovery learning' approaches showing that instructional methods do not always benefit learning motor skills [24, 25, 63, 64]. Second, this finding mirrors previous studies that have found better skill retention from reinforcement learning-based interventions [35, 37, 38, 65]. We expand these two lines of research to a complex sport involving skilled athletes and show that teaching strategies other than instruction can be effective for skill learning.

We proposed that the reinforcement learning group's improved performance compared to the supervised (free choice) learning group resulted from better strategy choices. However, we found no corroborating evidence for this hypothesis when better strategy choices were defined as choices of a single best strategy; the reinforcement learning group did not begin with or develop a greater probability of choosing the theoretically optimal strategy or the individual skier's estimated best strategy compared with the supervised (free choice) learning group. Instead, we found that the skiers in the reinforcement learning group who made 'suboptimal' choices during the retention session incurred a lower 'cost' for their choices than did the skiers in the supervised (free choice) learning group. The reinforcement learning group therefore appears to have learned to select strategies that offered comparable outcomes and to avoid those that carried the risk of substantially worse performance. One possibility is that the reinforcement learning group discovered that the "extend" strategy alone almost provided full benefits on its own and chose this strategy because it was simpler than "extend with rock skis forward".

Improved strategy selection appears not to be the sole explanation for the improved performance of the reinforcement learning group, however. Two other key findings were that the reinforcement learning group had a lower proportion of skiers who performed best with 'extend with rock skis forward' than did the supervised (free

choice) learning group, and that the reinforcement learning group improved more on the 'extend' strategy over the course of the sessions than did the supervised (free choice) learning group. In conjunction these findings suggest that the reinforcement learning group developed the 'extend' strategy in an unanticipated direction based on the race time feedback. Comments from a few coaches, who watched the retention and transfer from the sideline, mentioned that skiers in the reinforcement learning group used more forceful arm movements than did the skiers in the other groups, although the instructions did not explicitly tell them to do that. As such, it seems that the reinforcement learning group refined the 'extend' strategy to a new level. Unfortunately, we were not able to precisely quantify how this 'new' strategy was executed with our measurements.

What information guided the strategy decisions in the two learning groups? We found that both learning groups' decisions could be accounted for by a 'win-stay, lose-shift' heuristic, where the decision maker had a greater probability of continuing with a chosen strategy if the performance was better than the skier's average and conversely was more likely to switch strategies following negative feedback. Such 'win-stay, lose-shift' patterns have also been found in a previous study on skill learning [66] and appear to be a general characteristic of learning to make good decisions to achieve goals. Note that it is possible that learners used more that just the most recent feedback to inform strategy selection, as would be expected in reinforcement learning. Differentiating between 'win-stay, lose-shift' and more complex learning and decision-making strategies is beyond the scope of the current manuscript. Although the reinforcement learning group showed increased sensitivity to the most recent feedback compared to the supervised (free choice) group, this difference was not statistically significant. A potential reason for this is that the coaches in the supervised (free choice) learning group were highly experienced coaches with a great deal of knowledge, and they also had access to the time. Future studies might consider investigating different coach groups and with and without access to times to improve the understanding of how these decisions are made.

Contrary to our expectation, we did not find convincing evidence that reinforcement learning improved transfer compared with either of the two supervised learning groups. One explanation for this could be that reinforcement learning only improves performance in situations where rewards have been previously received [67], and aligns with [38] who found that reinforcement learning improved retention but not transfer compared with supervised learning. However, we can not argue for differential effects on retention and transfer in our data. The estimated mean difference between the reinforcement learning group and the supervised (free choice) learning group was quite similar to that during the retention session (0.12 sec. on retention versus 0.1 sec. on transfer), and the p value was also low (0.091). By definition, athletes are expected to show greater variability when transitioning from retention to transfer testing because the athlete are exposed to a new situation. Therefore, it is possible that our sample does not include a sufficient number of skiers to achieve adequate statistical power to detect such an effect. Alternatively, a more varied learning approach, where learners are exposed to frequent switches between situations (in our case, slalom courses), may be necessary to grasp the task's structure and promote transfer [68, 69]. Future

research should address the link between reinforcement learning and transfer more closely.

Interestingly, despite the broad consensus that 'extend with rock skis forward' was the best strategy upon the introduction of these strategies, coaches and skiers faced greater challenges in determining which of the two strategies, 'rock skis forward' and 'extend', was the second best. Over the course of the next sessions, we found that the reinforcement learning group learned to distinguish these strategies from each other, as evidenced by a significantly improved assessment of 'extend' and a poorer assessment of 'rock skis forward', which we did not find in the other two learning groups. In addition, we found that the reinforcement learning group ranked "rock skis forward" more poorly over the course of the sessions than did the supervised (target skill) learning group. Learners may miss potential learning by only being exposed to one strategy instead of a broader exploration of alternatives [70–72], as was the case with the supervised (target skill) learning group. Being exposed to reinforcement learning-type training may stimulate reasoning processes in learners due to task variability that enables them to cultivate a deeper comprehension of the relationship between their actions and performance outcomes. Over time, improving such reasoning capacity could prove crucial for developing innovative and truly brilliant solutions to achieve task goals [14]. It is possible that reinforcement learning, when used to train learners to choose between two or more explicit strategies, can promote reasoning processes that may be important for developing expertise. Future studies should test this idea.

Our findings have important implications for coaches when designing training sessions to enhance skills. We advise coaches to formulate explicit strategies to enhance performance in their sport and allow learners to directly learn the action values of these strategies by trying them out and comparing them. Doing so could help learners better understand the functions and effects of various strategies and enable them to make better strategy choices. However, we do not propose replacing traditional teaching methods with this approach but suggest integrating it as a supplementary tool to improve strategic decisions and enhance athletes' understanding of how their actions influence outcomes.

Before practitioners embrace our recommendation to incorporate more strategy evaluations into their coaching or teaching practices, it is important to consider the practical significance of the effect size and its potential amplifying and counteracting mechanisms[73]. Notably, the estimated effect size during the retention session was smaller than our predefined smallest effect size of interest. This benchmark, however, was set for a longer slalom course (that is, more gates) and more training sessions than we could execute due to space and time constraints in the ski hall. Consequently, we exercise caution in outright dismissing its practical significance. Instead, if we assume that a full slalom course of approximately 45 seconds has a flat section equivalent to approximately 15 seconds (which is quite typical) and that a slalom race consists of two runs, the 0.12-second difference can translate into an improved Fédération Internationale de Ski (FIS) world ranking of 27 positions for females and 65 for males, based on a median ranking of 600 in our sample (see Supplement Discussion F). Practitioners could therefore achieve meaningful gains by applying reinforcement learning to train strategic decisions. However, we must remember that sporting situations are

diverse and erratic, often requiring a athlete to switch between different strategies during a performance. The ability to successfully implement these adaptive switches is a hallmark of sporting expertise [1, 6, 8]. Our study did not capture such decisions because we focused only on flat slopes. Finally, it is also possible that the estimated group mean differences could have been greater had the reinforcement learning group and the supervised learning groups undergone retention and transfer sessions separately, thereby preventing any information leakage. With our current procedure, it is conceivable that the skiers observed each other, potentially leading to treatment diffusion. However, this decision was made to mirror the conditions of alpine competitions and gave us confidence that athletes experienced similar conditions during testing.

4.1 Limitation of the study

We did not have motion capture data available to analyze the skiers' execution of the strategies. Therefore, we cannot comment on the actual execution of the strategies, other than verifying that they were able to perform the strategies during the familiarization test. However, taking into account execution might have led to explosion of actually implemented strategies that would then be difficult to study. Indeed, considering a small set of discrete strategies, as we did in this study, has been the recommended approach for studying action selection for motor skill learning [9, 74]. Another limitation is the relatively short duration of the learning experiment. The idea, however, was to mimic a typical short training session for alpine skiers and study whether we could find meaningful effects within this short time window.

5 Conclusion

In summary, our data showed better learning with reinforcement learning than with the traditional instruction-based coaching approach, but the advantage did not transfer to a new slalom course. Only by informing coaches about what we believed to be the best strategy for improving race time on flat sections in slalom (based on mechanics and evidence from the field) were we able to achieve learning effects comparable to those of reinforcement learning. Interesting, although the supervised (target skill) learning group always picked the correct strategy, they did not improve more on this strategy than the reinforcement learning group. In accordance with previous suggestions [75, 76], our findings suggest that reinforcement learning can be an important learning strategy for improving strategy choices and learning in skilled athletes.

Supplementary information.

Acknowledgements. We thank SNØ (https://snooslo.no/) and Igloo Innovation (https://iglooinnovation.no/) for allowing us to use the indoor ski hall and facilities for testing. We thank Christian Mitter for helping us define the strategies and giving us ideas for how to communicate them to the skiers. We thank Erland Hoff Thommasen, Jeremy Phyffer, Mai-Sissel Linløkken, Erland Vedeler Stubbe, Johan Ansnes, Ørjan Lydersen, William Farstad Olsen, Victoria S. Placht, Einar Witteveen, Synne Sofie Stangeland, Kasper Sjøstrand, Tom Knudsen, and Tim Gfeller for helping with

watering the hill and assisting during the data collection. We thank Tine SA for providing the skiers with food and nutrients during the data collection. We thank all the ski coaches and skiers who participated and made this study possible. We also want to thank Andrea Pisauro, Doron Cohen, Jørgen Jensen, and Christian Thue Bjørndal for helpful comments and feedback on our preprint. The first author would also like to give special thanks to Cameron Patrick, Isabella R. Ghement and James Steele for their general methodological and statistical discussions.

Declarations

5.1 Data and code availability

All data and code are available on OSF under this link: https://osf.io/qdk5t/

5.2 Author contribution

Christian Magelssen: Led conception and design, led the data collection, analyzed and interpreted the data, drafted and revised the article, and agreed to the submitted version for publication. Matthias Gilgien: Contributed to the design, data collection, and revised and agreed to the submitted version for publication. Simen Leithe Tajet: Contributed to the design, data collection, interpreting of the data, and agreed to the submitted version for publication. Thomas Losnegard: Contributed to the design, data collection, and revised and agreed on the submitted version for publication. Per Haugen: Contributed to the design, data collection, and agreed to the submitted version for publication. Robert Reid: Contributed to the design, data collection, interpretation of the data, revised, and agreed to the submitted version for publication. Romy Frömer: Contributed to the conception and design, analyzed and interpreted the data, revised the article, and agreed to the submitted version for publication.

5.3 Funding

The project was funded by the Norwegian School of Sport Sciences

Appendix A Description of strategies

This supplementary describes each strategy in more detail. The "stand against" strategy emphasized maintaining a stable stance against external forces without body extension along the body's longitudinal axis or rocking skis forward. This term is frequently used by Norwegian ski coaches when communicating with skiers to help them improve their race times.

The "rock skis forward" strategy involves rocking the ski forward during the turning phase. This action effectively regulates the distribution of pressure over the skis. During the initiation and control phases of a turn, the pressure is generally shifted forward to bend the ski's forebody, increasing friction with the snow and enabling it to turn more sharply. However, at some point, during turn progression, the skier aims to stop turning and therefore shifts the pressure further back on the skis, thereby reducing turning and braking forces [47, 48]. Investigations of elite alpine ski racers

have shown that high performing skiers tend to rock skis more forward and pressure the back part of the ski for considerable longer time during a turn, than slower skiers [49]. To make the information more specific for the skiers, we communicated, that the maximum range of the rocking movement was about 30-50 centimeters from gate passage to completion of the turn, which is in correspondence with biomechanical evidence of elite ski racers [49].

The "extend" strategy involves extending the body from a laterally tilted position during the turn, closer to the turn's center of rotation. When skiers move their center of mass this way by extending their legs, they increase their kinetic rotational energy under certain situations. According to Lind and Sanders [40] model, the skier achieves this effect by shortening the radius of the axis about which they rotate, which will reduce the moment of inertia and consequently increase the rotational kinetic energy of the system under the assumption that angular momentum is conserved. In their model, the gain in rotational kinetic energy from this motion is proportional to the amount of work the skier does against the centrifugal force (from the skier's frame of reference); therefore, a larger extension movement will accomplish a greater increase in rotational kinetic energy. Simulation studies of individuals extending their body in rollers or during carved slalom turns have shown that this movement can increase speed [41, 42], and the effect has been observed in various ways with elite skiers during training and competition [44, 49, 77].

Finally, "extend with rocking skis forward" was expected to be the best strategy combining the two effects from extending and rocking skis forward, and we therefore defined this as the theoretical best strategy. Simulations of skiers extending their bodies in the bottom of a roller have observed an additional effect of rocking the skis forward [41]

Appendix B Snow preparation

We dedicated a substantial amount of time and effort to prepare the hill for our learning experiment. Our primary objective was to ensure that the snow conditions were as identical and fair as possible for all participants. To achieve this, we collaborated closely with the SNØ facility's staff and the coaches of the Norwegian Alpine Ski team. Together, we devised a comprehensive plan to achieve consistency before each group of skiers took to the slopes. This report will detail the steps we took to ensure the snow was ready for our skiers and provide insight into the reasoning behind our choices. The purpose of this supplementary note is to assist you in evaluating our study and to transparently document our commitment to delivering the best possible conditions for our skiers.

Ski group A

About two weeks before testing Group A, new snow was created on the racing hill. The evening before data collection, machines groomed the racing hill (Fig. **??**a:left). We then watered the hill and left it to freeze overnight to create a hard and firm surface (Fig. **??**a:right).

Ski group B

Ski group B began testing the day after ski group A completed their testing. Once group A finished, we inspected the race hill(Fig. **??**b:left), noticed some small holes, and decided to fill them before watering it again.

Ski group C

After ski group B, the race hill started getting icy, lacking grip in some areas. We were concerned that watering it again might worsen conditions, making it too challenging for skiers. Consequently, we opted to gently groom the hill with a machine and let the grooves set for a couple of days. (**The illustration image was not allowed in this preprint and was therefore removed)**.

Ski group D

After Ski Group C, the race hill needed new snow because some areas had become icy, with minimal grip (Fig. reffig:snowprepd:left)). Although the conditions were suitable for Ski Group C, they would not have worked for a new ski group undergoing testing. Consequently, we decided to produce new snow two days before Ski Group D started their training. This fresh snow was pushed into the racing hill the day before testing and groomed. Subsequently, we watered the hill and let it freeze overnight (Fig. **??**d:right).

Appendix C Course setting

We used a standard procedure to set the slalom courses, ensuring a fixed length and offset. First, we stretched a taut rope between two nails on either side of the ski hill. This rope helped us locate the exact starting line consistently from day to day. From the nail on the skier's right, we measured 6 meters into the slope. We did the same from a fixed point approximately 50 meters down the course, but here we measured 3.4 meters out. Then, we pulled a 50-meter-long measuring tape between the two points to establish the line down the hill. We chose a measuring tape over a rope because ropes tend to expand and contract when they get wet and dry, respectively (**the illustration image was not allowed in this preprint and was therefore removed)**.

Once the 50-meter straight line was established, we laid out a rope segment attached to a carabiner hooked onto the long rope. We moved this segment 1.9 meters out from the rope in one turn and 0 meters in the other, with a 10-meter vertical space. This ensured that the course followed a straight line down the slope and was set with the correct offset (see Fig. **??**b for an image illustrating this process). Once all the gates along the 50-meter measurement tape were set, we used a new fixation point down the slope and continued down the course. We practiced this procedure several times before the experiment, and the variation in course setting was a maximum of 10 cm at the end of the course.

Due to wear and tear on the trails, we opted to shift the course laterally or rotate it, depending on the situation. With the exception of one skiing group, we followed

the following practice. On day 1, we set the course as described above. On day 2, we rotated the course so that the first turn went in the opposite direction of day 1. On the third day, we shifted the course closer to the wall on skiers' right (**the illustration image was not allowed in this preprint and was therefore removed)**.

Appendix D Coach setup

We set up three coach stations in the finishing area for the supervised learning groups, one for each coach. The space between the coaches was approximately 3 to 5 meters. We used wall dividers to prevent information leakage between the coaches. In addition, the background noise in the ski hall was generally high, and it was difficult to perceive information without standing close to the person. The supervised (target skill) learning coach was behind the two other coaches to prevent the other coaches from seeing what he did. Each coach had a monitor where they could see their own (but not the other) skiers' race.

(**the illustration image was not allowed in this preprint and was therefore removed)**.

Appendix E Supplementary Table

E.1 Race time

E.2 Strategy choice

E.3 Strategy evaluations and outcomes

Appendix F Supplementary discussion

To convert race time to FIS World Ranking, we assumed that each second corresponded to a 7-point FIS. We could then multiply by 0.12 * 7 to find the difference in FIS points. Then, we used the median rank in our sample and calculated what this difference corresponded to in the World Ranking. We performed this analysis for both females and men. Due to confidentiality, we do not want to say which FIS list we used for this conversion. If the reader thinks 7 points is too small or large, then we welcome the reader to change this number up or down. We only used this conversion to help readers evaluating the effect.

References

[1] Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L. & Haith, A. M. Motor Learning. *Comprehensive Physiology* **9**, 613–663 (2019). URL https://doi.org/10.1002/cphy.c170043.

[2] Wolpert, D. M. & Flanagan, J. R. Motor learning. *Current Biology* **20**, R467–R472 (2010). URL https://www.sciencedirect.com/science/article/pii/S0960982210005129.

[3] Wolpert, D. M., Diedrichsen, J. & Flanagan, J. R. Principles of sensorimotor learning. *Nature Reviews Neuroscience* **12**, 739–751 (2011). URL https://doi.org/10.1038/nrn3112.

[4] Diedrichsen, J. & Kornysheva, K. Motor skill learning between selection and execution. *Trends in Cognitive Sciences* **19**, 227–233 (2015). URL https://www.sciencedirect.com/science/article/pii/S1364661315000273.

[5] Chen, X., Holland, P. & Galea, J. M. The effects of reward and punishment on motor skill learning. *Habits and Skills* **20**, 83–88 (2018). URL https://www.sciencedirect.com/science/article/pii/S2352154617301043.

[6] Stanley, J. & Krakauer, J. W. Motor skill depends on knowledge of facts. *Frontiers in human neuroscience* **7**, 503 (2013). Publisher: Frontiers Media SA.

[7] Du, Y., Krakauer, J. W. & Haith, A. M. The relationship between habits and motor skills in humans. *Trends in Cognitive Sciences* **26**, 371–387 (2022). URL https://www.sciencedirect.com/science/article/pii/S1364661322000389.

[8] Mangalam, M., Yarossi, M., Furmanek, M. P., Krakauer, J. W. & Tunik, E. Investigating and acquiring motor expertise using virtual reality. *Journal of Neurophysiology* **129**, 1482–1491 (2023). URL https://doi.org/10.1152/jn.00088.2023. Publisher: American Physiological Society.

[9] Taylor, J. A. & Ivry, R. B. The role of strategies in motor learning. *Annals of the New York Academy of Sciences* **1251**, 1–12 (2012). URL https://doi.org/10.1111/j.1749-6632.2011.06430.x. Publisher: John Wiley & Sons, Ltd.

[10] Tsay, J. S. *et al.* Strategic processes in sensorimotor learning: reasoning, refinement, and retrieval (2023). URL https://osf.io/preprints/psyarxiv/x4652. Publisher: PsyArXiv.

[11] Thorndike, E. L. *Educational Psychology...* Vol. 2 (Teachers college, Columbia university, 1913).

[12] Gray, W. D. & Lindstedt, J. K. Plateaus, Dips, and Leaps: Where to Look for Inventions and Discoveries During Skilled Performance. *Cognitive Science* **41**, 1838–1870 (2017). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12412. Dasd.

[13] Gray, W. D. Plateaus and asymptotes: spurious and real limits in human performance. *Current Directions in Psychological Science* **26**, 59–67 (2017). Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[14] Ericsson, K. A. The scientific study of expert levels of performance: general implications for optimal learning and creativity. *High Ability Studies* **9**, 75–100 (1998). URL https://doi.org/10.1080/1359813980090106. Publisher: Routledge.

[15] Ericsson, K. A. in *Development of elite performance and deliberate practice* 1 edn, (eds Starkes, J. L. & Ericsson, K. A.) *Expert performance in sports: Advances in research on sport expertise*, Vol. 1 49–83 (Human Kinetics, Champaign, 2003).

[16] Williams, A. M. & Hodges, N. J. Practice, instruction and skill acquisition in soccer: Challenging tradition. *Journal of Sports Sciences* **23**, 637–650 (2005). URL https://doi.org/10.1080/02640410400021328. Publisher: Routledge.

[17] Williams, A. M. & Hodges, N. J. Effective practice and instruction: A skill acquisition framework for excellence. *Journal of Sports Sciences* **41**, 833–849 (2023). URL https://doi.org/10.1080/02640414.2023.2240630. Publisher: Routledge.

[18] Hodges, N. J. & Franks, I. M. Modelling coaching practice: the role of instruction and demonstration. *Journal of sports sciences* **20**, 793–811 (2002). Publisher: Taylor & Francis.

[19] Jordan, M. I. & Rumelhart, D. E. Forward models: Supervised learning with a distal teacher. *Cognitive Science* **16**, 307–354 (1992). URL https://www.sciencedirect.com/science/article/pii/036402139290036T.

[20] Doya, K. Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology* **10**, 732–739 (2000). URL https://www.sciencedirect.com/science/article/pii/S0959438800001537.

[21] Supej, M., Nedergaard, N. J., Nord, J. & Holmberg, H.-C. The impact of start strategy on start performance in alpine skiing exists on flat, but not on steep inclines. *Journal of Sports Sciences* **37**, 647–655 (2019). URL https://doi.org/10.1080/02640414.2018.1522698. Publisher: Routledge.

[22] Cochrum, R. G. *et al.* Visual classification of running economy by distance running coaches. *European Journal of Sport Science* **21**, 1111–1118 (2021). URL https://doi.org/10.1080/17461391.2020.1824020. Publisher: Routledge.

[23] Popp, N. J., Yokoi, A., Gribble, P. L. & Diedrichsen, J. The effect of instruction on motor skill learning. *Journal of Neurophysiology* **124**, 1449–1457 (2020). Publisher: American Physiological Society Bethesda, MD.

[24] Wulf, G. & Weigelt, C. Instructions about Physical Principles in Learning a Complex Motor Skill: To Tell or Not to Tell.... *Research Quarterly for Exercise and Sport* (1997). URL https://www.tandfonline.com/doi/abs/10.1080/02701367.1997.10608018.

[25] Hodges, N. J. & Lee, T. D. The role of augmented information prior to learning a bimanual visual-motor coordination task: Do instructions of the movement pattern facilitate learning relative to discovery learning? *British Journal of Psychology* **90**, 389–403 (1999). URL https://onlinelibrary.wiley.com/doi/abs/10.1348/000712699161486.

[26] Sutton, R. & Barto, A. *Reinforcement Learning, second edition: An Introduction* 2 edn. Adaptive Computation and Machine Learning series (MIT Press, Cambridge, MA, 2018). URL https://books.google.no/books?id=sWV0DwAAQBAJ.

[27] Rescorla, R. & Wagner, A. A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. *Classical Conditioning: Current Research and Theory* (1972).

[28] Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412**, 43–48 (2001). URL https://doi.org/10.1038/35083500.

[29] Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997). URL https://doi.org/10.1126/science.275.5306.1593. Publisher: American Association for the Advancement of Science.

[30] Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**, 1042–1045 (2006). URL https://doi.org/10.1038/nature05051.

[31] Lee, D., Seo, H. & Jung, M. W. Neural Basis of Reinforcement Learning and Decision Making. *Annual Review of Neuroscience* **35**, 287–308 (2012). URL https://doi.org/10.1146/annurev-neuro-062111-150512. Publisher: Annual Reviews.

[32] Law, C.-T. & Gold, J. I. Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature Neuroscience* **12**, 655–663 (2009). URL https://doi.org/10.1038/nn.2304.

[33] Tobler, P. N., O'Doherty, J. P., Dolan, R. J. & Schultz, W. Human neural learning depends on reward prediction errors in the blocking paradigm. *Journal of Neurophysiology* **95**, 301–310 (2006). URL https://doi.org/10.1152/jn.00762.2005. Publisher: American Physiological Society.

[34] Mnih, V. *et al.* Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015). URL https://doi.org/10.1038/nature14236.

[35] Lior Shmuelof *et al.* Overcoming Motor "Forgetting" Through Reinforcement Of Learned Actions. *The Journal of Neuroscience* **32**, 14617 (2012). URL http://www.jneurosci.org/content/32/42/14617.abstract.

[36] Abe, M. *et al.* Reward improves long-term retention of a motor memory through Induction of offline memory gains. *Current Biology* **21**, 557–562 (2011). URL https://www.cell.com/current-biology/abstract/S0960-9822(11)00221-1.

[37] Truong, C. *et al.* Error-based and reinforcement learning in basketball free throw shooting. *Scientific Reports* **13**, 499 (2023). URL https://doi.org/10.1038/s41598-022-26568-2.

[38] Hasson, C. J., Manczurowsky, J. & Yen, S.-C. A reinforcement learning approach to gait training improves retention. *Frontiers in Human Neuroscience* **9** (2015). URL https://www.frontiersin.org/articles/10.3389/fnhum.2015.00459.

[39] Supej, M. & Holmberg, H.-C. A new time measurement method using a high-end global navigation satellite system to analyze alpine skiing. *Research Quarterly for Exercise and Sport* **82**, 400–411 (2011). Publisher: Taylor & Francis.

[40] Lind, D. A. & Sanders, S. P. *The physics of skiing: skiing at the triple point* 2 edn (Springer Science & Business Media, 2004).

[41] Mote, C. & Louie, J. Accelerations induced by body motions during snow skiing. *Journal of Sound and Vibration* **88**, 107–115 (1983). URL https://www.sciencedirect.com/science/article/pii/0022460X8390682X.

[42] Luginbühl, M., Gross, M., Lorenzetti, S., Graf, D. & Bünner, M. J. Identification of Optimal Movement Patterns for Energy Pumping. *Sports* **11** (2023).

[43] Reid, R. C., Haugen, P., Gilgien, M., Kipp, R. W. & Smith, G. A. Alpine ski motion characteristics in slalom. *Frontiers in Sports and Active Living* **2** (2020). URL https://www.frontiersin.org/articles/10.3389/fspor.2020.00025.

[44] Magelssen, C., Haugen, P., Reid, R. & Gilgien, M. Is there a contextual interference effect for sub-elite alpine ski racers learning complex skills? *Frontiers in Bioengineering and Biotechnology* **10** (2022).

[45] Lakens, D. Sample size justification. *Collabra: Psychology* **8**, 33267 (2022). URL https://doi.org/10.1525/collabra.33267.

[46] Minetti, A. E. in *Biomechanics of Alpine Skiing* (eds Schoenhuber, H., Panzeri, A. & Porcelli, S.) *Alpine Skiing Injuries: Prevention and Management* 1–7 (Springer International Publishing, Cham, 2018). URL https://doi.org/10.1007/978-3-319-61355-0_1.

[47] LeMaster, R. *The skier's edge* Jossey-Bass Business & Management (Human Kinetics, Champaign, IL, 1999).

[48] LeMaster, R. *Ultimate Skiing* (Human Kinetics, 2010).

[49] Reid, R. C. *A kinematic and kinetic study of alpine skiing technique in slalom.* PhD Thesis, Norwegian School of Sport Sciences (2010).

[50] Maxwell, S., Delaney, H. & Kelley, K. *Designing Experiments and Analyzing Data: A Model Comparison Perspective* 3 edn (Taylor & Francis, 2017). URL https://books.google.no/books?id=oNcrDwAAQBAJ.

[51] Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).

[52] R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2022). URL https://www.R-project.org/.

[53] Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**, 255–278 (2013). URL https://www.sciencedirect.com/science/article/pii/S0749596X12001180.

[54] Barr, D. J. Learning statistical models through simulation in R: An interactive textbook. Version 1.0.0 (2021). URL https://psyteachr.github.io/stat-models-v1.

[55] Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48 (2015).

[56] Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82**, 1 – 26 (2017). URL https://www.jstatsoft.org/index.php/jss/article/view/v082i13. Section: Articles.

[57] Mize, T. D. Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects. *Sociological Science* **6**, 81–117 (2019). URL http://dx.doi.org/10.15195/v6.a4.

[58] McCabe, C. J., Halvorson, M. A., King, K. M., Cao, X. & Kim, D. S. Interpreting Interaction Effects in Generalized Linear Models of Nonlinear Probabilities and Counts. *Multivariate Behavioral Research* **57**, 243–263 (2022). URL https://doi.org/10.1080/00273171.2020.1868966. Publisher: Routledge.

[59] Lenth, R. V. *emmeans: Estimated marginal means, aka least-squares means* (2023). URL https://CRAN.R-project.org/package=emmeans.

[60] Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* **364**, 56–58 (1993). URL https://doi.org/10.1038/364056a0.

[61] Worthy, D. A. & Todd Maddox, W. A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W.K. Estes. *Special Issue in Honor of William K. Estes* **59**, 41–49 (2014). URL https://www.sciencedirect.com/science/article/pii/S0022249613000874.

[62] Iyer, E. S., Kairiss, M. A., Liu, A., Otto, A. R. & Bagot, R. C. Probing relationships between reinforcement learning and simple behavioral strategies to understand probabilistic reward learning. *Journal of Neuroscience Methods* **341**, 108777 (2020). URL https://www.sciencedirect.com/science/article/pii/S0165027020302004.

[63] Hodges, N. J. & Franks, I. M. Learning a Coordination Skill: Interactive Effects of Instruction and Feedback. *Research Quarterly for Exercise and Sport* **72**, 132–142 (2001). URL https://doi.org/10.1080/02701367.2001.10608943. Publisher: Routledge.

[64] Vereijken, B. & Whiting, H. In defence of discovery learning. *Canadian Journal of Sport Sciences* **15**, 99–106 (1990).

[65] Therrien, A. S., Wolpert, D. M. & Bastian, A. J. Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain* **139**, 101–114 (2016). URL https://doi.org/10.1093/brain/awv329.

[66] Jordan A. Taylor, John W. Krakauer & Richard B. Ivry. Explicit and Implicit Contributions to Learning in a Sensorimotor Adaptation Task. *The Journal of Neuroscience* **34**, 3023 (2014). URL http://www.jneurosci.org/content/34/8/3023.abstract.

[67] Robertson, E. M. Memory instability as a gateway to generalization. *PLOS Biology* **16**, e2004633 (2018). URL https://doi.org/10.1371/journal.pbio.2004633. Publisher: Public Library of Science.

[68] Braun, D. A., Mehring, C. & Wolpert, D. M. Structure learning in action. *Behavioural Brain Research* **206**, 157–165 (2010). URL https://www.sciencedirect.com/science/article/pii/S0166432809005099.

[69] Braun, D. A., Waldert, S., Aertsen, A., Wolpert, D. M. & Mehring, C. Structure learning in a sensorimotor association task. *PLOS ONE* **5**, e8973 (2010). URL https://doi.org/10.1371/journal.pone.0008973. Publisher: Public Library of Science.

[70] Shea, J. B. & Zimny, S. T. in *Context effects in memory and learning movement information* (ed.Magill, R. A.) *Advances in Psychology*, Vol. 12 345–366 (North-Holland, 1983). URL https://www.sciencedirect.com/science/article/pii/S0166411508619986.

[71] Frömer, R., Stürmer, B. & Sommer, W. Come to think of it: Contributions of reasoning abilities and training schedule to skill acquisition in a virtual throwing task. *Acta Psychologica* **170**, 58–65 (2016). URL https://www.sciencedirect.com/science/article/pii/S0001691816301214.

[72] Shea, J. B. & Morgan, R. L. Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental psychology: Human Learning and memory* **5**, 179 (1979). Publisher: American Psychological Association.

[73] Anvari, F. *et al.* Not all effects are indispensable: Psychological science requires verifiable lines of reasoning for whether an effect matters. *Perspectives*

on Psychological Science **18**, 503–507 (2023). URL https://doi.org/10.1177/17456916221091565. Publisher: SAGE Publications Inc.

[74] Taylor, J. A. & Ivry, R. B. Flexible cognitive strategies during motor learning. *PLOS Computational Biology* **7**, e1001096 (2011). URL https://doi.org/10.1371/journal.pcbi.1001096. Publisher: Public Library of Science.

[75] Lohse, K., Miller, M., Bacelar, M. & Krigelson, O. in *Errors, rewards, and reinforcement in motor skill learning* 3 edn, (eds Hodges, N. J. & Williams, A. M.) *Skill acquisition in sport: research, theory and practice* 39–60 (Routledge, New York, 2020).

[76] Yarrow, K., Brown, P. & Krakauer, J. W. Inside the brain of an elite athlete: the neural processes that support high achievement in sports. *Nature Reviews Neuroscience* **10**, 585–596 (2009). URL https://doi.org/10.1038/nrn2672.

[77] Supej, M. Differential specific mechanical energy as a quality parameter in racing alpine skiing. *Journal of Applied Biomechanics* **24**, 121 – 129 (2008). URL https://journals.humankinetics.com/view/journals/jab/24/2/article-p121.xml. Place: Champaign IL, USA Publisher: Human Kinetics, Inc.

Table E1 Race time: Estimated improvement across sessions in acquisition

Term	Estimate	SE	df	CI	t	p
(Intercept)	16.62	0.13	3.93	16.27-16.97	132.69	< 0.001
supervised (free choice)	0.14	0.13	85.83	-0.12-0.39	1.04	0.299
supervised (target skill)	0.14	0.13	85.82	-0.12-0.40	1.10	0.276
reinforcement learning : free choice 1	-0.38	0.03	91.63	-0.45–0.31	-10.82	< 0.001
supervised (free choice) : free choice 1	-0.30	0.03	93.47	-0.37–0.23	-8.96	< 0.001
supervised (target skill) : free choice 1	-0.50	0.03	91.95	-0.57–0.44	-15.08	< 0.001
reinforcement learning : free choice 2	-0.45	0.05	95.16	-0.54–0.36	-9.99	< 0.001
supervised (free choice) : free choice 2	-0.31	0.04	96.39	-0.39–0.22	-7.17	< 0.001
supervised (target skill) : free choice 2	-0.43	0.04	96.20	-0.52–0.35	-10.07	< 0.001
sd(Intercept)	0.51					
cor(Intercept,free choice 1)	-0.03					
cor(Intercept,free choice 2)	0.23					
sd(free choice 1)	0.15					
cor(free choice 1, free choice 2)	0.73					
sd(free choice 2)	0.22					
sd(Intercept)	0.23					
sd(Observation)	0.21					

Formula: racetime ∼ treatment / session + (1 |skigroup) + (1 + session | skigroup:skier)

Table E2 Race time: Estimated difference between groups during acquisition

Term	Estimate	SE	df	CI	t	p
(Intercept)	16.62	0.13	3.93	16.27-16.97	132.68	< 0.001
free choice 1	-0.39	0.02	92.32	-0.43–0.35	-20.09	< 0.001
free choice 2	-0.40	0.03	95.89	-0.45–0.35	-15.74	< 0.001
forced exploration : supervised (free choice)	0.06	0.13	92.73	-0.20-0.32	0.48	0.631
free choice 1 : supervised (free choice)	0.14	0.13	86.15	-0.11-0.39	1.11	0.272
free choice 2 : supervised (free choice)	0.20	0.14	78.00	-0.08-0.49	1.43	0.157
forced exploration : supervised (target skill)	0.18	0.13	92.66	-0.08-0.44	1.37	0.174
free choice 1 : supervised (target skill)	0.05	0.13	86.12	-0.20-0.31	0.43	0.672
free choice 2 : supervised (target skill)	0.19	0.14	78.02	-0.09-0.48	1.37	0.176
sd(Intercept)	0.51					
cor(Intercept, free choice 1)	-0.03					
cor(Intercept, free choice 2)	0.23					
sd(free choice 1)	0.15					
cor(free choice 1, free choice 2)	0.73					
sd(free choice 2)	0.22					
sd(Intercept)	0.23					
sd(Observation)	0.21					

Formula: racetime ∼ session / treatment + (1 |skigroup) + (1 + session | skigroup:skier)

Table E3 Strategy choice: Probability change across sessions

Contrast	Treatment	Estimate	SE	df	CI	z	p
free choice 2 - free choice 1	rl	0.02	0.06	Inf	-0.11-0.14	0.26	0.791
retention - free choice 1	rl	0.07	0.07	Inf	-0.07-0.21	0.96	0.339
retention - free choice 2	rl	0.05	0.07	Inf	-0.09-0.19	0.72	0.469
transfer - free choice 1	rl	0.18	0.07	Inf	0.04-0.32	2.58	0.010
transfer - free choice 2	rl	0.16	0.07	Inf	0.03-0.30	2.34	0.019
transfer - retention	rl	0.11	0.08	Inf	-0.04-0.27	1.42	0.155
free choice 2 - free choice 1	slfc	0.08	0.06	Inf	-0.04-0.21	1.32	0.188
retention - free choice 1	slfc	0.31	0.07	Inf	0.18-0.44	4.59	< 0.001
retention - free choice 2	slfc	0.23	0.07	Inf	0.09-0.36	3.31	< 0.001
transfer - free choice 1	slfc	0.36	0.07	Inf	0.23-0.50	5.33	< 0.001
transfer - free choice 2	slfc	0.28	0.07	Inf	0.14-0.42	4.02	< 0.001
transfer - retention	slfc	0.05	0.06	Inf	-0.07-0.18	0.85	0.396

Formula: glmer(chosetheorybest $\sim$ treatment * session + (1 | skigroup/skier).

rl = reinforcement learning; slfc = supervised (free choice) learning

Table E4 Predicted probability difference between groups

Contrast	Session	Estimate	SE	df	CI	z	p
rl - slfc	free choice 1	0.01	0.13	Inf	-0.24-0.27	0.11	0.911
rl - slfc	free choice 2	-0.05	0.13	Inf	-0.31-0.20	-0.39	0.693
rl - slfc	retention	-0.23	0.13	Inf	-0.48-0.03	-1.75	0.079
rl - slfc	transfer	-0.17	0.12	Inf	-0.40-0.07	-1.41	0.159

Formula: glmer(chosetheorybest $\sim$ treatment * session + (1 | skigroup/skier).

rl = reinforcement learning; slfc = supervised (free choice) learning.

Table E5 Strategy choice: Predicted probability change in probability of choosing the individual skier's estimated best strategy

Contrast	Treatment	Estimate	SE	df	CI	z	p
free choice 2 - free choice 1	rl	0.20	0.06	Inf	0.09-0.32	3.39	< 0.001
retention - free choice 1	rl	0.13	0.06	Inf	0.00-0.26	2.02	0.043
retention - free choice 2	rl	-0.07	0.06	Inf	-0.19-0.04	-1.25	0.213
transfer - free choice 1	rl	0.21	0.06	Inf	0.09-0.34	3.30	< 0.001
transfer - free choice 2	rl	0.01	0.05	Inf	-0.09-0.11	0.18	0.854
transfer - retention	rl	0.08	0.06	Inf	-0.04-0.20	1.31	0.190
free choice 2 - free choice 1	slfc	0.09	0.06	Inf	-0.03-0.21	1.48	0.140
retention - free choice 1	slfc	0.21	0.07	Inf	0.08-0.34	3.11	0.002
retention - free choice 2	slfc	0.12	0.06	Inf	-0.00-0.24	1.88	0.060
transfer - free choice 1	slfc	0.23	0.07	Inf	0.10-0.36	3.39	< 0.001
transfer - free choice 2	slfc	0.14	0.06	Inf	0.02-0.26	2.21	0.027
transfer - retention	slfc	0.02	0.06	Inf	-0.09-0.13	0.35	0.727

Formula: glmer(choseestimatedbest treatment * session + (1 | skigroup/skier).

rl = reinforcement learning; slfc = supervised (free choice) learning

Table E6 Strategy choice: Predicted probability difference of choosing the individual skier's estimated best strategy at each session

Contrast	Session	Estimate	SE	df	CI	z	p
rl - slfc	block2	0.01	0.12	Inf	-0.22-0.24	0.12	0.904
rl - slfc	block3	0.13	0.10	Inf	-0.06-0.32	1.30	0.194
rl - slfc	retention	-0.06	0.10	Inf	-0.26-0.14	-0.61	0.544
rl - slfc	transfer	0.00	0.09	Inf	-0.17-0.17	0.00	0.999

Formula: glmer(choseestimatedbest $\sim$ treatment * session + (1 | skigroup/skier).

rl = reinforcement learning; slfc = supervised (free choice) learning

Table E7 Strategy evaluation: Estimated differences between strategy ranking during forced exploration

Term	Estimate	SE	CI	t	p
(Intercept)	2.50	0.03	2.44-2.56	85.64	< 0.001
slfc	0.00	0.08	-0.16-0.16	0.00	1.000
slts	0.00	0.05	-0.11-0.11	0.00	1.000
ranktime	0.00	0.01	-0.02-0.02	0.00	1.000
rl : b	-1.72	0.11	-1.94–1.50	-15.47	< 0.001
slfc : b	-1.43	0.19	-1.81–1.05	-7.38	< 0.001
slts : b	-1.20	0.11	-1.41–0.99	-11.32	< 0.001
rl : c	-1.05	0.11	-1.27–0.83	-9.42	< 0.001
slfc : c	-0.70	0.19	-1.08–0.32	-3.60	< 0.001
slts : c	-0.95	0.11	-1.16–0.74	-8.94	< 0.001
rl : d	-2.06	0.11	-2.28–1.84	-18.49	< 0.001
slfc : d	-2.09	0.19	-2.47–1.71	-10.77	< 0.001
slts : d	-2.41	0.11	-2.61–2.20	-22.67	< 0.001
slfc : ranktime	-0.00	0.02	-0.04-0.04	-0.00	1.000
slts : ranktime	-0.00	0.02	-0.04-0.04	-0.00	1.000
rl : b : ranktime	-0.15	0.04	-0.22–0.07	-3.95	< 0.001
slfc : b : ranktime	-0.12	0.05	-0.22–0.02	-2.45	0.014
slts : b : ranktime	-0.13	0.04	-0.20–0.06	-3.65	< 0.001
rl : c : ranktime	0.05	0.04	-0.03-0.12	1.21	0.225
slfc : c : ranktime	-0.01	0.05	-0.11-0.09	-0.27	0.791
slts : c : ranktime	-0.06	0.04	-0.13-0.01	-1.78	0.076
rl : d : ranktime	-0.09	0.04	-0.17–0.02	-2.51	0.012
slfc : d : ranktime	-0.09	0.05	-0.19-0.01	-1.85	0.065
slts : d : ranktime	-0.15	0.04	-0.22–0.08	-4.18	< 0.001

Formula: lm(value $\sim$ treatment/strategy * ranktime)

rl = reinforcement learning; slfc = supervised (free choice) learning, slts = supervised (target skill) learning, b="extend", c="rock skis forward", d="extend with rock skis forward"

Table E8 Strategy evaluation: Estimated differences between groups on strategy ranking during forced exploration

Term	Estimate	SE	CI	t	p
(Intercept)	2.50	0.03	2.44-2.56	85.64	< 0.001
b	-1.45	0.08	-1.61–1.29	-17.58	< 0.001
c	-0.90	0.08	-1.06–0.74	-10.88	< 0.001
d	-2.18	0.08	-2.35–2.02	-26.46	< 0.001
ranktime	-0.00	0.01	-0.02-0.02	-0.00	1.000
a : slfc	-0.15	0.16	-0.46-0.16	-0.96	0.340
b : slfc	0.14	0.16	-0.17-0.45	0.87	0.385
c : slfc	0.20	0.16	-0.11-0.51	1.26	0.209
d : slfc	-0.19	0.16	-0.50-0.13	-1.17	0.242
a : slts	-0.07	0.11	-0.28-0.15	-0.63	0.531
b : slts	0.45	0.11	0.24-0.67	4.17	< 0.001
c : slts	0.03	0.11	-0.18-0.24	0.29	0.772
d : slts	-0.42	0.11	-0.63–0.20	-3.83	< 0.001
b : ranktime	-0.13	0.02	-0.18–0.09	-5.56	< 0.001
c : ranktime	-0.01	0.02	-0.06-0.04	-0.43	0.665
d : ranktime	-0.11	0.02	-0.16–0.06	-4.65	< 0.001
a : slfc : ranktime	0.01	0.04	-0.08-0.10	0.19	0.847
b : slfc : ranktime	0.03	0.04	-0.05-0.12	0.73	0.465
c : slfc : ranktime	-0.05	0.04	-0.14-0.04	-1.13	0.259
d : slfc : ranktime	0.01	0.04	-0.08-0.10	0.21	0.837
a : slts : ranktime	0.04	0.04	-0.03-0.11	1.00	0.317
b : slts : ranktime	0.05	0.04	-0.02-0.13	1.48	0.140
c : slts : ranktime	-0.07	0.04	-0.14–0.00	-1.97	0.049
d : slts : ranktime	-0.02	0.04	-0.09-0.05	-0.50	0.615

Formula: lm(value $\sim$ strategy / treatment * ranktime)

rl = reinforcement learning; slfc = supervised (free choice) learning, slts = supervised (target skill) learning, a="stand against", "b="extend", c="rock skis forward", d="extend with rock skis forward"

Table E9 Strategy evaluation: Estimated slope for each strategy across session

Term	Estimate	SE	CI	t	p
(Intercept)	2.50	0.03	2.44-2.56	85.64	< 0.001
slfc	0.00	0.08	-0.16-0.16	0.00	1.000
slts	0.00	0.05	-0.11-0.11	0.00	1.000
b	-1.45	0.08	-1.61–1.29	-17.58	< 0.001
c	-0.90	0.08	-1.06–0.74	-10.88	< 0.001
d	-2.18	0.08	-2.35–2.02	-26.46	< 0.001
slfc : b	0.29	0.22	-0.15-0.73	1.29	0.197
slts : b	0.52	0.15	0.22-0.82	3.39	< 0.001
slfc : c	0.35	0.22	-0.09-0.79	1.56	0.118
slts : c	0.10	0.15	-0.20-0.40	0.65	0.517
slfc : d	-0.03	0.22	-0.47-0.41	-0.15	0.879
slts : d	-0.35	0.15	-0.65–0.05	-2.26	0.024
rl : a : ranktime	0.05	0.03	-0.00-0.10	1.86	0.064
slfc : a : ranktime	0.06	0.04	-0.01-0.13	1.61	0.107
slts : a : ranktime	0.09	0.03	0.04-0.13	3.40	< 0.001
rl : b : ranktime	-0.10	0.03	-0.15–0.05	-3.73	< 0.001
slfc : b : ranktime	-0.07	0.04	-0.14-0.00	-1.85	0.065
slts : b : ranktime	-0.04	0.03	-0.09-0.00	-1.77	0.077
rl : c : ranktime	0.09	0.03	0.04-0.15	3.57	< 0.001
slfc : c : ranktime	0.04	0.04	-0.03-0.11	1.24	0.216
slts : c : ranktime	0.02	0.03	-0.03-0.07	0.89	0.376
rl : d : ranktime	-0.04	0.03	-0.10-0.01	-1.70	0.090
slfc : d : ranktime	-0.04	0.04	-0.11-0.03	-1.00	0.317
slts : d : ranktime	-0.06	0.03	-0.11–0.01	-2.51	0.012

Formula: lm(value ∼ treatment * strategy / ranktime)

rl = reinforcement learning; slfc = supervised (free choice) learning, slts = supervised (target skill) learning, a="stand against", b="extend", c="rock skis forward", d="extend with rock skis forward"

Table E10 Strategy outcome: Estimated difference between strategies during forced exploration

Term	Estimate	SE	df	CI	t	p
(Intercept)	16.88	0.11	3.99	16.56-17.20	147.94	< 0.001
supervised (free choice)	0.06	0.13	91.87	-0.21-0.32	0.42	0.674
supervised (target skill)	0.12	0.13	91.72	-0.14-0.37	0.90	0.370
reinforcement learning : c-a	-0.22	0.03	641.78	-0.29–0.15	-6.39	< 0.001
supervised (free choice) : c-a	-0.16	0.03	641.98	-0.23–0.10	-4.80	< 0.001
supervised (target skill) : c-a	-0.22	0.03	641.72	-0.28–0.15	-6.53	< 0.001
reinforcement learning : b-c	-0.22	0.03	641.75	-0.29–0.16	-6.47	< 0.001
supervised (free choice) : b-c	-0.15	0.03	641.93	-0.22–0.09	-4.51	< 0.001
supervised (target skill) : b-c	-0.23	0.03	641.75	-0.30–0.17	-7.00	< 0.001
reinforcement learning : d-b	0.00	0.03	641.72	-0.07-0.07	0.02	0.984
supervised (free choice) : d-b	-0.05	0.03	641.97	-0.12-0.01	-1.53	0.125
supervised (target skill) : d-b	-0.01	0.03	641.80	-0.07-0.06	-0.21	0.834
sd(Intercept)	0.52					
sd(Intercept)	0.20					
sd(Observation)	0.19					

Formula: lmer(racingtime ∼ treatment / strategy + (1 — skigroup / skier)

a="stand against", b="extend", c="rock skis forward", d="extend with rock skis forward"

For this model we used a sliding difference contrast coding scheme

Table E11 Strategy outcome: Estimated difference between groups per strategy

Term	Estimate	SE	df	CI	t	p
(Intercept)	16.88	0.11	3.99	16.56-17.20	147.94	< 0.001
b	-0.40	0.02	641.79	-0.44–0.37	-20.63	< 0.001
c	-0.20	0.02	641.83	-0.24–0.16	-10.22	< 0.001
d	-0.42	0.02	641.78	-0.46–0.38	-21.72	< 0.001
a : supervised (free choice)	-0.01	0.13	101.33	-0.28-0.26	-0.07	0.945
b : supervised (free choice)	0.12	0.13	101.55	-0.15-0.39	0.87	0.386
c : supervised (free choice)	0.05	0.14	101.61	-0.22-0.32	0.36	0.716
d : supervised (free choice)	0.06	0.13	101.04	-0.20-0.33	0.48	0.632
a : supervised (target skill)	0.17	0.13	101.01	-0.09-0.44	1.28	0.203
b : supervised (target skill)	0.17	0.13	100.98	-0.10-0.44	1.26	0.209
c : supervised (target skill)	0.18	0.13	101.24	-0.09-0.45	1.33	0.188
d : supervised (target skill)	0.16	0.13	101.33	-0.11-0.43	1.21	0.231
sd(Intercept)	0.52					
sd(Intercept)	0.20					
sd(Observation)	0.19					

Formula: lmer(racingtime ∼ strategy / treatment + (1 | skigroup / skier)

a="stand against", b="extend", c="rock skis forward", d="extend with rock skis forward"

Table E12 Strategy outcome: Estimated change for each strategy

Term	Estimate	SE	df	CI	t	p
(Intercept)	16.80	0.11	4.05	16.48-17.12	146.23	< 0.001
slfc	0.06	0.13	60.23	-0.21-0.32	0.44	0.661
b	-0.44	0.02	1378.01	-0.48–0.40	-20.15	< 0.001
c	-0.21	0.02	1378.03	-0.25–0.16	-8.82	< 0.001
d	-0.45	0.02	1378.05	-0.50–0.41	-21.04	< 0.001
slfc : b	0.17	0.04	1378.01	0.09-0.26	3.98	< 0.001
slfc : c	0.07	0.05	1378.03	-0.02-0.16	1.47	0.142
slfc : d	0.08	0.04	1378.05	-0.00-0.17	1.95	0.051
rl : a : session number	0.01	0.05	1378.53	-0.09-0.10	0.18	0.860
slfc : a : session number	-0.07	0.03	1378.40	-0.13–0.01	-2.36	0.018
rl : b : session number	-0.11	0.01	1378.80	-0.13–0.09	-9.80	< 0.001
slfc : b : session number	-0.07	0.01	1379.01	-0.10–0.05	-5.89	< 0.001
rl : c : session number	-0.12	0.03	1380.31	-0.17–0.07	-4.58	< 0.001
slfc : c : session number	-0.13	0.02	1378.76	-0.17–0.09	-6.13	< 0.001
rl : d : session number	-0.11	0.01	1378.82	-0.13–0.08	-9.49	< 0.001
slfc : d : session number	-0.10	0.01	1378.28	-0.12–0.08	-10.59	< 0.001
sd(Intercept)	0.52					
sd(Intercept)	0.19					
sd(Observation)	0.19					

Formula: lmer(racetime ∼ treatment * strategy / sessionnumber + (1 | skigroup / skier))

rl = reinforcement learning; slfc = supervised (free choice) learning, a="stand against", b="extend", c="rock skis forward", d="extend with rock skis forward"

Table E13 Strategy outcome: Estimated difference in change between groups for each strategy

Term	Estimate	SE	df	CI	t	p
(Intercept)	16.80	0.11	4.05	16.48-17.12	146.23	< 0.001
b	-0.44	0.02	1378.01	-0.48–0.40	-20.15	< 0.001
c	-0.21	0.02	1378.03	-0.25–0.16	-8.82	< 0.001
d	-0.45	0.02	1378.05	-0.50–0.41	-21.04	< 0.001
session number	-0.09	0.01	1378.67	-0.10–0.07	-9.99	< 0.001
a : slfc	-0.02	0.14	66.50	-0.30-0.25	-0.17	0.866
b : slfc	0.15	0.14	64.08	-0.12-0.42	1.12	0.268
c : slfc	0.05	0.14	65.91	-0.23-0.32	0.33	0.740
d : slfc	0.06	0.13	63.63	-0.21-0.33	0.45	0.653
b : session number	-0.06	0.03	1378.48	-0.12–0.00	-1.98	0.047
c : session number	-0.09	0.03	1378.72	-0.16–0.03	-2.78	0.005
d : session number	-0.07	0.03	1378.41	-0.13–0.01	-2.44	0.015
a : slfc : session number	-0.08	0.06	1378.54	-0.19-0.03	-1.40	0.163
b : slfc : session number	0.04	0.02	1378.88	0.01-0.07	2.34	0.020
c : slfc : session number	-0.01	0.03	1379.85	-0.07-0.06	-0.23	0.819
d : slfc : session number	0.00	0.01	1378.56	-0.03-0.03	0.20	0.842
sd(Intercept)	0.52					
sd(Intercept)	0.19					
sd(Observation)	0.19					

Formula: lmer(racetime ∼ strategy / treatment * sessionnumber + (1 | skigroup / skier))

rl = reinforcement learning; slfc = supervised (free choice) learning, a="stand against", b="extend", c="rock skis forward", d="extend with rock skis forward"

Table E14 Strategy outcome: Estimated change on "extend with rock skis forward"

Term	Estimate	SE	df	CI	t	p
(Intercept)	16.59	0.16	6.81	16.20-16.98	101.76	< 0.001
supervised (free choice)	0.08	0.13	95.97	-0.18-0.35	0.64	0.522
supervised (target skill)	0.06	0.13	94.98	-0.20-0.32	0.44	0.660
reinforcement learning : session number	-0.09	0.01	1065.01	-0.11–0.07	-8.15	< 0.001
supervised (free choice) : session number	-0.10	0.01	1063.28	-0.11–0.08	-10.01	< 0.001
supervised (target skill) : session number	-0.08	0.01	1058.03	-0.09–0.06	-9.25	< 0.001
sd(Intercept)	0.51					
sd(Intercept)	0.27					
sd(Observation)	0.18					

Formula: lmer(racetime ∼ treatment / sessionnumber + (1 | skigroup / skier))

rl = reinforcement learning; slfc = supervised (free choice) learning, slts = supervised (target skill) learning, a="stand against", b="extend", c="rock skis forward", d="extend with rock skis forward"