

**Cognitive function in simulated paragliding flight.***Authors*

Dr. Matt Wilkes MSc FRCA

Mr. Geoff Long MPhil

Dr. Heather Massey PhD

Dr. Clare Eglin PhD

Professor Michael J. Tipton PhD

*Department, institution and affiliation*

Extreme Environments Laboratory, University of Portsmouth, Portsmouth, UK

*Short title (30 char)*

Cognition in paragliding

*Corresponding author (reprints)*

Matt Wilkes, Extreme Environments Laboratory, University of Portsmouth, Portsmouth, UK

*Metrics*

Word count (abstract): 250

Word count (main text): 4768

Number of references: 38

Number of tables: 4

Number of figures: 3

## Abstract

**Introduction:** Paragliding is an emerging discipline of aviation, with recreational pilots flying distances over 100 km. It remains risky. Accidents typically relate to pilot error rather than equipment failure. We measured cognition and physiological responses during simulated flight, to investigate whether errors might be due to pilot impairment, rather than misjudgment.

**Methods:** Ten male paraglider pilots (aged 19-58 years) undertook a simulated flight in an environmental chamber from sea level (0.209 FiO<sub>2</sub>) to 1524 m (0.174 FiO<sub>2</sub>), 2438 m (0.156 FiO<sub>2</sub>), and 3658 m (0.133 FiO<sub>2</sub>), over approximately two hours. They experienced normobaric hypoxia, environmental cooling and headwind, completing logical reasoning, mannikin, mathematical processing, Stroop Color-Word and Tower Puzzle tasks; alongside measures of risk-taking (BART), mood (POMS) and subjective experience. Results were compared to ten controls, matched by age, sex and flying experience. Physiological measures were oxygen consumption, carbon dioxide production, ventilation, heart rate, oxygen saturation, rectal and skin temperature, blood glucose, blood lactate and urine production.

**Results:** There were no significant differences between pilots and controls at any altitude. Results were heterogenous within and between individuals. As altitude increased, oxygen consumption and minute volume increased significantly, while oxygen saturations fell (98.3% [baseline] to 88.5% [peak]). Rectal temperatures fell by a statistically (but not clinically) significant amount (37.6°C to 37.3°C), while finger skin temperatures dropped steeply (32.2°C to 13.9°C).

**Conclusions:** Results suggest cognitive impairment is unlikely to be a primary cause of pilot error during paragliding flights (of less than two hours, below 3658 m), though hand protection requires improvement.

Keywords: Hypoxia, cold, performance, low altitude, temperature

## Introduction

Paragliding is an emerging discipline of aviation, with at least 127,000 paraglider pilots flying worldwide.<sup>25</sup> Rapid advances in wing and harness design have allowed recreational pilots to regularly fly distances of over 100 km. The open distance record presently stands at 568 km in a single, unpowered, 11-hour flight and the altitude record is 8,157 m (26,800 feet) without supplementary oxygen.<sup>12</sup> Paragliding, once a descent technique for mountaineers, has now become aviation.

Despite these leaps in performance, paragliding remains risky, though the risks are difficult to quantify. Hazards are often site-specific, and incident databases suffer from underreporting and a lack of denominators: we do not always know how many flights are taking place, by whom, and under what circumstances. However, analyses of past incidents indicated that accidents were rarely due to equipment failure (though poor equipment design and incorrect use may have been factors). Rather, a retrospective analysis of paragliding incidents in the U.K. between 2011-2015 identified omission of pre-flight checks (10%), poor glider control (44%), poor awareness of other aircraft (11%) and misjudgments of conditions in flight (19%) as the primary precursors to accidents.<sup>6</sup> The review pointed to failure of the operators, rather than their equipment, as the key cause for concern.

It may be that paraglider pilots lack sufficient training, or that the safety culture surrounding paragliding is inadequate given the inherent risks of the activity. However, such conclusions rely on the assumption the operators are not physically or cognitively impaired, situations described in the Human Factors Analysis and Classification System (HFACS) as 'physiological or mental states or limitations [that might be] preconditions for error'.<sup>33</sup>

In a 2017 study, we measured cardiorespiratory parameters and oxygen consumption of paraglider pilots in flight.<sup>38</sup> Paragliding proved to be minimally demanding

(approximately 1.7 Metabolic Equivalents) except at take-off, the extremes of altitude, or during maneuvers that generated high acceleration forces.<sup>38</sup> We did not record any physiological states or limitations likely to have caused operator impairment in cross-country flight.

McMorris et al. concluded in a 2017 meta-analysis that an abrupt reduction in inspired oxygen sufficient to lower PaO<sub>2</sub> below 60 mmHg impaired both central executive and non-executive cognitive performance.<sup>21</sup> The deterioration seen may have resulted directly from tissue hypoxia or hypocapnic cerebral hypoperfusion secondary to the hypoxic ventilatory response.<sup>27</sup> However, paragliders typically fly at lower altitudes with higher inspired oxygen fractions than those in studies reviewed by McMorris. The effects of gradual-onset hypoxia caused by extended periods at these lower altitudes remain contentious. A number of studies have demonstrated changes in complex reaction times<sup>9</sup>, mood<sup>17, 26</sup>, subjective symptoms<sup>23, 30, 35</sup>, risk behavior<sup>29</sup> and complex cognition<sup>5, 15, 16</sup> at altitudes below 3658 m (12,000 ft). However, small sample sizes and interindividual variability, in combination with the heterogeneity, practice effects and uncertain operational relevance of cognitive testing have often hampered their interpretation.

These difficulties are compounded when considering the paragliding flight environment, as hypoxia is not the only stressor. In paragliding, pilots face a cocktail of environmental cooling, headwind and hypoxia, all with the potential to affect performance.<sup>36</sup> Furthermore, unlike in general aviation and in the studies listed above, ascent rates are also relatively slow (typically 0.5-3 m·s<sup>-1</sup>, 100-600 feet·min<sup>-1</sup>) and paragliders fly a linked series of climbs and descents, rather than simply climbing to a cruising altitude and then descending to a landing.

To test the hypothesis that the paragliding flight environment might, in HFACS terminology, 'lead to mental states or limitations that might be preconditions for error', we simulated the hypoxia, cold and headwind of a paragliding flight. Pilots, suspended in their paragliding harnesses, ascended from sea level (0.209 FiO<sub>2</sub>) in a ramped altitude profile to 1524 m (5000 ft., 0.174 FiO<sub>2</sub>), 2438 m (8000 ft., 0.156 FiO<sub>2</sub>) and 3658 m (12,000 ft., 0.133 FiO<sub>2</sub>) over the time-course of a typical flight. At each altitude, they undertook validated tests of cognitive function, mood and risk-taking behavior, with their results compared to a group of matched controls.

We did not aim to draw generalizable conclusions about the role of each separate environmental stressor in cognitive performance. Instead, we took a pragmatic and context-specific approach: to simply investigate if the combination of stressors typical of paragliding flight would lead to a deterioration sufficient to explain past incidents, be a target for further investigation or for risk mitigation. If we did not find such a deterioration then, while not dismissing cognitive impairment as a factor in complex decision-making, we could shift the focus of our safety efforts to other (potentially more impactful) parts of the HFACS taxonomy.

## **Methods**

### *Subjects*

The simulator group consisted of ten male paraglider pilots recruited by word of mouth. We screened potential subjects with an exercise and health history questionnaire, 12-lead electrocardiogram and baseline hemoglobin. Exclusion criteria were: sojourns to altitudes above 3,500 m of longer than 48 hours within the last three months (excluding commercial air travel); cigarette smoking; anemia or polycythemia; blood donation or lower respiratory

tract infection within the preceding six weeks or symptomatic upper respiratory tract infection; blood dyscrasias or suspected pregnancy. Two female pilots were consented but excluded prior to participation (active pregnancy and intercurrent illness, respectively) and one male pilot was excluded following an incidental diagnosis of Wolff-Parkinson-White syndrome. Three male pilots were subsequently recruited in their place, making a total of ten. They were subsequently matched to a group of controls for cognitive testing, by age (intervention group 42.7 (13.2) years vs. controls 46.9 (12.2) years); years of paragliding experience (10.8 (8.6) years vs. 12.5 (10.8) years), hours paragliding (859 (684) hours vs. 692 (897) hours), and maximum altitude flown in a paraglider (3481 (1305) m vs. 3585 (1217) m).

The study protocol was approved in advance by the University of Portsmouth Science Faculty Ethics Committee (SFEC 2018-006). Each subject provided written informed consent and the study conformed to the Declaration of Helsinki.

### *Equipment and materials*

Our experiments took place in the Fiennes Chamber of the Extreme Environments Laboratory (University of Portsmouth). We simulated altitude using whole-chamber reduction in oxygen fraction (normobaric hypoxia), environmental cooling through the introduction of cooled air, and headwind with an industrial box fan (Man Cooler, Colt International, Petersfield UK).

Equivalent inspired oxygen fractions ( $FiO_2$ ) were 0.209 for baseline, 0.174 for 1524 m, 0.156 for 2438 m and 0.133 for 3658 m. The ascent rate was  $2 \text{ m}\cdot\text{s}^{-1}$ . Establishing equivalent environmental temperatures was more complex. We first estimated temperature at each altitude according to the International Standard Atmosphere (ISA)

lapse rate.<sup>1</sup> However, we were aware that if this were real flight, pilots would experience a balance of environmental cooling and heating from solar radiation, which would vary as they climbed up under clouds and then glided out into direct sunlight. Consequently, we elected to cool the chamber down to the ISA estimate during each simulated climb, then turned off the cooling system, allowing the chamber to naturally warm up during the 'glide' phase of flight, until the next ascent. This approach felt most authentic to pilots during practice runs in the chamber.

To calculate simulated headwind, we began with the equivalent airspeed of a commercially-available midrange paraglider, approximately  $10 \text{ m}\cdot\text{s}^{-1}$ .<sup>22</sup> We then increased it according to true air speed at each altitude and decreased it according to the density ratio (assuming constant humidity). The target fan speeds were therefore  $10 \text{ m}\cdot\text{s}^{-1}$  (sea level),  $9.3 \text{ m}\cdot\text{s}^{-1}$  (1524 m),  $8.9 \text{ m}\cdot\text{s}^{-1}$  (2438 m) and  $8.3 \text{ m}\cdot\text{s}^{-1}$  (3658 m); average speed  $9.1 \text{ m}\cdot\text{s}^{-1}$ .

Subjects were suspended in appropriately-sized, semi-recumbent Supair Strike cocoon harnesses (Supair VLD, Annecy, France), holding paragliding brake handles (Gin Gliders, Yongin, South Korea), tensioned to provide a similar 'weighted' feel to real brake lines (Figure 1). They were permitted their own choice of base and mid-layers, but all wore appropriately-sized Rab Neutrino Endurance 850 fill-power down jackets (Rab Equipment Ltd, Ripley, UK) as outer layers, Guide Gloves (Mountain Equipment Ltd, Hyde, UK), Bamboo Balaclava (Bamboo Clothing Ltd, Plymouth, UK) and Supair Pilot Helmets (Supair VLD, Annecy, France). To maintain attention, and to simulate the exercise demands of flight, the subjects mimicked the brake line inputs and weight-shift turns made by a pilot in a chase-camera video projected on the wall in front of them (when not undertaking cognitive tests).

[Fig. #1 here]



We recorded  $F_{IO_2}$  and  $F_{ICO_2}$  fractions with a Rapidox 3100 Multigas Analyzer (Cambridge Sensotec Ltd., St Ives, Cambridge) calibrated to the sea level environment and a reference gas (10%  $O_2$ , 5%  $CO_2$ ); all temperatures using thermistor probes (environmental temperature, skin temperature of forehead, little finger and great toe, and rectal temperature), logged via Squirrel 2040 data logger (Grant Instruments, Shepreth, UK). We measured physiological variables ( $V_E$ ,  $VO_2$ ,  $VCO_2$  and heart rate) using a Metamax 3b system (CORTEX Biophysik GmbH, Leipzig, Germany) and peripheral oxygen saturation ( $SpO_2$ ) with a Nonin 7500 table-top pulse oximeter with ear lobe sensor (Nonin Medical, Plymouth, USA). We measured blood glucose and lactate concentrations with a Biosen C-line Sport (EKF Diagnostics, Cardiff, UK) and urine osmolality with an Osmocheck handheld digital refractometer (Vitech Scientific, Horsham, UK).

We chose the ANAM 4 battery (Cognitive Science Research Center, University of Oklahoma, OK), previously assessed as having good construct validity for its specified cognitive domains<sup>34</sup> and for its mood scale<sup>14</sup>, as well as minimal practice effects following second administration<sup>4,11</sup>. Given the variety of cognitive tests used in past studies, we deliberately chose our altitudes and test battery to be comparable with those used in two of the most rigorous previous studies: Legg et al.<sup>17</sup> and Pilmanis et al.<sup>30</sup>

We selected the individual tests from the battery based on our experience of the demands of paragliding: logical reasoning (abstract reasoning), mannikin variation (spatial awareness and attention), mathematical processing (computation, concentration, working memory), Stroop Color-Word test (selective attention, interference and executive function) and Tower Puzzle (visual spatial ability, motor control, planning). The primary outcome metric for logical reasoning, mannikin variation and mathematical processing was throughput (number of correct responses per unit of available response time); for Stroop it

was interscore (number of correct responses on block 3 minus the Predicted Color-Word Score) and for Tower it was the move ratio (ratio of actual moves to minimum number of moves required to solve the puzzle).

The subjects also undertook a Profile of Mood States (POMS)<sup>24</sup> test within the ANAM battery, where they were asked to describe how much a mood adjective applied on a categorical scale of 0 ('not at all') to 6 ('very much'), from the subcategories of anger, anxiety, depression/dysphoria, fatigue, happiness, restlessness and vigor.

Finally, we measured risk-taking behavior using a Balloon Analogue Risk Task (Automatic BART, Inquisit 5 environment, Millisecond Software, Seattle, USA) which asked subjects to balance potential reward with loss over 10 trials.<sup>20</sup> To increase sensitivity, we asked the subjects indicate their desired level of risk by deciding how much to blow up the balloon as a single total at the beginning of each trial (the 'Automatic BART'), rather than in small increments, and added a relatable reward.<sup>19, 31</sup>

As we only had one day with each subject, we paid particular attention to mitigating the practice effects of the cognitive tests. We used pseudo-randomization of stimuli<sup>2, 32</sup>, took two baselines (discarding the results from the first) and also discarded the first six trials of each subsequent iteration of the test.<sup>7</sup> To address any remaining practice effects, we subtracted the median score from the control group from the individual scores of the chamber subjects in a second analysis (see statistical analysis, below). We used these techniques in preference to the 'practice to asymptotic' approach, out of concern that boredom and fatigue might influence results if the cognitive tests had to be repeatedly practiced on the same day as the study.<sup>7</sup>

We asked subjects to subjectively rate the following symptoms before and at the end of their simulated flight (at peak altitude): headache, nausea, shortness of breath, loss

of coordination, hands uncomfortably cold, hands uncomfortably hot, body uncomfortably cold, body uncomfortably hot. We also asked subjects to subjectively rate the realism of their simulated flight experience. They used a categorical scale from 1-5: "None", "Slight", "Moderate", "Very", "Extremely".

### *Procedure*

Following baseline testing, the pilots ascended to 1524 m, 2438 m and 3658 m of simulated altitude (Figure 2). Ascent rate was  $2 \text{ m}\cdot\text{s}^{-1}$  and, on arrival at each target altitude, the pilots were given five minutes for equilibration before beginning the tests. They spent 25 minutes at each altitude, with a total flight time of 115-120 minutes (subjects were allowed to leave the chamber once the final tests were completed). The timepoints of testing in the control group were matched to the median times of testing in the simulator group. The control group undertook the cognitive tests seated, indoors (sea level altitude, ambient temperature  $18^{\circ}\text{C}$ ).

[Fig. #2 here]

### *Statistical analysis*

We downloaded monitoring data into LabChart via Powerlab 4/30 (AD Instruments, Oxford, UK) and Metasoft Studio (CORTEX Biophysik GmbH, Leipzig, Germany). We then collated and analyzed the ANAM, POMS, BART and monitoring data on a 1 Hz timebase in R Studio (Version 1.0.143, R Core Development Team, version 3.4.1). We chose non-parametric tests, given the small sample sizes and the heterogenous distributions of the results, assessed using descriptive methods (skewness, outliers, and distribution plots) and inferential statistics (Shapiro–Wilk test). We compared the raw scores from ANAM, POMS and BART in

the simulator and control groups using the Wilcoxon Rank Sum (Mann-Whitney  $U$ ) test with Holm-Bonferroni correction for multiple comparisons. We then performed a second analysis of the ANAM results from the simulator group, this time subtracting the median score of the control group at each altitude from the scores of each subject in the simulator group to mitigate any residual practice effects.<sup>10</sup> These corrected scores are displayed in Figure 3 and were summarized using Friedman followed by Wilcoxon Signed Rank tests (comparison of 0 m, 1524 m, 2438 m, 3658 m with Holm-Bonferroni correction). We calculated effect sizes in the Friedman tests using Kendall's coefficient of concordance ( $W_t$ ). In Figure 3, and where confidence intervals were required for comparisons with previous papers, we derived the pseudomedian and 95% confidence intervals from the Wilcoxon Signed Rank test. We also used the Wilcoxon Signed Rank test (Holm-Bonferroni correction) to compare the pre- and post- values for physiological parameters and symptom scores. We looked for correlations between the ANAM, POMS and BART results directly, and between the ANAM, POMS and BART results and SpO<sub>2</sub> and finger and rectal temperatures using Spearman's rank correlation. This necessitated 54 separate comparisons, and so where relevant, we present the results both uncorrected, and corrected for multiple comparisons (Holm-Bonferroni correction). We set significance as  $p < 0.05$  and reported descriptive statistics as median (IQR).

## Results

Results from the ANAM cognitive tests and BART are displayed in Table I and the POMS in Table II.

In the ANAM cognitive tests, there were no significant differences in raw scores between the simulator and the control groups at any time point. In the second analysis, using the corrected scores, Friedman tests identified significant differences in Logical Reasoning, Mathematical Processing, Manikin Variations and Stroop Tasks in the simulator group at the different altitudes. Post hoc comparison tests indicated significant differences in Logical Relations (Figure 3A) scores between baseline and 1524 m ( $p = 0.012$ ), 2438 m ( $p = 0.012$ ), and 3658 m ( $p = 0.012$ ). Peak scores in Logical Relations were at 1524 m, rather than baseline, likely representing a residual learning effect, then fell with increasing altitude (but without statistical significance). In Mathematical Processing (Figure 3B), there were significant decrements between baseline and 1524 m ( $p = 0.035$ ) and 2438 m ( $p = 0.049$ ); In Manikin Variations (Figure 3C), there was a significant decrement between baseline and 1524 m ( $p = 0.023$ ); in Stroop (Figure 3D) between 1524 m and 2438 m ( $p = 0.029$ ) and 2438 m and 3658 m ( $p = 0.012$ ). However, the pattern in each of these was a decrement between baseline and 1524 m, then an improvement on climbing to 2438 m, and then a further decrement between 2438 m and 3658 m (Figure 3). The ANAM results, both uncorrected and corrected for practice effects were not correlated with SpO<sub>2</sub> or rectal or finger skin temperatures, with the exception of the Tower Puzzle. Uncorrected for multiple comparisons, performance in the Tower puzzle was correlated with SpO<sub>2</sub> ( $r = 0.36$ ,  $p = 0.021$ ), rectal temperature ( $r = 0.44$ ,  $p = 0.0044$ ) and finger skin temperature ( $r = 0.34$ ,  $p = 0.040$ ). However, as performances improved with falling SpO<sub>2</sub>, rectal and finger skin temperatures (and therefore, with each further iteration of the test), this was likely to have been a residual practice effect. Indeed, when looking for correlations using the ANAM scores corrected for practice effects, this effect was abolished.

[Table I here]

[Fig. #3 here]

In the BART, we identified no significant differences in behavior with increasing altitude or between the simulator and control groups, nor was subjects' behavior correlated with any of the mood dimensions measured in the POMS, rectal, or finger skin temperatures.

[Table II here]

In the POMS, we also identified no significant differences between the simulator and the control groups. Within the simulator group, Friedman tests indicated significant differences in anxiety, happiness, and restlessness scores (Table II). However, in post hoc pairwise tests, following correction for multiple comparisons, none of the differences were significant. Uncorrected for multiple comparisons, finger skin temperatures were negatively correlated with the anger metric of the POMS ( $r_s = -0.45$ ,  $p = 0.006$ ). In other words, the three subjects with the coldest hands at peak altitude were the only ones to report anger scores  $> 0$ . Rectal temperature was also correlated with fatigue ( $r_s = -0.49$ ,  $p = 0.001$ ). However, rectal temperature declined steadily from supranormal to normal (see below), so this may have represented increased fatigue with time, rather than with deep body temperature.

The metabolic parameters recorded during the simulated flights are displayed in Table III. Baseline values were somewhat elevated in the simulator group, perhaps in anticipation of their upcoming 'flight'. However, as altitude then increased from 1524 to 3658 m there were statistically significant increases in oxygen consumption ( $p < 0.001$ ) and minute volume ( $p < 0.001$ ), with corresponding falls in oxygen saturations ( $p < 0.001$ ), though RER remained relatively static.

Deep body temperature fell by a statistically significant ( $p < 0.001$ ) amount from baseline to peak altitude (37.6 (0.4) to 37.3 (0.1) °C). However, as 37.3°C is closer to normal body temperature and the pilots were all hot at the start of their flight (having been dressed warmly), this likely represented a return to normothermia, rather than incipient hypothermia. Peripheral (finger) temperature fell much more steeply, from 32.2 (7.6) to 13.9 (9.0) °C ( $p < 0.001$ ).

There was a small, but statistically significant fall in blood glucose between baseline (4.7 (0.2) mmol·L<sup>-1</sup>) and the end of the flight (4.3 (0.3) mmol·L<sup>-1</sup>,  $p = 0.013$ ) but not in blood lactate. Urine output during the flight varied considerably between pilots (median 301 (287) mL), equating to 1.4 (1.3) mL·kg<sup>-1</sup>·hr<sup>-1</sup>, with a statistically insignificant median change in urine osmolality from 650 (475) to 425 (290) mOsm·kgH<sub>2</sub>O<sup>-1</sup>. Other insensible losses, derived from pre- and post-flight, post-voiding weights, equated to 2.3 (2.0) mL·kg<sup>-1</sup>·hr<sup>-1</sup>.

The simulator values are presented in Table III next to values recorded during a previous study at similar altitudes (simulated altitude  $\pm 100$  m) by four pilots in real flight.<sup>38</sup> Direct comparisons should be made with some caution, as the live flight pilots were passing through the altitude range (either ascending or descending) at various points during their flights, rather than following a standardized ramped profile, as in the simulator experiments. Equally, the baseline values from the live flight recordings were measured in camp, several hours removed from the flight environment, unlike in the simulator group. However, the results from both groups followed similar trends.

[Table III here]

Subjects did not report statistically significant changes in any of the symptom scores, though scores for cold hands increased from 1.0 (0) to 2.5 (2.5) between baseline and peak altitude ( $p = 0.168$ ). They scored the realism of the simulator experience as: overall realism

= 3.7 (0.8), body position = 4.7 (0.5), arm position = 4.5 (0.5), headwind sensation = 4.0 (0.8), temperature sensation = 4.2 (1.0), physical comfort = 4.2 (0.6), mental effort = 3.3 (0.2).  
(Categorical scale, 1-5, "Not at all" to "Extremely".)

Peripheral oxygen saturations in our simulator group closely matched those recorded at equivalent altitudes in hypobaric hypoxia by Legg et al.<sup>17</sup> and Pilmanis et al.<sup>30</sup> (Table IV). We plotted 95% confidence intervals for the metrics we measured in common with both authors: Logical Relations and Mathematical Processing (Accuracy, Mean Reaction Time, Throughput), Mannikin (Throughput), Stroop (Block Ratio 3:2), Tower Puzzle (Actual Number of Moves Made, Mean Reaction Time). All overlapped with our results, with the exception of the mean reaction time values in Pilmanis et al.<sup>30</sup> (this was to be expected: our subjects were wearing thick gloves and operating in a cold environment) and the baseline values for Logical Relations (ours were lower).

[Table IV here]

In the Profile of Mood States, Legg et al.<sup>17</sup> reported increased fatigue and decreased vigor at peak altitude (3658 m). We also observed similar trends in fatigue ( $p = 0.85$ ) and vigor ( $p = 0.35$ ), but these were not statistically significant. (Table II). Confidence intervals again overlapped for all the metrics except happiness: our subjects reported higher levels of happiness at all time points.

## Discussion

We investigated whether the unique combination of environmental stressors in the paragliding flight environment might elicit 'mental states or limitations [that might be] preconditions for error'.<sup>33</sup>



We found no significant differences between the simulator and control groups at any altitude. As expected from previous studies, individual performances were variable: some subjects tended to be relatively consistent performers, others more erratic. It may be that the circumstances exaggerated subjects' natural cognitive strengths and weaknesses, but no individuals were found to be grossly outlying.

We also noted two small, but useful, secondary outcomes. The first was that paraglider pilots may be particularly vulnerable to peripheral cold injury. Paraglider pilots fly with their hands above their heads, armpits and wrists exposed to the apparent wind. When first trialing the simulator, we risked breaching ethical limits for finger skin temperature as subjects' digits cooled so rapidly, despite wearing 'standard' flying equipment. For the main studies, we added additional hand protection (the only part of the flying set up that differed from the norm) so subjects would not have to be withdrawn from the study, an enlarged keyboard and a stylus so inputting test answers remained straightforward despite cold fingers. While flying a paraglider does not require much fine motor control, emergency procedures do (for example, deploying the reserve parachute) and unpleasant cold can also be a significant distraction. The second finding was that the pilots lost around  $4 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{hr}^{-1}$  of fluid. As only some pilots drink in flight, and some even actively fluid-restrict to avoid the need to urinate, thirst may also become a distraction over time.

Our results led us to reject our hypothesis that the paragliding flight environment might lead to mental states that may be preconditions for error, though in so doing we remained mindful of both the strengths and limitations of our study, and of cognitive testing in general. Drawing mechanistic conclusions on the role of low-grade environmental stressors on performance is challenging. For example, in simply studying hypoxia alone, there is interindividual variability in the hypoxic ventilatory response; while duration and

speed of exposure, as well as exercise, are all additional variables. Degrees of hypoxia cannot be easily quantified, especially on a regional rather than whole-body resolution: for example, peripheral oxygen saturation is rarely a satisfactory acute measure, as the results can be masked by hyperventilation and the shape of the oxyhemoglobin dissociation curve at low-moderate altitudes.<sup>13</sup>

The significance of cognitive tests as outcome measures has also been uncertain. Practice effects are a perennial problem, as are the wide variety of different tests used. It is also hard to relate a measured reduction in cognitive function on a test battery to real-world flight performance. (However, if an unvalidated (but more ecologically valid) live or simulated flight task were to be used instead, variation in performance may be due to skill differences, distractions or flying currency, as much as the environmental variables in question.)

Our aims were limited to 'ruling in' or 'ruling out' cognitive impairment as a significant source of pilot error in paragliding flight, rather than searching for mechanistic insights. Nonetheless, we had to consider many of the factors listed above in our study design. In so far as we were able, we built on the foundations laid by previous work. While our target altitudes were mainly chosen as being typical of good cross-country paragliding days in the U.K (1524 m), the Alps (2438 m), and the United States / Himalayas (3658 m), they also allowed us to compare our results (with the additional stressors, ramped profile and slower ascent rate of our particular flight environment) with those from two other authors who focused on hypoxia alone<sup>17, 30</sup>. We used the same validated test battery as those authors and included their chosen test outcome metrics in our data collection. However, for our primary outcome metrics, we picked measures that placed less emphasis on reaction times, knowing that subjects' muscle responses might be slowed due to

peripheral cooling. We added the BART test for risk-taking behavior and made strenuous efforts in both the data collection and the analysis to control for practice effects. With regard to recording subjects' physiology, we went beyond oxygen saturations, measuring oxygen consumption, carbon dioxide production and ventilation, deep body and peripheral temperatures. Through the use of real equipment and chase-cam footage, we recreated the feel of the paragliding flight environment as best we could, given the resources available to us. Our efforts were reflected in the subjects' positive subjective scores of the simulator's realism, as well as in comparisons with previously recorded live flight data.

However, a simulator is not real flight. We have alluded to the risks of paragliding throughout the paper, but our subjects were always safe. This knowledge may have affected their performance, but we cannot quantify to what extent. The acute stress of an in-flight emergency would undoubtedly have reduced cognitive 'bandwidth', but the interaction of environmental stressors with a low-grade awareness of vulnerability (ever-present in live flight) lay beyond our laboratory model. Our study also gave no indications of the margins of 'cognitive safety': for example, if the pilots had begun to shiver in the cold, so increasing their oxygen consumption beyond any spare capacity in their arterial oxygen content, they may have deteriorated at that point.

Our study was further limited by its small number of (only male) subjects. If we had more time with the subjects, we would have preferred them to have practiced the cognitive tests until asymptotic on a separate occasion, and to have included more trials in the BART (ten is typically considered the minimum).<sup>37</sup> Conditions in the chamber were normobaric rather than hypobaric, though the degree of distinction between the two at lower altitudes is debated.<sup>8</sup> We would also have liked to include descent as a component of the study protocol<sup>28</sup>, but the chamber venting system limited our ability to do this in an ecologically

valid way. Each flight was limited to just over two hours and we may have seen more profound changes with longer times at each target altitude.<sup>3</sup>

Our study represented a major effort to recreate the paragliding flight environment in a laboratory context. It is concluded that cognitive abilities or risk-taking behavior are not grossly distorted by the simulated flight environment. This is a useful finding; before we began, we knew that pilot error, rather than equipment failure, was (and is) the main cause of paragliding accidents, but until this study, we did not know whether these errors were due to pilot impairment or pilot misjudgment. Based on our results, we are drawn to conclude it is the latter, though we remain mindful of the limitations of our study, in particular its lack of risk. While we cannot entirely disregard the effects of hypoxia, cold and headwind on the performance of paraglider pilots at low-to-moderate altitudes, we have shown that the discipline's limited resources may be better invested in improving pilots' judgement and actions. We should be mindful of the risk of cold injury in pilots, but a focus purely on the mitigation of environmental stressors is unlikely to prevent pilot error.

**Acknowledgements**

We gratefully acknowledge the kind efforts of the study participants and of Dr. Joe Costello (University of Portsmouth, statistical advice and review of the manuscript); Przemek Marek, Dr John Leach and Dr Eddy Davelaar (technical assistance); Sup'Air (material support: harnesses and helmets); Outdoors Company (material support: pilot clothing); Royal Aeronautical Society GP Olley Award (financial support to MW).

## References

1. International Organization for Standardization (ISO). Standard Atmosphere. Geneva, Switzerland: ISO; 1975.
2. Anastasi A, Urbina S. Psychological testing. 7 ed. Upper Saddle River, NJ: Prentice-Hall; 1997.
3. Angerer P, Nowak D. Working in permanent hypoxia for fire protection-impact on health. *Int Arch Occup Environ Health*. 2003;76(2):87-102.
4. Benedetto JM, Harris WC, Goernert PN. Automated Neuropsychological Assessment Metrics (ANAM) repeated assessment stability. Mankato, MN: Engineering Psychology and Human Performance Laboratory, 1995.
5. Bouak F, Vartanian O, Hofer K, Cheung B. Acute Mild Hypoxic Hypoxia Effects on Cognitive and Simulated Aircraft Pilot Performance. *Aerosp Med Hum Perform*. 2018;89(6):526-35.
6. British Hang Gliding and Paragliding Association. BHPA incident analysis 2011-2015. Leicester, UK: British Hang Gliding and Paragliding Association (BHPA), 2016.
7. Collie A, Maruff P, Darby DG, McStephen M. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc*. 2003;9(3):419-28.
8. Coppel J, Hennis P, Gilbert-Kawai E, Grocott MP. The physiological effects of hypobaric hypoxia versus normobaric hypoxia: a systematic review of crossover trials. *Extrem Physiol Med*. 2015;4:2.
9. Denison DM, Ledwith F, Poulton EC. Complex reaction times at simulated cabin altitudes of 5000 feet and 8000 feet. *Aerosp Med*. 1966(37):1010-3.

10. Duff K. Evidence-Based Indicators of Neuropsychological Change in the Individual Patient: Relevant Concepts and Methods. *Arch Clin Neuropsychol*. 2012;27(3):248-61.
11. Eonta SE, Carr W, McArdle JJ, Kain JM, Tate C, Wesensten NJ, et al. Automated neuropsychological assessment metrics: repeated assessment with two military samples. *Aviat Space Environ Med*. 2011;82:34-9.
12. Hang gliding and paragliding records [Internet]. Fédération Aéronautique Internationale (FAI). 2017 [cited 14 March 2017]. Available from: <http://www.fai.org/civil-records>.
13. Hodkinson P. Acute Exposure to Altitude. *J R Army Med Corps*. 2011;157(1):85-91.
14. Johnson DR, Vincent AS, Johnson AE, Gilliland K, Schlegel RE. Reliability and construct validity of the Automated Neuropsychological Assessment Metrics (ANAM) mood scale. *Arch Clin Neuropsychol*. 2008;23(1):73-85.
15. Kim CH, Ryan EJ, Seo Y, Peacock C, Gunstad J, Muller MD, et al. Low intensity exercise does not impact cognitive function during exposure to normobaric hypoxia. *Physiol Behav*. 2015;151:24-8.
16. Legg S, Hill S, Mundel T, Gilbey A, Schlader Z, Raman A. Could mild hypoxia impair pilot decision making in emergencies? *Work*. 2012;41 Suppl 1:198-203.
17. Legg SJ, Gilbey A, Hill S, Raman A, Dubray A, Iremonger G, et al. Effects of mild hypoxia in aviation on mood and complex cognition. *Appl Ergon*. 2016;53 Pt B:357-63.
18. Legg SJ, Hill S, Gilbey A, Raman A, Schlader Z, Mundel T. Effect of mild hypoxia on working memory, complex logical reasoning and risk judgement. *Int J Aviat Psychol*. 2014;24(2):126-40.
19. Lejuez CW, Aklin W, Daughters S, Zvolensky M, Kahler C, Gwadz M. Reliability and Validity of the Youth Version of the Balloon Analogue Risk Task (BART-Y) in the Assessment

of Risk-Taking Behavior Among Inner-City Adolescents. *J Clin Child Adolesc Psychol.*

2007;36(1):106-11.

20. Lejuez CW, Read JP, Kahler CW, Richards JB, Ramsey SE, Stuart GL, et al. Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *J Exp Psychol Appl.* 2002;8(2):75-84.

21. McMorris T, Hale BJ, Barwood M, Costello J, Corbett J. Effect of acute hypoxia on cognition: A systematic review and meta-regression analysis. *Neurosci Biobehav Rev.* 2017;74(Pt A):225-32.

22. Miller H. How fast does it go? Testing paraglider speeds. *Cross Country.* 2016:60-8.

23. Nishi S. Effects of altitude-related hypoxia on aircrews in aircraft with unpressurized cabins. *Mil Med.* 2011;176(1):79-83.

24. Nyenhuis DL, Yamamoto C, Luchetta T, Terrien A, Parmentier A. Adult and geriatric normative data and validation of the profile of mood states. *J Clin Psychol.* 1999;55(1):79-86.

25. Paragliding Manufacturers' Association. PMA research: Numbers of Paraglider Pilots Worldwide. PMA News [Internet]. 2014 April 27 2017. Available from: <http://www.p-m-a.info/news/article/pma-research-numbers-of-paraglider-pilots-worldwide.html>.

26. Peacock CA, Weber R, Sanders GJ, Seo Y, Kean D, Pollock BS, et al. Pilot physiology, cognition and flight performance during flight simulation exposed to a 3810-m hypoxic condition. *Int J Occup Saf Ergon.* 2016:1-6.

27. Petrassi FA, Hodkinson PD, Walters PL, Gaydos SJ. Hypoxic Hypoxia at Moderate Altitudes: Review of the State of the Science. *Aviation, Space, and Environmental Medicine.* 2012;83(10):975-84.



28. Phillips JB, Horning D, Funke ME. Cognitive and perceptual deficits of normobaric hypoxia and the time course to performance recovery. *Aerosp Med Hum Perform.* 2015;86(4):357-65.
29. Pighin S, Bonini N, Savadori L, Constantinos H, Antonetti T, Schena F. Decision making under hypoxia: Oxygen depletion increases risk seeking for losses but not for gains. *Judgment and Decision Making.* 2012;7(4):472-7.
30. Pilmanis AA, Balldin UI, Fischer JR. Cognition Effects of Low-Grade Hypoxia. *Aerosp Med Hum Perform.* 2016;87(7):596-603.
31. Pleskac TJ, Wallsten TS, Wang P, Lejuez CW. Development of an automatic response mode to improve the clinical utility of sequential risk-taking tasks. *Exp Clin Psychopharmacol.* 2008;16(6):555-64.
32. Roebuckspencer T, Sun W, Cernich A, Farmer K, Bleiberg J. Assessing change with the Automated Neuropsychological Assessment Metrics (ANAM): Issues and challenges. *Arch Clin Neuropsychol.* 2007;22:79-87.
33. Shappell S, Detwiler C, Holcomb K, Hackworth C, Boquet A, Wiegmann DA. Human Error and Commercial Aviation Accidents: An Analysis Using the Human Factors Analysis and Classification System. *Human Factors: The Journal of the Human Factors and Ergonomics Society.*
34. Short P, Cernich A, Wilken JA, Kane RL. Initial construct validation of frequently employed ANAM measures through structural equation modeling. *Arch Clin Neuropsychol.* 2007;22 Suppl 1:S63-77.
35. Smith AM. Acute Hypoxia and Related Symptoms on Mild Exertion at Simulated Altitudes Below 3048 m. *Aviation, Space, and Environmental Medicine.* 2007;78(10):979-84.

36. Taylor L, Watkins SL, Marshall H, Dascombe BJ, Foster J. The Impact of Different Environmental Conditions on Cognitive Function: A Focused Review. *Front Physiol.* 2015;6:372.
37. Wallsten TS, Pleskac TJ, Lejuez CW. Modeling behavior in a clinically diagnostic sequential risk-taking task. *Psychol Rev.* 2005;112(4):862-80. Epub 2005/11/03.
38. Wilkes M, MacInnis MJ, Hawkes LA, Massey H, Eglin C, Tipton MJ. The Physiology of Paragliding Flight at Moderate and Extreme Altitudes. *High Alt Med Biol.* 2017.

## Tables

Test	Metric	Baseline (FiO <sub>2</sub> 0.209)		1524 m (FiO <sub>2</sub> 0.174)		2438 m (FiO <sub>2</sub> 0.156)		3658 m (FiO <sub>2</sub> 0.133)		$\chi^2$	p	Wt
		Raw	-pe	Raw	-pe	Raw	-pe	Raw	-pe			
Logical Reasoning	TP	21 (7.3)	-7 (7.3)	28 (7.2)	3.1 (7.2)	25 (7)	-1.4 (7)	23 (8.6)	-3.4 (8.6)	22	<0.001	0.72
Maths Processing	TP	22 (6.4)	1.4 (6.4)	21 (11)	-4.6 (11)	22 (15)	-3.5 (15)	23 (7.1)	-3.2 (7.1)	10	0.018	0.34
Mannikin Variations	TP	52 (25)	13 (25)	55 (18)	5.4 (18)	55 (25)	10 (25)	57 (6.3)	3.6 (6.3)	13	0.005	0.42
Stroop	IS	16 (10)	1.8 (10)	20 (6)	1.4 (6)	21 (6)	3 (6)	20 (6.3)	-2.2 (6.3)	15	0.002	0.49
Tower	MR	1.1 (0.19)	-0.06 (0.19)	1.2 (0.25)	0.01 (0.24)	1.1 (0.25)	-0.11 (0.26)	1 (0.06)	-0.05 (0.06)	1.6	0.670	0.05
BART	AAPC	64 (13)	-	55 (18)	-	50 (11)	-	60 (18)	-	3.0	0.390	0.1

Table I. Results for the ANAM ( $n = 10$ , raw scores and scores corrected for practice effects ‘(-PE)’) and BART (raw scores only). Presented as median (IQR), followed by the outcomes of Friedman tests and Kendall's coefficient of concordance (Wt) on the ANAM corrected scores (3 df) and BART raw score. Metrics: TP (throughput), IS (interscore), MR (move ratio), AAPC (adjusted average pump count).

Test	Baseline (FiO <sub>2</sub> 0.209)	1524 m (FiO <sub>2</sub> 0.174)	2438 m (FiO <sub>2</sub> 0.156)	3658 m (FiO <sub>2</sub> 0.133)	$\chi^2$	p	Wt
Anger	0 (0)	0 (0)	0 (0)	0 (0.1)	7.2	0.066	0.24
Anxiety	0.2 (0.6)	0 (0)	0 (0)	0 (0)	9.4	0.025	0.31
Depression	0 (0)	0 (0)	0 (0)	0 (0)	3.0	0.390	0.1
Fatigue	0.8 (0.7)	0.7 (0.5)	0.6 (1.1)	1.1 (1.6)	1.1	0.790	0.04
Happiness	4.6 (0.6)	4.7 (1.0)	4.3 (0.6)	4.1 (1.2)	11	0.011	0.37
Restlessness	0.6 (0.8)	0.3 (0.3)	0.3 (0.6)	0.6 (0.8)	9.0	0.029	0.30
Vigour	3.7 (1.7)	3.5 (1.1)	2.9 (1.1)	3 (1.1)	4.2	0.240	0.14

Table II. Results for the Profile of Mood States (POMS),  $N = 10$ , raw scores presented as median (IQR), followed by the outcomes of Friedman tests and Kendall's coefficient of concordance (3 df).

Parameter	Baseline		1524 m		2438 m		3658 m	
	(FiO <sub>2</sub> 0.209)		(FiO <sub>2</sub> 0.174)		(FiO <sub>2</sub> 0.156)		(FiO <sub>2</sub> 0.133)	
	Sim	Live	Sim	Live	Sim	Live	Sim	Live
VO <sub>2</sub> mL·kg <sup>-1</sup> ·min <sup>-1</sup>	6.2 (0.6)	3.9 (0.6)	5.9 (0.6)	5.0 (2.6)	6.10 (0.7)	5.1 (2.1)	6.8 (0.5)	-
VCO <sub>2</sub> mL·kg <sup>-1</sup> ·min <sup>-1</sup>	6.2 (0.5)	3.6 (0.6)	5.7 (0.5)	5.7 (3.1)	5.8 (0.7)	5.2 (1.7)	6.5 (0.5)	-
VE L·min <sup>-1</sup>	16 (1.5)	8.2 (1.3)	13 (0.8)	22 (7.9)	13 (1.0)	20 (5.9)	16 (0.9)	-
TV L·min <sup>-1</sup>	0.74 (0.04)	0.56 (0.07)	0.72 (0.06)	0.78 (0.24)	0.73 (0.07)	0.77 (0.2)	0.87 (0.13)	-
RR Breath·min <sup>-1</sup>	22 (2.4)	16 (2.3)	20 (2.0)	29 (5.8)	19 (1.2)	28 (4.5)	19 (3.4)	-
HR Beats·min <sup>-1</sup>	82 (4.3)	75 (2.4)	76 (1.7)	100 (9.0)	73 (3.2)	91 (10)	74 (3.4)	-
SpO <sub>2</sub> %	98 (1.0)	-	97 (1.1)	-	94 (2.2)	-	88 (4.2)	-
Rectal temp °C	37.6 (0.4)	-	37.5 (0.3)	-	37.4 (0.2)	-	37.3 (0.1)	-
Finger temp °C	33.7 (7.5)	-	32.9 (11.0)	-	22.4 (14.4)	-	13.7 (10.6)	-

Table III. Metabolic (median (IQR)) values recorded in the environmental simulator (n = 10), alongside those recorded at similar altitudes (simulator altitude ± 100 m, n = 4) in real flight.<sup>38</sup>

Group	Baseline (FiO <sub>2</sub> 0.209)	1524 m (FiO <sub>2</sub> 0.174)	2438 m (FiO <sub>2</sub> 0.156)	3658 m (FiO <sub>2</sub> 0.133)
Simulator group	98.3 (0.5)	96.5 (0.8)	93.2 (2.2)	88.5 (4.0)
Pilmanis et al.	97.5 (0.8)	96.0 (-)	93.5 (-)	86.6 (2.9)
Legg et al.	99.0 (1.0)	-	95.0 (3.0)	88 (3)

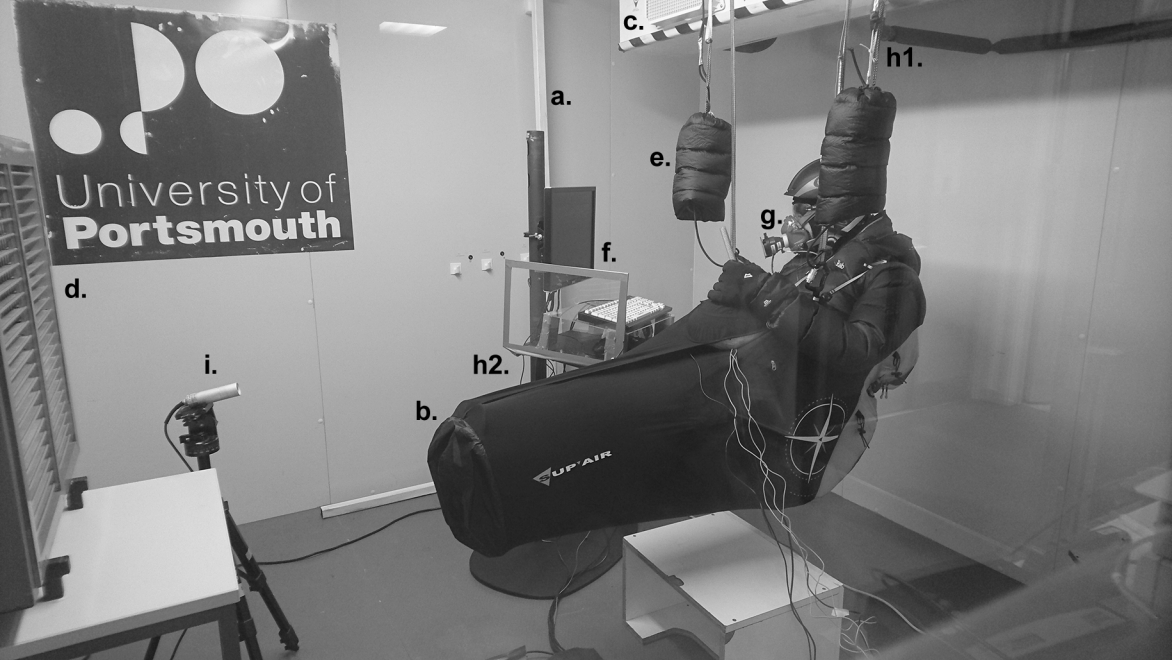
*Table IV. Peripheral oxygen saturations recorded during the flights in the environmental simulator in normobaric hypoxia (n = 10), alongside those reported at equivalent altitudes in hypobaric hypoxia by Legg et al. (n = 36)<sup>17</sup> and Pilmanis et al. (n = 91)<sup>30</sup> (Mean and sd).*

### Captions for figures

**Figure 1.** The flight environment simulator: (a) Custom-welded hollow steel frame; (b) Semi-recumbent cocoon harness; (c) Cooling system; (d) Fan unit; (e) Down sleeves, containing weighted brake lines and brake handles; (f) Computer monitor, Perspex shield, enlarged keyboard and stylus; (g) Metamax 3b; (h 1-2) Ambient temperature sensors; (i) Monitoring camera; (Out of shot) Projector and chase cam video, oxygen saturation logger with ear probe, skin and rectal temperature sensors.

**Figure 2.** Measured oxygen concentration (% , solid line) and chamber temperature (°C, dotted line) during a simulated flight.

**Figure 3.** ANAM Cognitive Tests and BART results. ANAM scores have been corrected for residual practice effects by subtracting the median score for each test achieved by the control group from individual subjects' scores. Each dot represents an individual's score, the black cross the pseudomedian and the error bars 95% confidence intervals. (BL = 'baseline').



c.

h1.

a.

e.

g.

f.

University of  
**Portsmouth**

d.

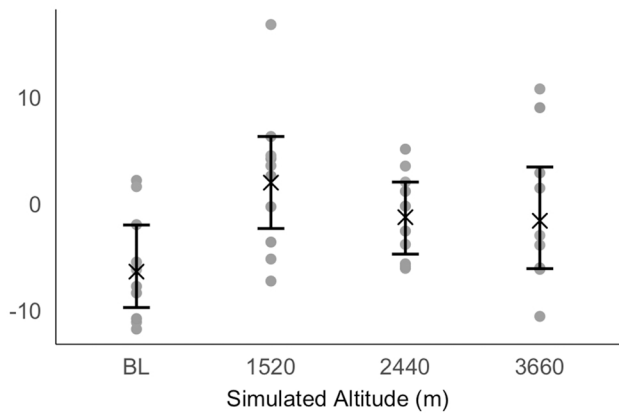
i.

h2.

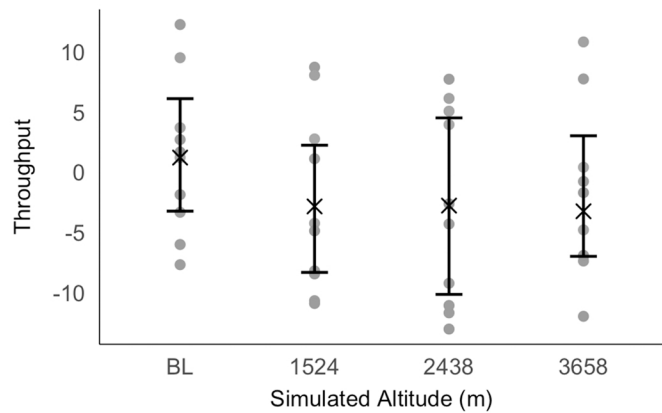
b.

SUPAIR

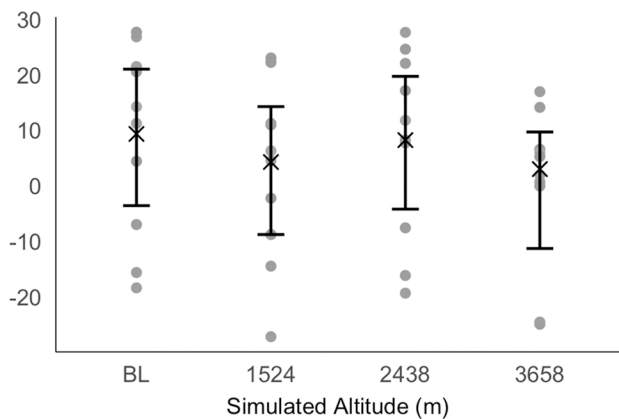
(A) Logical Relations



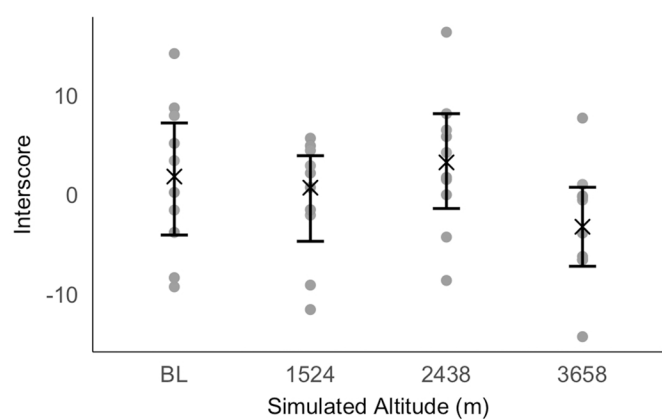
(B) Mathematical Processing



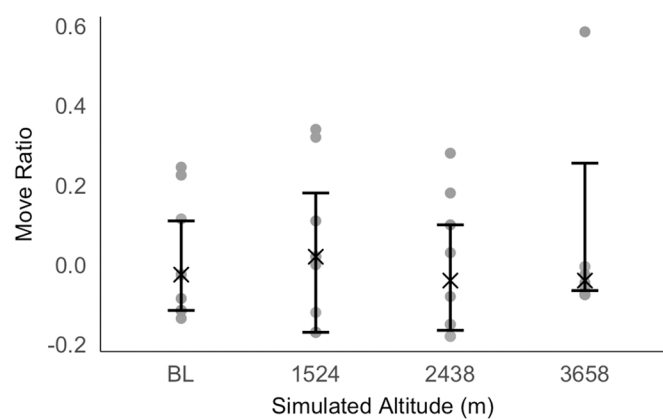
(C) Mannkin



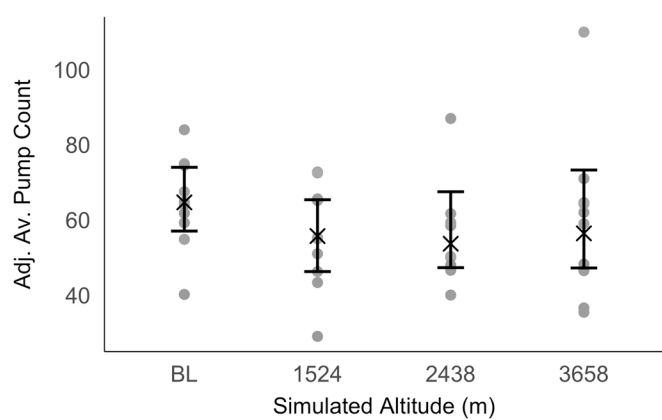
(D) Stroop



(E) Tower



(F) Balloon Analogue Risk Task





FiO<sub>2</sub> (%) / Temperature (°C)

