

The respiratory resistance sensitivity task: An automated method for quantifying respiratory interoception and metacognition

Journal Article**Author(s):**

Nikolova, Niiia; Harrison, Olivia; Toohey, Sophie; Brændholt, Malthe; Legrand, Nicolas; Correa, Camile; Vejlø, Melina; Jensen, Martin Snebjerg; Fardo, Francesca; Allen, Micah

Publication date:

2022-04

Permanent link:

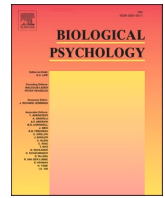
<https://doi.org/10.3929/ethz-b-000545759>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Originally published in:

Biological Psychology 170, <https://doi.org/10.1016/j.biopsycho.2022.108325>



The respiratory resistance sensitivity task: An automated method for quantifying respiratory interoception and metacognition

Niia Nikolova^{a,*}, Olivia Harrison^{b,c,d}, Sophie Toohey^b, Malthe Brændholt^a, Nicolas Legrand^a, Camile Correa^e, Melina Vejlø^a, Martin Snebjerg Jensen^a, Francesca Fardo^{a,f}, Micah Allen^{a,g,h}

^a Center of Functionally Integrative Neuroscience, Aarhus University, Denmark

^b Department of Psychology, University of Otago, New Zealand

^c Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Switzerland

^d Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom

^e Department of Food Science, Aarhus University, Denmark

^f Danish Pain Research Center, Aarhus University, Denmark

^g Aarhus Institute of Advanced Studies, Aarhus University, Denmark

^h Cambridge Psychiatry, University of Cambridge, United Kingdom

ARTICLE INFO

Keywords:

Respiration
Interoception
Metacognition
Psychophysics

ABSTRACT

The ability to sense, monitor, and control respiration – e.g., respiratory interoception (henceforth, respiratory interoception) is a core homeostatic ability. Beyond the regulation of gas exchange, enhanced awareness of respiratory sensations is directly related to psychiatric symptoms such as panic and anxiety. Indeed, chronic breathlessness (dyspnea) is associated with a fourfold increase in the risk of developing depression and anxiety, and the regulation of the breath is a key aspect of many mindfulness-based approaches to the treatment of mental illness. Physiologically speaking, the ability to accurately monitor respiratory sensations is important for optimizing cardiorespiratory function during athletic exertion, and can be a key indicator of illness. Given the important role of respiratory interoception in mental and physical health, it is unsurprising that there is increased interest in the quantification of respiratory psychophysiology across different perceptual and metacognitive levels of the psychological hierarchy. Compared to other more popular modalities of interoception, such as in the cardiac domain, there are relatively few methods available for measuring aspects of respiratory interoception. Existing respiratory loading tasks are difficult to administer and frequently require expensive medical equipment, or offer poor granularity in their quantification of respiratory-related perceptual ability. To facilitate the study of respiratory interoception, we here present a new, fully automated and computer-controlled apparatus and psychophysiological method, which can flexibly and easily measure respiratory-related interoceptive sensitivity, bias and metacognition, in as little as 30 min of testing, using easy to make 3D printable parts.

1. Introduction

In a most general sense, perception is the detection, organization and manipulation of sensory information. The study of internal perception, hereafter interoception, targets sensations arising from within the body, and can be defined to include perceptions originating from the visceral organs (e.g., heart, stomach) as well as processes involving them, such as respiration (Sherrington, 1952; Vaitl, 1996). Interoception influences diverse cognitive processes including sensory perception, decision-making, memory, and emotion (Al et al., 2020; Critchley & Garfinkel, 2017; James, 1884, 1894; Nikolova, Waade, Friston, & Allen,

2021; Park et al., 2014), and is increasingly seen as a core factor in psychiatric illness (Allen, 2020; Allen, Legrand, Correa, & Fardo, 2020; Khalsa et al., 2018; Khalsa & Lapidus, 2016; Owens, Allen, Ondobaka, & Friston, 2018). To date however, the large majority of interoception research is based within the cardiac domain, limiting our understanding of other modalities such as respiratory or gastric interoception.

Respiratory interoception (respiroception) is an interoceptive modality of particular interest due to its close linkage to affect (Guz, 1997; Meuret, Rosenfield, Hofmann, Suvak, & Roth, 2009). Similar to cardiac sensation, respiroception is fundamentally a mixture of a variety of different visceral, tactile, and exteroceptive sensory cues including those

* Corresponding author.

E-mail address: niia@cfni.au.dk (N. Nikolova).

<https://doi.org/10.1016/j.biopsycho.2022.108325>

Received 19 October 2021; Received in revised form 21 March 2022; Accepted 24 March 2022

Available online 28 March 2022

0301-0511/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

arising from the smooth muscle of the diaphragm, the detection of CO₂, and the activation of stretch and baroreceptors throughout the neck and chest wall (Khalsa, Rudrauf, Feinstein, & Tranel, 2009; Knapp-Kline, Ring, Emmerich, & Brener, 2021). While exteroceptive cues can be involved in breathing perceptions, here we refer to this process as respiratory interoception as these signals are largely denoting the state of the body, and thus interoceptive in nature. Feelings of breathlessness, or dyspnea, are a core symptom of panic disorder and anxiety (Bailey, 2004; Manning & Schwartzstein, 1995; Morélot-Panzini et al., 2007), and a variety of cognitive and mindfulness-based interventions seek to improve respiratory awareness and control (Tweeddale et al., 1994). Unlike gastric or cardiac modalities, respiration is often amenable to direct conscious access, by increasing or decreasing the breathing rate, by holding the breath, or by combinations of these actions (e.g., syncope breathing). The possibility to retrain or otherwise improve respiratory awareness and control is therefore of substantive clinical interest. Furthermore, respiratory cycles are known to directly affect both cardiac and gastric (Porges, 1992) activity through basic reflexes in the autonomic nervous system (an example being the respiratory sinus arrhythmia). A more thorough understanding of respiratory perception may thus offer new avenues for treating known interoceptive deficits associated with cardiac and gastric sensations (Khalsa & Lapidus, 2016; Quadt et al., 2018), as seen in anxiety, panic disorder or medically unexplained symptomatology (Bogaerts et al., 2008; Harrison, Marlow, Finnegan, Ainsworth, & Pattinson, 2021; Harrison, Nanz, et al., 2021; Tiller et al., 1987; van Dyck et al., 2021). For example, both carbon dioxide inhalation (Gorman et al., 2001) and isoproterenol (Pohl et al., 1988) have been shown to induce anxiety in patients with panic disorder, indicating that the symptom is amenable to modulation through both respiratory and cardiac modalities.

Several methods have been used to study analogs of different breathing-related symptoms. Carbon-dioxide inhalation has been used to effectively simulate hypercapnia, the build-up of carbon dioxide in the blood, due for example to chronic obstructive pulmonary disease (COPD) or extended breath hold (e.g., during diving). Carbon dioxide inhalation has been used to demonstrate the contribution of carbon dioxide regulation in anxiety and eating disorders (Eke & McNally, 1996; Pappens, De Peuter, Vansteenwegen, Van den Bergh, & Van Diest, 2012; Perna et al., 2004; Vickers, Jafarpour, Mofidi, Rafat, & Woznica, 2012; Woznica, Vickers, Koerner, & Fracalanza, 2015). Monitored exercise, for example on a cycle ergometer, has been used to study exertional dyspnea and onset of hyperventilation (Meyer, 2004; Takano & Deguchi, 1997; Ward et al., 2001).

Respiratory-related interoception has often been measured using resistance-detection paradigms, in which either inspiratory or expiratory breaths through a circuit can be made more or less difficult (Bennett, Jayson, Rubenstein, & Campbell, 1962; Dahme et al., 1996; Garfinkel et al., 2016; Harrison, Garfinkel, et al., 2021; Harver et al., 1993; Wiley & Zechman, 1966). In fact, studies have used resistive loads for over 50 years with a wide variety of methodologies ranging from a circuit of oil drums with varying volumes (Campbell et al., 1961) to mesh filters (Harrison, Garfinkel, et al., 2021; Noble, Frankel, Else, & Guz, 1971). For example, breathing difficulty can be manipulated by adding or subtracting static filters containing a small amount of resistance to the circuit. Resistive loads have been previously used to study dyspnea, airway obstruction and anxiety (Gottfried, Altose, Kelsen, Fogarty, & Cherniack, 1978; Paulus, 2013). Some studies have measured how magnitude estimation (i.e., using the Borg scale for physical exertion, on which exertion on each trial is rated on a scale from 'none' to 'maximal', (Borg, 1982)) and perceived unpleasantness relate to stimulus intensity as an index of reactivity (Kifle et al., 1997; Tsai, Chan, von Leupoldt, & Davenport, 2013; von Leupoldt & Dahme, 2005). Although similar in spirit to resistance sensitivity paradigms, these reactivity measures have aimed to assess the subjective experience of stimuli, or how steeply subjective perception increases with stimulus intensity, rather than objective sensitivity to detecting the presence of a stimulus,

or how steeply detection increases with stimulus intensity. Still others have investigated resistive expiratory loads, but while expiratory resistive loads affect blood carbon dioxide levels, inspiratory loading gives stable levels in carbon dioxide levels (Lofaso, Isabey, Lorino, Harf, & Scheid, 1992; Lopata, La Fata, Evanich, & Lourenço, 1977). This technique can be paired with signal-theoretic approaches and psychophysical methods such as adaptive staircase procedures to measure perceptual sensitivity towards resistive loads, indexing the individual ability to detect pressure in the airway and muscle force in the diaphragm.

However, these and similar tasks suffer from several drawbacks. Previous methods for altering resistance suffer from a lack of stimulus granularity, such that the derived thresholds are extremely noisy or unreliable, and require many trials to stabilize (Harver et al., 1993). Further, resistive loads can be inherently aversive, obscuring the relationship between the objective physical detection of respiratory sensations and associated subjective affect. This combination of coarse stimulus granularity resulting in larger than necessary resistive loads, as well as the inherent aversiveness of high resistive loads, and long testing times (typically 60+ minutes in total) cumulatively make these tasks difficult for healthy participants. Crucially, these tasks are out of reach for many of the clinical populations for whom these measures would be of most importance, such as those suffering from anxiety or respiratory distress disorders (e.g., moderate to severe asthma). Additionally, classical methods only allow for the estimation of thresholds, and do not provide information regarding the slope of the psychometric function for respiratory resistance, which relates to the precision or uncertainty of the perceptual process. Clearly a more reliable, precise, and automated procedure is called for.

To achieve these aims, we designed a fully automated, 3D printable respiratory apparatus for delivering precise inspiratory and expiratory loads. Our apparatus builds on previous approaches (Garfinkel et al., 2016; Harrison, Garfinkel, et al., 2021) to enable fully computer-controlled estimation of respiroceptive thresholds, as well as improved estimation of other signal theoretic parameters such as sensitivity, bias, and metacognition. We further developed an accompanying Bayesian adaptive psychophysical approach, the respiratory resistance sensitivity task, to estimate the full respiratory psychometric function (PMF) relating stimulus level to probability of a correct response. As an initial validation of our approach, we applied this method in a sample of 32 healthy subjects. Our findings demonstrate that the respiratory resistance sensitivity task (RRST) can quickly and reliably estimate respiratory thresholds in under 30 min, with minimal subjective aversiveness. We additionally ran a direct comparison between the RRST and a recently published method (the Filter Detection Task, or FDT) in 15 healthy individuals, demonstrating much improved control in task accuracy (see [Supplementary material](#) for details).

2. Methods

2.1. Participants

Thirty-three participants (20 females) were recruited through the Aarhus University Centre for Functionally Integrative Neuroscience Sona system participant pool. Their ages ranged from 19 to 66 years, with a mean of 28.09 years and standard deviation of 9.08 years. One participant was excluded from all data analysis due to non-convergence of the psychophysical staircases, resulting in 32 participants for the final analyses. For the analysis of metacognition performance, a further two participants were excluded due to too many extreme confidence ratings. All participants had no current psychiatric diagnosis, did not use drugs or medication that can alter psychological functions, had normal or corrected-to-normal vision, and fluent English or Danish proficiency. Although we did not exclude persons with a possible history of prior respiratory illness, no participants indicated they had any such history and only two participants provided asthma symptom severity ratings

above 10% (see Section 3.5).

In total, the experiment lasted 60–75 min, and participants were reimbursed for their time. The study was approved by the local Region Midtjylland Ethics Committee and was carried out in accordance with the Declaration of Helsinki.

2.2. Device specification

To enable the precise, automated delivery of respiratory resistive loads, we developed a novel apparatus based on previous resistive load tasks (Garfinkel et al., 2016; Harrison, Garfinkel, et al., 2020). The primary mechanism of the device is a 3D-printed housing which secures a section of flexible tubing against a wedge. This tubing is then connected to a sanitary, hospital grade respiratory circuit by custom-fit couplers. The wedge is connected by a screw to a step motor, and is then driven forwards or backwards against the flexible tubing by the motor. The motor itself is connected to an Arduino circuit board which is programmed with the instructions for converting digital inputs into discrete steps along the screw. Thus, by delivering electronic commands to the step motor, the wedge moves forwards or backwards against the length of tubing, resulting in a reliable stepwise compression or relaxation, i.e., an increase or decrease of the static resistance through the full respiratory circuit. At the participant end of the device a Hans Rudolph 2-way non-rebreathing t-valve connects the respiratory tubing to a PowerBreathe TrySafe filtered mouthpiece, through which the participant breathes (see Supplementary Fig. 1 – Detailed schematic for details).

A stepper motor is controlled by a TMC driver and Teensy 3.1 microcontroller. The motor is securely attached to a threaded TR8 × 2 leadscrew by a coupler. The TR8 × 2 leadscrew and nut are used due to their design for use in 3D printers, and are therefore made to withstand repetitive and continuous movement. A motor damper and rubber feet are further used to reduce vibration and noise from the device. Commands to move to positions resulting in varying percentages of airway obstruction were determined by the psychophysical staircase (see Section 2.5 below) and were sent to the microcontroller by Psychtoolbox through a USB-A to micro-USB connection.

The RRST software was written using MATLAB (2020a), using Psychtoolbox 3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and the Palamedes toolbox (Prins & Kingdom, 2018). All task code, associated data analysis scripts, and anonymized participant data are publicly available on our GitHub page: <https://github.com/embodied-computat-ion-group/RespiroceptionMethodsPaper>.

2.3. Set-up

Participants performed the RRST while seated in front of a computer at a height-adjustable desk. The user-end of the breathing circuit with the single-use mouthpiece was positioned and held in place in front of the participant using a desktop microphone stand (see GitHub project repository for a photo of the set-up). Participants were encouraged to adjust the height of the desk so as to be able to comfortably lean forward and inhale through the mouthpiece while performing the task. Several precautions were taken to eliminate non-respiratory cues, such as visual, auditory and tactile cues associated with the movement of the device. First, participants were fitted with over-ear headphones playing continuous “rain” noise selected to mask sounds generated by the device. Second, the compressed end of the breathing circuit and the device were placed within a box located to the side of the participant, such that they were not able to see any movement. Third, to eliminate vibrations caused by the stepper motor, it was fitted with a damper and rubber feet were attached to the device. The apparatus was further placed on top of thick foam padding inside the box, which was positioned on a surface detached from the participants’ desk. Finally, the position of the load was changed to a random value during each inter-trial interval to decorrelate the duration of movement to the stimulus intensity on the

subsequent trial. These measures were taken as a result of extensive piloting using, for example, different acceleration settings for the motor driver, and varying noise sounds to play over the headphones, in order to ensure that participants could not use cues from vibrations and sound as an aid for detecting differences in resistance.

At the start of the first session, participants completed a tutorial session. This introduced the trial structure in a stepwise manner, using two trials with accuracy feedback, and six additional trials of varying difficulty without feedback. This ensured consistency in the task instructions, and gave participants a chance to learn and adapt to the trial structure and breathing pace before the start of the adaptive staircases. The tutorial lasted about 5 min, the QUEST staircase 15 min, and the Psi staircase 30 min including breaks.

2.4. Trial structure

To measure respiratory resistance sensitivity, we used a two-interval forced-choice (i.e., temporal two-alternative forced-choice) design (2IFC). When interested in determining measures of perceptual sensitivity, a 2IFC design is preferable to one where a single stimulus is presented on each trial because they are less susceptible to biases in the decision process. Indeed, it has been shown that participants can voluntarily shift their psychometric curve on a single-interval forced-choice (1AFC) task, without altering their sensitivity (Morgan, Dillenburg, Raphael, & Solomon, 2012).

On each trial, participants took two breaths, where one (the stimulus) had a resistive load applied to the breathing circuit while the other (the standard) did not. Which of the two breaths contained the resistive load was determined pseudo-randomly. After taking the two consecutive inhalations, participants decided whether the first or the second breath was more difficult, and entered their response using the left and right buttons of a computer mouse. They were then asked to rate how confident they were in their decision by using the mouse to move a slider on a visual analog scale (VAS) ranging from “Guess” to “Certain” and confirming the answer by a left-click on the mouse. The maximum response time for the decision was 3 s, and 5 s for the confidence rating. If no response was made during this time, the trial was repeated (see Fig. 1).

Each stimulus interval began with a 1500 ms “prepare to breathe” instruction. The inhalation pace was cued by a visual stimulus: the outline of a circle and a “first inhale or second inhale” cue appeared on the screen for 200 ms, then an expanding gaussian ring was displayed for 800 ms, during which time the ring grew to occupy the space within the circle. Participants were instructed to pace their breaths to the expansion of the ring, to breathe “sharply and shallowly” and that inhaled should last just under a second. We recommended participants to keep their mouth on the mouthpiece for the duration of each trial and exhaling through the nose between the two intervals, similar to breathing through a snorkel. This instruction was provided to minimize rebreathing of air from within the circuit. Aside from this suggestion, we instructed participants to breathe as was most comfortable for them and did not control the breathing strategy further (e.g., participants were free to exhale through the breathing circuit or outside it if they wished). To avoid hyperventilation, the task contained forced breaks of at least 2 min in duration every 20 trials (ca. every 5 min of testing). Participants were instructed to use these breaks to stand up, move around and breathe naturally. Prior to each break, participants rated how aversive they found the task stimuli using a VAS from “Not unpleasant” to “Very unpleasant”. At the end of the session, participants responded to subjective experience questions assessing “dizziness/light-headedness”, “breathlessness”, and “presence of asthma symptoms”.

2.5. Psychophysical methods

To estimate participants’ sensitivity to detecting small increments of obstruction of airflow on inhalations, we used established psychophysical methods to measure the PMF threshold and slope. The threshold

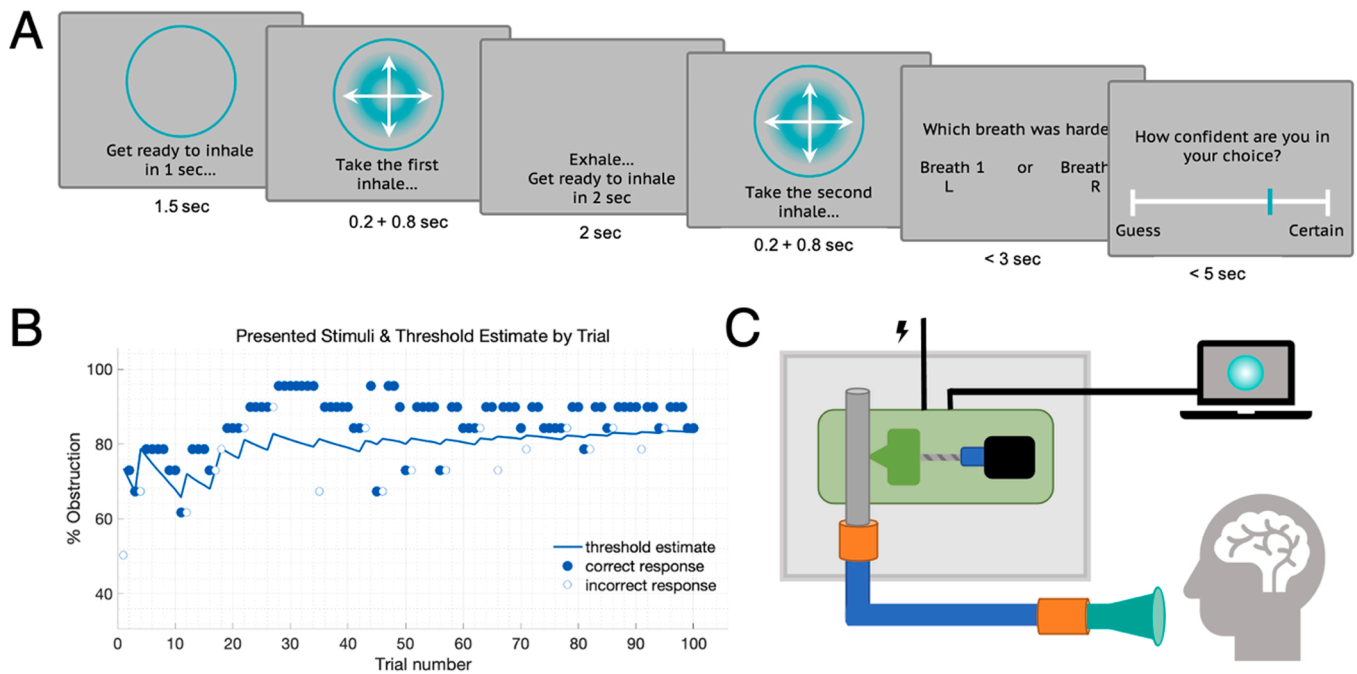


Fig. 1. The respiratory resistance sensitivity task. **A)** Trial schematic depicting the 2-interval forced choice (2IFC) design of the task. On each trial participants view a circular cue instructing them to prepare to inhale. The circle then blinks and begins expanding, with the participant instructed to sharply inhale with the expansion of the circle. The participant then exhales, and a second similarly guided breath is conducted. This procedure of pacing the participant’s breathing via visual cues is a novel feature of the RRST, and is intended to reduce intra- and inter-subject variance in respiratory effort. Following the two breaths, the participant indicates by keyboard press whether the first or second breath was more difficult. **B)** Sample single subject data, illustrating the psychophysical procedure. On each trial, one of two breaths is the standard, such that the compression wedge is at resting baseline (0% obstruction) with no added resistance, and the other is the stimulus with some level of compression determined by the staircase procedure. The procedure rapidly hones in on a threshold estimate using a Bayesian procedure (psi); in this example the participant threshold of approximately 80% obstruction is found within 20 trials. **C)** Schematic illustrating the design of the automated resistive load apparatus (see Supplementary Fig. 1 – Detailed schematic for details).

corresponds to the Weibull parameter α , denoting the stimulus value at which the probability of responding correctly is 75%. The slope corresponds to the Weibull parameter β , approximating the signal uncertainty. β is proportional to the gradient of the function at stimulus $x = \alpha$ when Ψ is plotted on log units. Varying α produces a left - or rightward shift of the PMF, while varying β changes its steepness, with larger β values resulting in a steeper function. In the context of the RRST, the threshold corresponds to the detection sensitivity, while the slope gives information about their uncertainty, or the response variance with changes in stimulus intensity.

The Weibull function is often modified to include a lapse parameter λ so that the function asymptotes at a performance level of $(1 - \lambda)$, allowing for lapses due to inattention or motor error. The Weibull function Ψ_W relating the predicted proportion correct $P(\text{correct})_W$, is then given by:

$$P(\text{correct})_W = \Psi_W(x; \alpha, \beta, \lambda) = (1 - \lambda) - (5 - \lambda) \exp\left(\left(-\frac{\Delta x}{\alpha}\right)^\beta\right),$$

where x represents a value on the stimulus dimension, α denotes the threshold, β the slope, and λ the lapse rate.

Each participant completed two sessions of the RRST, using different psychometric methods with counterbalanced order to evaluate the internal reliability of the task, namely the QUEST staircase (Watson & Pelli, 1983) and Psi (Kontsevich & Tyler, 1999). Next, we describe these two methods, and our use of them for the RRST. Both QUEST and Psi are Bayesian adaptive psychophysical methods, and use prior information provided by the experimenter (from previous experiments, or from literature) in addition to information from all preceding trials in order to guide the placement of the next stimulus level in order to efficiently obtain estimates of the PMF parameters of interest. Both methods therefore estimate the threshold online (i.e., based on the responses from all preceding trials), and determine the stimulus level to be presented on

the subsequent trial. Whereas QUEST estimates the threshold only, Psi can be used to estimate both threshold and slope, as well as the guess and lapse rates. Here for simplicity we fixed these parameters based on piloted values, although they could be estimated as free parameters in future studies.

The staircase procedures were initiated to run in units of wedge displacement (i.e., proportional to units of motor rotation) ranging between 0 and 18 mm in steps of 1 mm, and these units were also used for all analyzes. To aid interpretability, we transform these values into percentage obstruction for the figures. We emphasize that while the Psi method can select from these 18 preset stimuli, the threshold estimate is obtained by fitting a continuous probability function to response accuracy on multiple presentations of these stimuli. It can thereby take on any value in the specified range, and is not limited to the 18 stimulus levels. Assessment of the standard errors of the threshold and slope parameters for each participant indicated that they generally showed good convergence. This supports that the chosen level of granularity was appropriate for threshold estimates, whereas the slope parameter estimate could benefit from increased number of trials and coarser stimulus granularity. While we chose to explore the full possible stimulus range here for the purpose of validating the RRST, it is possible in future studies to restrict the stimulus range based on the data presented here, and to then use below-millimeter step sizes to yield even greater granularity.

In this study, QUEST was initiated with a uniform prior distribution (range 0–18 mm) over the threshold. Two interleaved QUEST staircases of 30 trials each were run within a single (60 trial long) session in order to reduce trial-to-trial dependencies and to evaluate if the two independent staircases converged on the same threshold. The mean of the posterior distribution was used as the threshold estimate for each staircase (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994).

Subsequently, the mean of the two interleaved staircases was taken as the QUEST threshold.

For every trial, Psi considers the range of possible stimulus intensities to present. For each intensity, Psi computes the probability of a correct versus incorrect response as well as the expected entropy resulting from either response. The stimulus level for which the expected entropy is lowest is then presented on each trial. Following the observer's response, Psi then uses the posterior distribution to recalculate the PMF that best fits the data from all previous trials. Here, Psi was initiated with the following prior parameters: Weibull PMF, α uniform [0,1,18], β uniform $\log(1)$ to $\log(16)$, guess rate $\gamma = 0.5$, and lapse rate $\lambda = 0.02$. The guess and lapse rate parameters are fixed and determine the lower and upper asymptotes, respectively. Individual and group PMFs relating stimulus level (% obstruction) to the probability of a correct response were fit using a Bayesian criterion. The search grid was defined with the same parameters used for the priors for the Psi staircase, with wide distributions for the threshold and slope, and fixed values for the guess and lapse rates. Since Psi estimates two free parameters (slope and threshold) whereas QUEST estimates only threshold, it can generally be assumed that Psi needs more trials to reach full convergence (Kontsevich & Tyler, 1999). As such we included 100 trials for Psi to ensure that we achieve reliable estimates of the threshold as well as an estimate of the slope parameter (note however, that Psi often requires upwards of 100 trials to accurately estimate slope).

2.6. Physiological measures

To determine the relationship between the degree of tube obstruction and effective static resistance produced in the airway, it is important to obtain measures of the mechanical properties of the air circuit. To this end, we measured airflow and differential pressure at a constant airflow of 0.95 L/s, and used these to calculate the resistance at each level of obstruction. Resistance corresponds to the force opposing the flow of air through the circuit, and remains constant for a given pipe diameter (i.e., obstruction level in the RRST). However, the force of inspiration can change the flow of air through the area of compression (the silicone tubing segment of the breathing circuit that is being compressed), leading to changes in measured differential pressure. The relationship between resistance, pressure and flow is given by Ohm's law.

$$R = \frac{\Delta p}{F},$$

Where resistance R is determined by pressure Δp and flow rate F . Measures of differential pressure, and airflow were collected both for human participants performing the task and using a constant flow device. Physiological measures of mouth pressure were collected using an ADInstruments pressure gauge coupled to a bridge amplifier, and inspiratory flow was measured using an ADInstruments spirometer with a 300 L flow head. Physiological traces were recorded via a 16-channel PowerLab (ADInstruments, New Zealand), and analyzed within LabChart (version 8; ADInstruments).

The constant airflow was generated using a custom-built device that blew a steady stream of room-temperature air through the breathing circuit. Since the effective pressure depends on the characteristics (e.g., depth, strength, acceleration) of each inhale, the constant flow data was used to analyze the resistance generated at different levels of obstruction without breath-to-breath variability. This also allowed us to visualize the variance and drift in resistance across trials. The constant flow device was attached to the circuit instead of the mouthpiece section. LabChart software was used to record the pressure and flow data while the RRST device moved in increments between positions corresponding to 0% and 100% obstruction. Each obstruction measure was interspersed with a 0% obstruction position, such that the device moved in a similar manner to what is employed during experimental tests. Therefore, each of 17 obstruction steps were interleaved with the device

returning to 0% obstruction, and the process was repeated three times in total. In future work, we will record physiological data as described here alongside the task in order to ensure that only trials performed correctly are included in analysis. The GitHub repository will then be updated with functions for automatically detecting and excluding missed trials.

2.7. Comparison to Filter Detection Task

The data for the comparison to the Filter Detection Task (FDT, Harrison, Garfinkel, et al., 2021) was collected at a second site (University of Otago, New Zealand, Ethics number 20/CEN/168; Approval given by the New Zealand Health and Disability Ethics Committee). Fifteen participants (12 females, average age of 23.1 years \pm 5.5 (SD)) completed a session of the RRST and FDT, while differential pressure and flow generated in the respiratory circuit were recorded. Briefly, in the FDT between 0 and 12 spirometry filter head filters (each providing a resistance load of 0.42 cm H₂O/L s⁻¹) are attached to a circuit, which the participant inhales through. The baseline resistance consisting of an empty/dummy filter was applied either in the first interval or second interval of three breaths, with the resistance load applied in the other interval of three breaths, for a total of six breaths in each trial. Calibration trials were performed before the task to determine the starting filter number at the participant's perceptual threshold. A total of 60 trials were completed by each participant. On each trial, participants take two sets of three breaths on the circuit, and then decide whether the first or second set of breaths carried a greater resistance (i.e., a 2IFC task). The number of inspiratory resistance filters was determined using a staircase method implemented in MATLAB, and was readjusted throughout the trials to maintain a task accuracy of 65–80%.

2.8. Analysis

Statistical analyses and fitting of PMFs were performed using MATLAB (2020a) and the Palamedes toolbox (Prins & Kingdom, 2018). We used Pearson correlation to evaluate the correlations between perceptual and metacognition variables, and to determine between-method agreement. The effect of trial accuracy on response times and stimulus level was tested using paired-sample Student's *t*-tests assuming unequal variances (*F*-test for equality of variance). Exploratory correlations between threshold and mean aversiveness judgments and subjective ratings of dizziness, breathlessness were tested using Spearman correlation. The figures were created using MATLAB, and the distributions for threshold, slope, response times and stimulus levels were made using raincloud plots (Allen, Poggiali, Whitaker, Marshall, & Kievit, 2019).

We used a signal theoretic approach to evaluate perceptual and metacognitive sensitivity, to reduce the influence of response biases and estimate metacognitive performance independently from perceptual sensitivity. Traditionally in psychophysics, a "Type 1" task denotes those regarding first order or perceptual decisions, e.g. about some stimulus, whereas a "Type 2" task involves a second order metacognitive judgment (Galvin, Podd, Drga, & Whitmore, 2003; Maniscalco & Lau, 2012). For type 1 performance, trials on which the presence of a resistance was identified were coded as correct, for example if the first alternative was chosen, given that the resistance load was indeed applied on the first breath. Conversely, incorrect trials were those on which the standard interval (without resistance load) was chosen.

For type 2 performance, the confidence ratings were first converted from continuous-scale ratings of 0–100 into 4 equally spaced bins. Here we chose 4 intervals because previous research has suggested that this division most closely reflects the subjective granularity for confidence ratings (Sandberg, Timmermans, Overgaard, & Cleeremans, 2010). We used the binning strategy of Fleming (2017, p. 201) which normalizes the confidence quantiles for each participant. This has the benefit of reducing the possibility of empty bins, which confound *m*-ratio estimation. We further checked for any participants with overly extreme confidence ratings (e.g., more than 90% 100% rating) as this can further

confound the binning procedure, resulting in 2 exclusions. These steps all follow established guidelines for best practices in SDT metacognition research (see Fleming & Lau, 2014). Then metacognitive signal theoretic measures can be defined; “hits” are marked as trials on which the type 1 response was correct, and the confidence rating was high, whereas “misses” are trials on which the type 1 response was incorrect but the confidence was high (Fleming & Lau, 2014; Maniscalco & Lau, 2012). In this way the area under the type 2 receiver operating characteristic (aROC) curve can be determined, which estimates the metacognitive sensitivity while accounting for a bias to over or under confidence. The type 1 (e.g., d') and 2 parameters (aROC and mRatio) were calculated using the HMeta-d toolbox (Fleming, 2017). The mRatio here refers to the ratio of $d'/\text{meta-}d'$, where meta- d' is a d' estimate fit only to the confidence ratings. We used the area under the type 2 receiver operating characteristic (aROC) as the main metacognitive measure here because performance was tightly controlled at around 80% by the Psi staircase (Fleming & Lau, 2014; Maniscalco & Lau, 2012).

For analysis of the physiological measures recorded during constant airflow application, pressure and flow measures were averaged across a 2 s interval at each level of obstruction for each of three experimental runs. Resistance was calculated by dividing the change in pressure by the average flow at each obstruction step. Percentage obstruction was then plotted against the measured resistance values. Additionally, as an exponential relationship was observed between percentage obstruction and resistance, the resistance values were log-transformed and then re-plotted against percentage obstruction, where a linear relationship could then be quantified.

For the comparison to the Filter Detection Task, we first related average physiological measures of pressure and resistance, as well as

type 1 and type 2 performance variables between the two tasks using paired-samples t-tests. We further conducted exploratory Pearson correlational analyses inter-relating FDT and RRST average resistance and pressure, as well as subjective confidence ratings and metacognition scores (aROC). All analyses were Bonferroni corrected for multiple comparisons. For raw confidence analyses, RRST values were first divided by 10 to equate them to the 10 - point FDT scale. These analyses can be reproduced using the FDTvsRRST.jasp JASP data file found on the project GitHub: <https://github.com/embodied-computation-group/RespiroceptionMethodsPaper/blob/main/suppAnalyses/FDTvsRRST.jasp>.

3. Results

3.1. Psychometric results

We first evaluated the staircase convergence and threshold estimates for the QUEST and Psi staircase methods. As similar results were observed from the two methods, the following results are presented for the Psi staircase session only (see Section 3.1 – Performance for Psi and QUEST staircases for details). Individual and group PMFs relating stimulus level (% obstruction) to probability of a correct response were fitted using a Bayesian criterion. The search grid was defined with the same parameters used for the priors for the Psi staircase, with wide distributions for the threshold and slope, and fixed values for the guess and lapse rates (see Section 2.5 above). The mean group threshold (α) was 65.61% obstruction (SD = 8.47%), and the individual threshold values ranged from 56.71% to 89.79%, as shown in Fig. 2 C and D. Slope estimates also exhibited large inter-subject variability, with a mean of 7.80, standard deviation of 2.61, and range from 3.12 to 12.37.

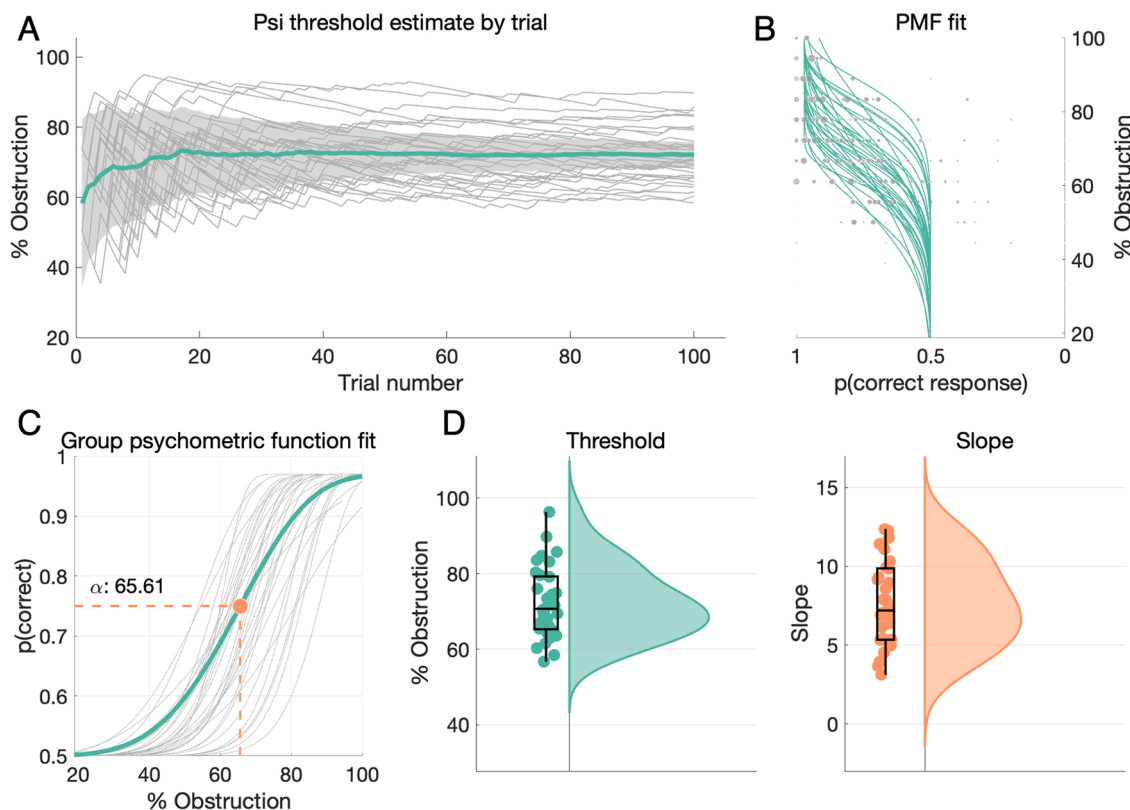


Fig. 2. Psychometric task results. A) Plot depicting trial-by-trial Psi threshold estimates for all participants. Light gray lines depict individual stimulus traces indicating the % of tube obstruction on the stimulus breath for each trial; the thick green line represents grand mean stimulus on each trial \pm SEM. In general, threshold estimates stabilize around trials 20–40 for all participants. B) PMF fits for all subjects. The green lines depict individuals' PMF fits, and gray points show stimulus levels presented, where the dot size indicates the number of times presented. C) Grand mean psychometric fit (green) overlaid on individual PMF fits (gray), demonstrating that average respiratory thresholds are around 66% airway obstruction, with substantive inter-individual variance around this value. D) Raincloud plots (Allen, Poggiali, et al., 2019) depicting individual threshold (green) and slope (orange) estimates for all subjects.

3.2. Type 1 performance – perception

The RRST method estimates subject thresholds by presenting stimuli at various points around the estimated decision function. As a validation of the procedure and test of the robustness of the task to the choice of the psychometric staircase, we examined the relationship between choice accuracy, stimulus intensity, and reaction time. If estimated thresholds are reasonably well estimated, then we would expect responses to be significantly faster for correct versus error trials, and to observe that correct versus incorrect trials are generally associated with higher stimulus intensity levels. Indeed, when examining the effect of accuracy on reaction times (RT) for Psi staircase sessions, we observe that RTs are significantly lower for correct versus incorrect trials, $t(42.25) = -3.34$, $p < 0.01$ (Fig. 3 A). Stimulus intensity levels on correct trials were significantly greater than on incorrect trials, $t(60) = 4.66$, $p < 0.01$ (Fig. 3 B). These results indicate that near-threshold signal-present stimuli (i.e., resistances just above and below a participants' threshold) modulate both processing time and response accuracy as expected.

Similar results are obtained for the Quest staircase, with lower RTs for correct compared to incorrect trials $t(49.8) = -3.35$, $p < 0.01$ (Fig. 3 C), and higher stimulus intensities on correct versus incorrect trials, $t(60) = 3.82$, $p < 0.01$ (Fig. 3 D). This similarity in the results obtained using the Psi and QUEST staircases further underscores the robustness of the task to the chosen staircase method, as both the cognitive difficulty and speed-accuracy tradeoff for peri-threshold stimuli are highly similar between the two methods.

3.3. Task reliability

To determine the number of trials needed for staircase convergence, we analyzed measures of how far-removed current staircase estimates were from the final threshold estimate at the end of the session. On each trial, the Psi method determines a standard error (SE) for each PMF

parameter it is estimating. The progression of SE estimates for the threshold and slope over trials are shown in Fig. 4 A and B. Visual inspection shows that threshold SEs shrink the most up to around trial 50, whereas slope SEs show a decreasing trend even at trial 100. This suggests that over 100 trials may be necessary to accurately estimate the slope parameter of the PMF. For QUEST, we calculated the difference between the stimulus (% obstruction) level presented on each trial and the final threshold estimate (see Fig. 4 C). This measure asymptotes around trial 20, however care should be taken in the interpretation as this measure is not directly comparable to the SE estimates from Psi above.

The within-subject reliability of the thresholds obtained using the RRST was assessed by a Pearson correlation on the estimates obtained from the Psi and QUEST staircase methods. Thresholds on the two interleaved sessions were found to be strongly correlated, $r(31) = 0.88$, $p < 0.001$ indicating high consistency between the two methods (see Fig. 4 D). Threshold estimates obtained using the two staircase methods were found to not differ significantly from one another ($t(55) = -0.97$, $p = 0.34$), further demonstrating good robustness of the RRST to the choice in psychometric procedure.

3.4. Type 2 performance – metacognition

To assess respiration metacognition, we asked participants to rate their confidence in their decision on every trial. Here we describe the results of the metacognition analysis, based on trials from the Psi session only. Confidence ratings were generally higher for easier trials (i.e., trials on which the obstruction was greater, see Fig. 5 A), and there was a dissociation between ratings on correct and incorrect trials, such that confidence was lower on incorrect trials and higher on correct trials (see Fig. 5 B). To estimate individuals' metacognitive ability, we calculated the area under the type 2 ROC curves (see Fig. 5 C, the aROC is the area between the identity diagonal and the ROC curve), which corresponds to

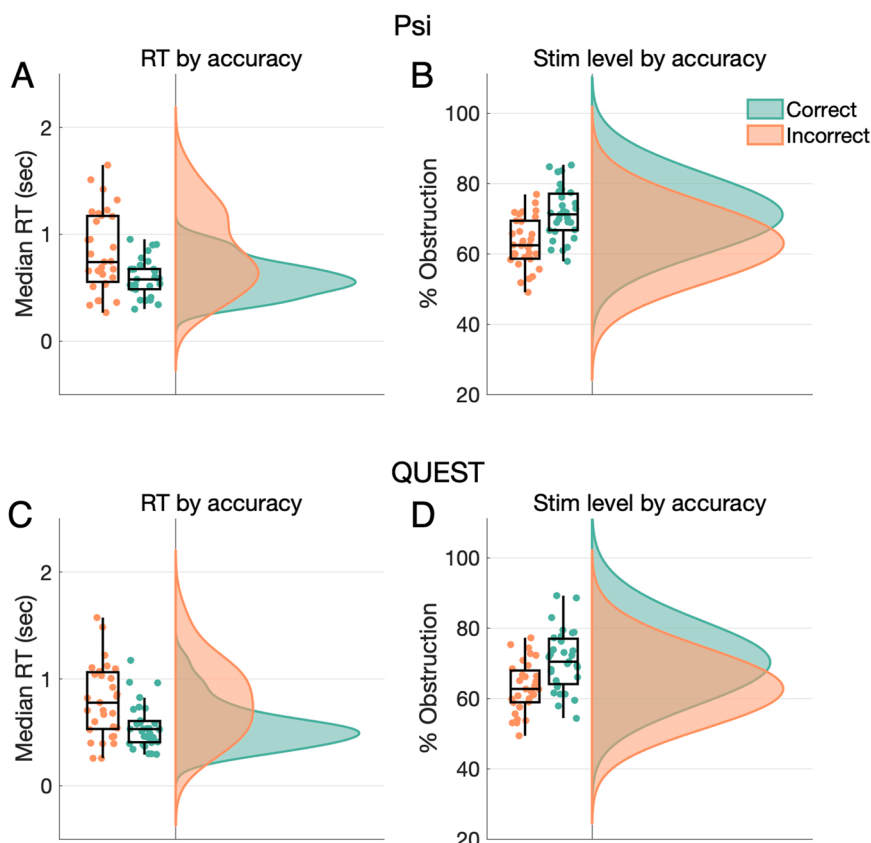


Fig. 3. Type 1 performance on Psi & QUEST methods. Raincloud plots of reaction times (RT, left panel) and stimulus level (right panel) by accuracy (correct vs. incorrect) for the Psi (upper panel) and QUEST (lower panel) staircase methods. **A) and C)** Median RTs presented for each subject, for correct (green) and incorrect (orange) trials showing that RTs on correct trials are lower than on incorrect trials. **B) and D)** Average stimulus levels presented for each subject, for correct and incorrect trials, showing that stimuli were higher (i.e., easier) on correct trials.

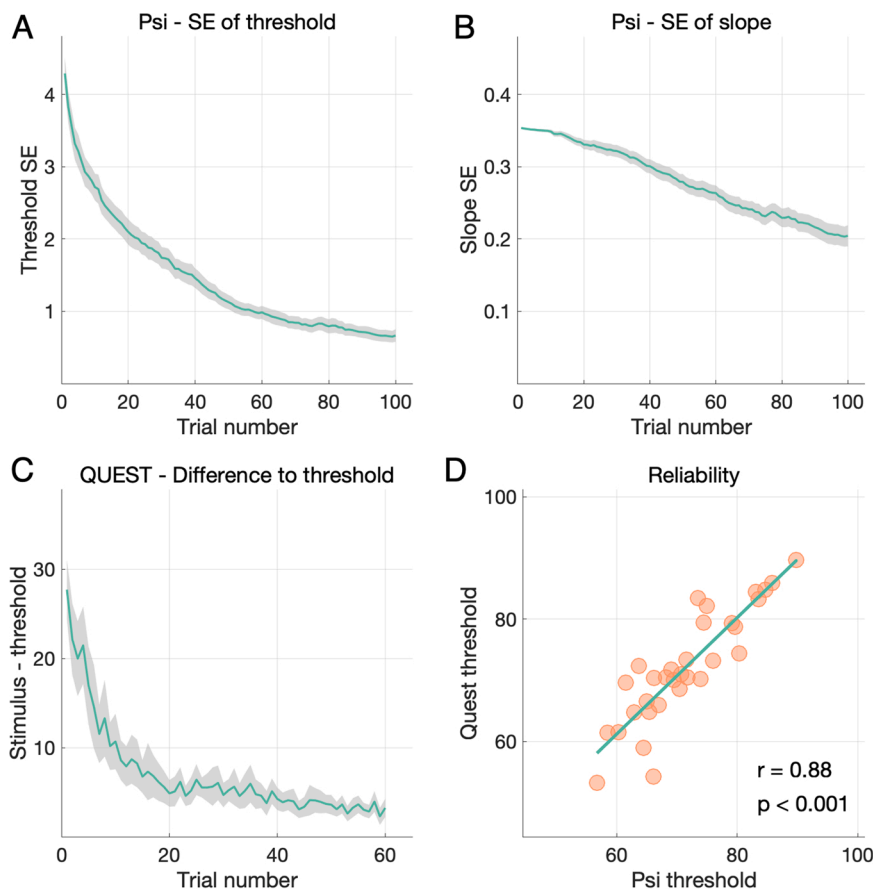


Fig. 4. Staircase convergence and task reliability. **A)** Standard errors of the threshold estimate by trial number obtained with Psi indicate that reliable threshold estimates are derived within 20–50 trials. **B)** Standard errors of the slope estimate by trial number show that slope uncertainty drops linearly as a function of trials, indicating that slope estimates may benefit from higher trial numbers and/or hierarchical modeling. **C)** The difference to the final threshold estimates by trial number obtained with QUEST indicates that threshold estimates also converge within 20–50 trials. **D)** Correlation plot across the two (counter-balanced) QUEST and Psi threshold estimates indicates a high within-subject reliability of respiroceptive thresholds, regardless of estimation technique.

the ability to associate confidence to perceptual performance. While the Psi staircase held perceptual performance constant at a level of 76–84% task accuracy, metacognitive ability (aROC) varied substantially, ranging from 0.48 to 0.85 (mean = 0.70, SD = 0.09, see Fig. 5 D). Furthermore, the type 1 and type 2 performance measures were not correlated ($r(31) = -0.03$, $p = 0.86$). These results highlight the unique ability of the RRST approach to titrate respiroceptive performance by quickly identifying the approximate threshold value (usually within 20 trials), and thereafter presenting stimuli around this level. This approach results in a tightly controlled accuracy of around 80%, which is typically desired for estimates of metacognitive ability (e.g., Guggenmos, 2021).

3.5. Task tolerance

We wanted to assess whether participants found the task unpleasant or aversive, and whether this changed over the course of the testing session. Fig. 6 A shows the aversiveness ratings over 10 blocks (6 blocks of Psi and 4 blocks of QUEST), or 45 min, of testing. Mean displeasure ratings were 24.69% (SD across the means = 1.61), indicating that the task was on average mildly aversive and that this remained constant throughout the testing session. Average dizziness/light-headedness (18.92%), breathlessness (29.08%), and severity of asthma symptoms (2.22%) are shown in Fig. 6 B, along with data points indicating individual ratings for each session. Mean ratings of aversiveness were found to correlate strongly with those of dizziness, but not with other subjective ratings, or with perceptual or metacognitive task parameters (see Supplementary Fig. 2 for details).

3.6. Inspiratory resistance and pressure

The precise measurement of respiratory resistance thresholds is

dependent on both reproducible resistive stimuli delivered by the apparatus, as well as reliable effective resistance values despite individual variability in breathing dynamics. To evaluate resistances produced by the RRST during real-life testing, we measured pressure and flow throughout sessions of the task.

In Fig. 7, we show the measured resistance over levels of obstruction. The relationship between obstruction and resistance is exponential, with resistance increasing dramatically from about 70% obstruction (Fig. 7A). The relationship between the natural log of resistance and obstruction is linear (Fig. 7 B, slope = 0.44, $p < 0.0001$, $R^2 = 0.88$), with some variability at extremely low circuit obstruction and deviation from linearity at extremely high circuit obstruction. Variability at low circuit obstruction levels is likely due to measurement-associated error at very low levels of resistance.

Since we observed that the resistance produced on each step of airway obstruction is well fit by a log-linear function (Section 3.6), the results here are represented on a log scale as well. Fig. 8 shows resistance (A) and pressure (B) measurements on individual trials for 15 sessions of the RRST completed by different participants. During natural performance of the task, the logarithm of both resistance and pressure is found to increase linearly with obstruction level.

3.7. Comparison to Filter Detection Task

We wanted to compare performance on the RRST to an existing measure of respiratory interoception, the Filter Detection Task (Harrison, Garfinkel, et al., 2021), to evaluate whether psychometric estimates (i.e., thresholds), metacognitive variables and accuracy relate between testing methods. Summary statistics describing physiological variables, accuracy, and metacognition are described below in Table 1:

In general, average performance parameters across the two tasks were highly similar, with the exception of increased average resistance

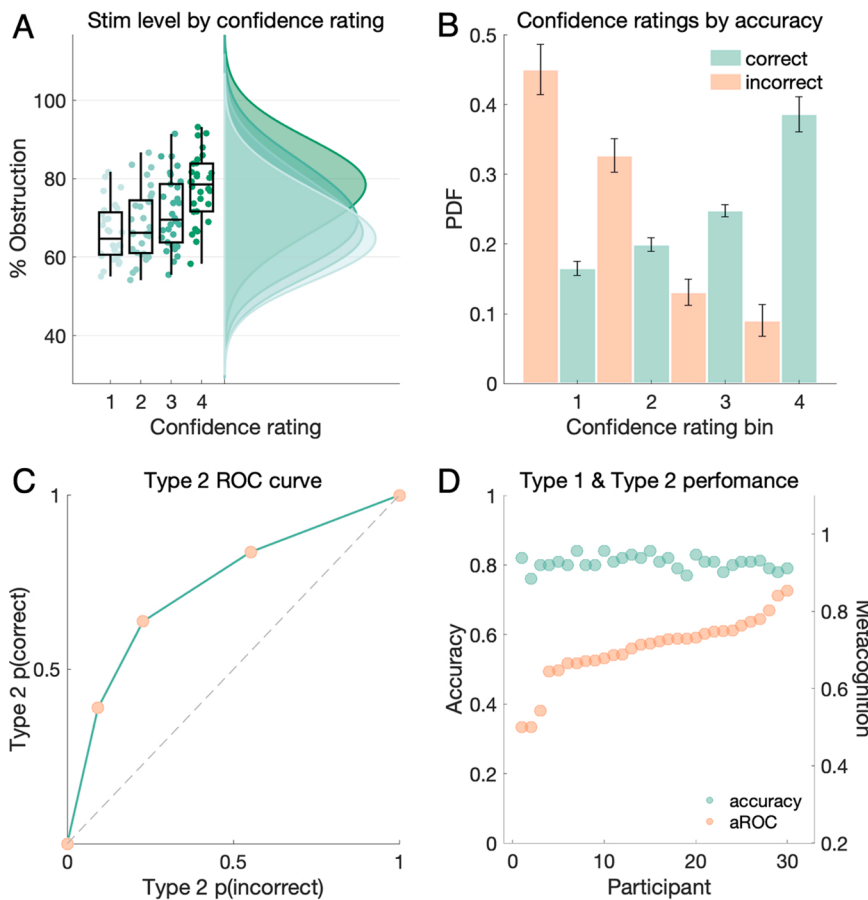


Fig. 5. Type 2 performance. A) Raincloud plot showing how stimulus intensity (i.e., the percentage obstruction) covaries with each confidence bin input to the metacognition model, showing the subjective correspondence between stimulus intensity and confidence. B) Histogram of binned confidence ratings for correct (green) and incorrect (orange) trials. Generally, participants showed high metacognitive sensitivity, as seen in the dissociation between the correct and incorrect trial ratings. C) Type 2 ROC curve, averaged over participants, showing good respiratory type 2 performance. D) Type 1 (accuracy, green) and type 2 (aROC, orange) performance, sorted by each participant's aROC. Participants show substantial variations in metacognition while type 1 accuracy is held relatively constant by the Psi staircase procedure.

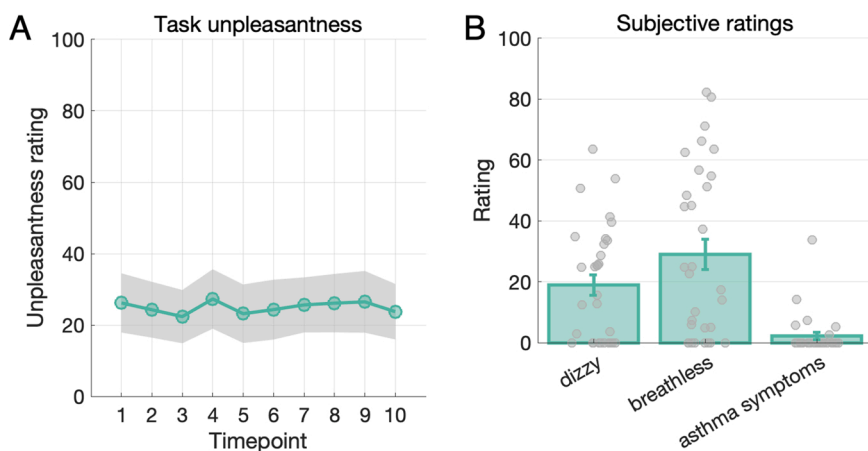


Fig. 6. Task tolerance & subjective ratings. A) Plot depicting unpleasantness ratings across all 10 blocks (timepoints) of testing. Mean unpleasantness ratings after each block are shown in green, and the shaded gray area represents the standard error of the mean (SEM). Each block comprised 20 trials, for a total testing time of approximately 45 min. Participants on average reported 24.96% stimulus unpleasantness, which remained stable throughout the testing period. This indicates that the stimuli were mildly unpleasant and that extended testing time did not increase task adversity within these limits. B) Plot depicting mean dizziness, breathlessness and asthma symptoms across participants. Bar height represents mean ratings, error bars denote SEM, and gray circles show individual participants' ratings. In general, participants showed low levels of these adverse effects following 1 full hour of testing, indicating good tolerability of the task.

and accuracy in the RRST. No significant correlations were found between perceptual or metacognitive sensitivity, average resistance, or pressure on the two tasks (all $p > 0.05$). We did observe a significant correlation between subjective confidence in the two tasks, Pearson's $r(14) = 0.75, p = 0.002$.

While the RRST by design holds accuracy (hit rate) at 79% ($\pm 1.5\%$ SD), performance varies more widely in the FDT (mean accuracy 71% $\pm 9\%$ SD), with 6 times higher standard deviation of accuracy on the FDT as compared to the RRST (Fig. 9 A). A tight control of accuracy is desirable for analyses interested in type 2 performance, as it allows for the estimation of metacognitive variables, independently of contamination by differences in type 1 performance (Guggenmos, 2021; Xue

et al., 2021). Mean confidence ratings are found to correlate between the RRST and FDT (Fig. 9 B). This is in agreement with previous results showing that confidence ratings relate highly across exteroceptive modalities (e.g., Mazancieux, Fleming, Souchay, & Moulin, 2018).

4. Discussion

We present a novel psychophysical method and custom experimental apparatus for the streamlined and efficient estimation of respiratory sensitivity. To achieve this, we used accessible and low-cost 3D printing methods to develop an apparatus which can flexibly, safely, and reliably deliver inspiratory or expiratory loads entirely through a stimulus PC

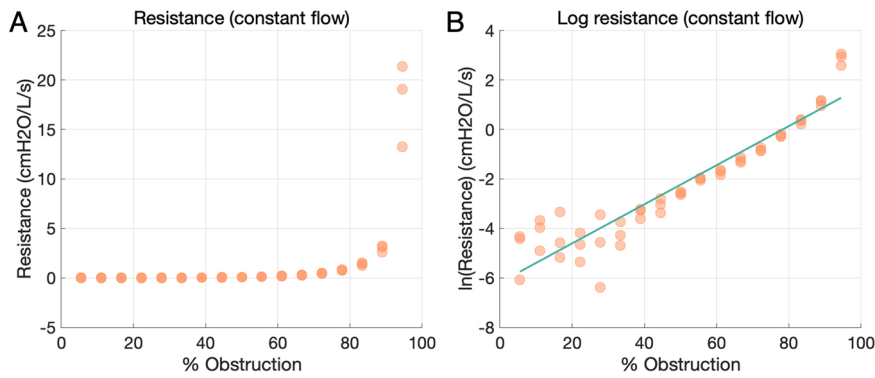


Fig. 7. Inspiratory resistance as a function of obstruction under constant airflow. A) Effective resistance is plotted in terms of the percentage of airway obstruction, based on 3 trials per obstruction level. Resistance is seen to increase exponentially with obstruction from a value of about 70%. B) Log resistance as a function of airway obstruction. The magnitude of resistance is well fitted as a log-linear function of airway obstruction.

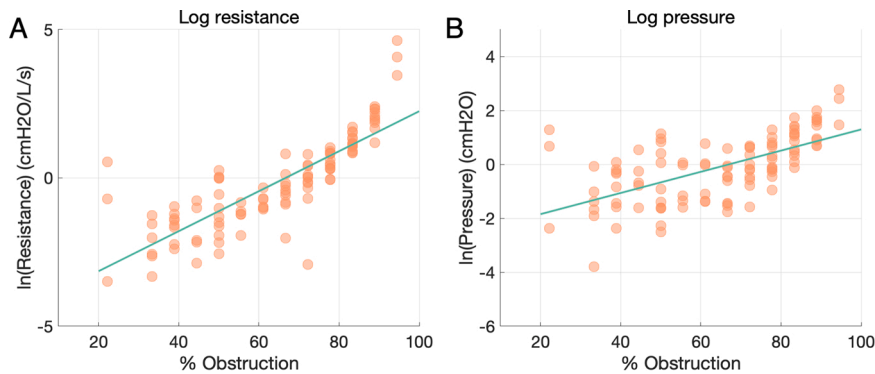


Fig. 8. RRST resistance and pressure as a function of obstruction for participants. A) Natural log resistance ($\text{cm H}_2\text{O/L s}^{-1}$) as a function of airway obstruction ($r = 0.82, p < 0.0001$). B) Natural log pressure ($\text{cm H}_2\text{O}$) as a function of airway obstruction. Data points are participant means across trials at each level of obstruction ($r = 0.62, p < 0.0001$).

Table 1
Means and variances for key FDT and RRST measures.

	FDT (SD)	RRST (SD)	t (13)	p	Cohen's d
Pressure (cmH_2O)	1.21 (0.66)	2.75 (2.27)	-2.386	0.033	-0.638
Resistance ($\text{cmH}_2\text{O/L/s}$)	1.82 (1.21)	7.92 (13.16)	-1.766	0.101	-0.472
Accuracy	71% (9%)	79% (1.5%)	-3.196	0.007	-0.854
Confidence	5.2 (1.6)	5.7 (1.4)	-1.161	0.267	-0.310
Metacognition (aROC)	0.63 (0.07)	0.61 (0.05)	0.759	0.462	0.203

Note. Student's t-test. FDT = Filter Detection Task, RRST = Respiratory Resistance Sensitivity Task. Normality was determined via the Shapiro-Wilk test, which was not significant for any variable pair except for the FDT and RRST resistance comparison ($W = 0.631, p < 0.001$). Follow-up analysis with a Wilcoxon signed-rank test gave the same qualitative results for resistance ($W = 25, p = 0.091, \text{Rank-biserial correlation} = -0.524$).

and sanitary circuit. By expanding well-established Bayesian psychometric approaches, we further developed a novel psychophysical task for estimating respiroceptive thresholds, the respiratory resistance sensitivity task (RRST). Our results show that the RRST provides a reliable measure of respiroceptive thresholds in under 30 min, as assessed by a same-day test-retest, and is mildly aversive for participants. Further, by enabling the rapid collection of many near-threshold trials in a 2IFC design, the RRST can flexibly dissociate interoceptive sensitivity, precision, and bias, and is well suited for model-based research in interoceptive perceptual decision-making over both perceptual and

metacognitive dimensions (see e.g., Allen, Levy, Parr, & Friston, 2019; Allen et al., 2021; Unal et al., 2021). Confidence ratings were collected after each trial in order to allow for estimates of metacognitive insight over different levels of stimulus difficulty. These trial-by-trial estimates allow for measures of metacognitive bias, sensitivity and efficiency - variables which may reflect meaningful individual variations in interoception-related metacognitive insight. Here we focused on metacognitive ratings and only collected measures of subjective affect and negative symptoms at the end of each staircase session to determine the extent of these symptoms as a result of exposure to RRST resistive loads. Future studies could benefit from also tracking these subjective variables on a trial by trial basis, to better understand how the evolution of subjective respiratory sensations relates to objective detection performance.

4.1. The perception of breathing

A common issue in the development of new tasks measuring interoceptive ability is the inability to dissociate sensitivity from response bias. This is particularly salient for example, in studies utilizing either purely subjective measures, or in the cardiac domain where the ability of existing tasks to dissociate sensitivity and bias has been questioned (Brener & Ring, 2016). This dissociation can be made by controlling the information present in the stimulus (i.e., the signal-to-noise ratio) in a precise way and quantifying the change in sensitivity resulting from increases in signal. However, the signal in most interoceptive sensory modalities, such as cardiac or gastric sensations, are not directly amenable to experimental control in the absence of invasive procedures such as distention of the esophagus, stomach and rectum (Jones, Hoffman, Shah, Patel, & Ebert, 2003; van Dyck et al., 2016; Yuan et al.,

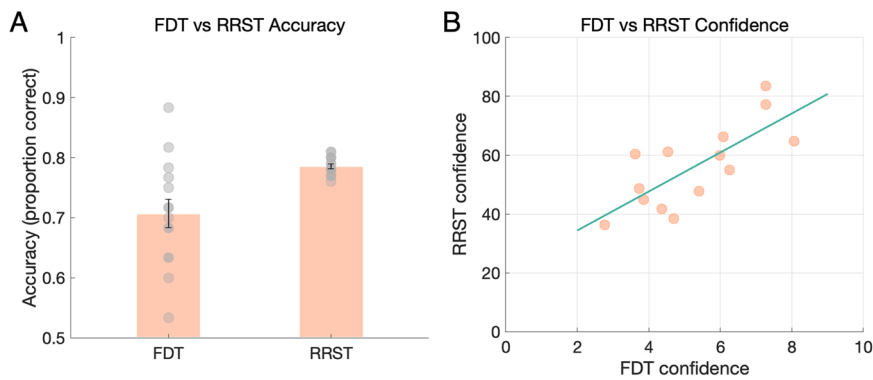


Fig. 9. : Comparison of FDT and RRST type 1 and type 2 variables. **A)** The RRST achieves a greater degree of control over accuracy, facilitating the comparison of metacognitive variables. Points denote overall task accuracies for each of 15 participants, bar height indicates mean accuracy across participants, and error bars indicate standard deviation. **B)** Confidence ratings correlate between FDT and RRST. Points represent mean confidence scores for each of 15 participants. Line denotes best fit of a linear regression ($R = 0.75$, $p = 0.002$).

2003), which require medical expertise and cannot readily be administered in healthy participants. While such measures are very important to brain-body research, their inherent practical limitations render them difficult to administer in certain contexts and populations. This limits studies of interoception, as well-established techniques for studying perception (e.g., psychophysics, signal detection theory) become difficult to apply in the absence of control over the physical signal. Further, human introspection is rife with cognitive and perceptual biases, meaning that purely subjective methods cannot identify the channel capacity or true sensitivity of the interoceptive system. While subjective biases themselves are of interest, this is a fundamental limitation for interoception research, where identifying the objective sensitivity to interoceptive stimuli may prove valuable both for unraveling the fundamental neural circuits underlying interoceptive perception, and for identifying maladaptive interoceptive sensitivity, which could be wholly orthogonal to interoceptive belief. As such, predominant approaches to measuring interoceptive perception either rely on difficult to interpret measures such as heart-beat counting (Brener & Ring, 2016; Schandy, 1981), or use invasive procedures such as gastric intubation (Stephan et al., 2003).

The respiratory domain presents interoception research both with unique opportunities and challenges. Respiratory sensation is itself a complex mix of interoceptive and exteroceptive sensory-motor and chemosensitive channels. During normal ventilation, air is mechanically pumped into the lungs by the distension and contraction of the diaphragm, which pushes air in and out of the lungs much like a blacksmith's bellows. Throughout the inspiratory-expiratory cycle, the passage of air through the mouth, nose, and upper airway is communicated by thermosensory and tactile receptors, and the rhythmic expansion and contraction of the lungs is associated with the activation of stretch receptors located throughout the diaphragm, chest wall, and surrounding organs (Schroijen, Davenport, Bergh, & Diest, 2020). This rhythmic information is communicated to the medullary and pontine brainstem nuclei, the somatosensory cortex, and higher-order structures such as the insula (Davenport & Vovk, 2009; Schroijen et al., 2020; von Leupoldt et al., 2008). Additionally, deviations in the concentrations of blood gasses (carbon dioxide and oxygen) are communicated to the viscerosensory nuclei of the brain by chemosensory pathways, generating the interoceptive sensation of "air hunger" (Guz, 1997; Manning & Schwartzstein, 1995).

The complex physiological and neuroanatomical pathways subscribed by respiration are therefore fundamentally multi-modal, and a mixture of interoceptive and exteroceptive sensations arising from the airway, blood, and skeletal-muscular system. Alterations in either inspiratory or expiratory resistance can be expected to interact with this system at multiple levels: through the sensation of pressure at the lips, mouth, and upper airway, through the sensation of increased diaphragmatic effort, and in the case of sustained (i.e., non-discrete) loads, through the sensation of (likely) increased carbon dioxide in the bloodstream. A measure aiming at characterizing respiration

therefore needs to be flexible and fine-grained enough to resolve the sensations produced by these convergent breathing-related signals.

4.2. Benefits of the RRST

A primary benefit of the RRST approach is that resistive stimuli can be flexibly manipulated to identify the minimal stimulus a participant can reliably discriminate. Ecologically speaking, the process of detecting respiratory loads operationalized by the RRST and similar procedures is not unlike that which accompanies a serious chest cold, where we might notice that our breathing has become more labored due to the partial obstruction of the airway by mucous or inflammatory swelling. Respiratory resistance tasks therefore present an ecologically valid means by which to dissociate the objective sensitivity of airway monitoring.

However, as others have noted (Miller & Davenport, 2015), increasing the effort associated with respiration is for most individuals an inherently aversive process. This, coupled with the coarse granularity of most previous resistance-based tasks, meant that the estimation of resistive thresholds was a slow, painstaking process requiring many trials. By leveraging our unique apparatus, the RRST is able to automatically adjust the intensity of the respiratory stimulus (e.g., the amount of airway resistance) in a highly granular manner. Compared to previous tasks using inspiratory resistive loads, which typically had between 4 and 7 stimulus levels (Garfinkel et al., 2016; Harrison, Garfinkel, et al., 2021), the RRST as presented here offers a 2-to-3-fold increase in granularity. This opens the possibility to use Bayesian or other adaptive psychophysical techniques to optimize the delivered stimulus on each trial, greatly increasing the speed and precision with which respiratory psychophysical parameters such as threshold and precision can be estimated. An adaptive psychophysical staircase such as Psi or QUEST estimates the threshold within about 20 trials, and thereafter presents stimuli very close to the threshold for the remainder of the session; this minimizes exposure to high resistance loads improving clinical utility. Further, once threshold has been identified, the experimenter can flexibly adapt stimuli to each participant, delivering fine-grained control over task difficulty (i.e., error rates) and ensuring that task stimuli are equivalent across participants while controlling for variations in sensitivity (see Fig. 9). This, together with the 2IFC design of the RRST, is an important feature in particular for the estimation and modeling of respiratory metacognition where uncontrolled type 1 error rates can strongly bias estimates of metacognitive sensitivity and/or efficiency (Fleming & Lau, 2014; Guggenmos, 2021). Further, by maintaining fine-grained control over stimulus magnitude, the RRST ensures that excessively large resistive loads can be avoided and keeps total data acquisition times for threshold and metacognitive estimates under 30 min per participant, limiting the aversiveness of the task. This is especially important for clinical populations, such as persons with asthma or anxiety, where previous methods are challenging to apply in the very populations within whom respiration abnormalities are likely to be of the greatest interest. A further benefit of the RRST in this

that it provides a non-invasive means of dissociating sensitivity and bias by precisely and flexibly controlling the signal. Previous work using the FDT has already shown that perceptual sensitivity and metacognitive measures differ with anxiety scores (Harrison, Köchli, et al., 2021). Future work will then be able to investigate how different sources of sensory and cognitive input contribute to RRST parameters, for example by combining within-subject manipulations (e.g., topical anesthesia of the chest wall) with the task.

By use of a 2IFC design, our task is optimized for the measure of respiratory perceptual sensitivity. Single alternative (1IFC, or yes/no) tasks, in which a single stimulus is presented and the observer decides whether a signal was present or absent, are criterion-dependent. This reflects the observation that the participant needs to adopt a criterion, or a stimulus level at which they will begin to respond “yes”. Due to this dependency, 1IFC tasks characterize response bias as the propensity to respond that a signal is present, reflecting either an unconscious bias, or a response strategy (Morgan et al., 2012). In a 2IFC design such as that used in the RRST, this response bias manifests as a tendency to respond “first interval” or “second interval” more often. In both of these cases, the use of the signal detection theoretic d' can allow for the dissociation of sensitivity from response bias.

The RRST brings several further practical benefits. The introduction of a visual pacing aid contributes to respiratory entrainment, assisting participants to pace their inhalations in a consistent manner. This likely contributes to the good same-day retest reliability we observed here (see Fig. 4D), as well as to the relatively low levels of aversiveness reported by the participants. By encouraging quick, shallow and even breaths, most of the participants in the study were able to avoid hyperventilation and feelings of breathlessness, even following 30 min of testing. Finally, the components necessary in order to build the RRST apparatus are inexpensive and easily accessible through online vendors, and the shell can be 3D printed using standard commercial 3D printers, such as those available at the maker labs found at many universities. We make all components, schematics and software available on GitHub, with the intention that others can build the device independently.

4.3. Limitations

The task presented here evaluates inspiratory resistance sensitivity across cognitive levels, by measuring both type 1 perceptual performance (i.e., resistance sensitivity as measured by % airway obstruction at threshold, and psychometric function slope parameter) and type 2 metacognitive performance (measured by mRatio, or aROC). It is possible that variation in breathing patterns between participants results in substantial variations in airflow and therefore, in different values of pressure, at the same airway obstruction level (see Fig. 8). Here we introduced a number of features to mitigate the effect of variations in respiratory pattern. First, we designed a visual respiratory entrainment stimulus to standardize depth of inspirations between participants. Second, each trial of the task requires two short inhalations in quick succession, thereby regularizing respiratory frequency. Nevertheless, the variation in respiratory patterns, especially within clinical populations, leads us to recommend caution when interpreting type 1 results from the RRST, especially when the perceptual sensitivity is the primary variable of interest. In these cases, it would be recommended to record pressure and flow data during the task, to enable estimates of the psychometric parameters as a function of inspiratory pressure as a more nuanced measure, rather than of % airway obstruction, since it accounts for the inter- and intra- individual variability of air flow and likely reflects the effective stimulus intensity on each trial more accurately.

A temporal 2AFC design was chosen for this study, as it was best suited for the purpose of estimating perceptual sensitivity to resistive inspiratory loads. It should be noted however that there may be an uneven distribution of attention across the two stimulus intervals,

and that this design poses a working memory load such that the second stimulus may be easier to recall than the first at the time when

the response is made. An advantage of the adaptive staircase approach used here is that trial-wise differences in these effects on choice accuracy can be controlled by the thresholding procedure, because it automatically adjusts the stimulus difficulty throughout the session. If a participant's attention begins to falter after 30 trials, for example, they will likely make more incorrect responses on stimulus levels that would previously have been detected reliably, and the staircase procedure will adjust the threshold estimate up as a result of this. This is particularly beneficial for the estimation of metacognitive sensitivity within and across groups, where uncontrolled error rates can be a major confound (Fleming & Lau, 2014).

A further potential concern is the tolerance to the task, as prolonged exposure to inspiratory resistances was previously shown to be aversive to participants. Here, we regularly recorded subjective ratings of task aversiveness, as well as subjective ratings of breathlessness, dizziness and asthma symptoms at the end of each session. While a small subset of participants did find the task aversive, our results show that the RRST is well tolerated by most participants. Furthermore, individual levels of self-reported aversiveness were found to be uncorrelated with perceptual and metacognitive parameters, indicating that these parameters are orthogonal with respect to task aversiveness.

Future work should examine the test-retest reliability of the task over several days or weeks in order to better characterize the consistency of results and constancy in respiratory interoception. While our results from two sessions performed over the course of an hour using different psychophysical staircase methods are reassuring, the conclusions we can draw from this test are limited as we only examined the immediate task reliability.

4.4. Conclusion

Here we have presented the RRST, a novel method for measuring respiratory interoception-related factors across perceptual and metacognitive levels. By leveraging a custom-designed apparatus and Bayesian adaptive psychophysical algorithm, the RRST can reliably estimate threshold, slope and signal theoretic measures alongside metacognitive factors in under 30 min of testing. This short testing time combined with minimal aversiveness, opens up the possibility of studying respiratory interoception in clinical populations.

Acknowledgements

NN, MB, NB, NL, and MA are supported by a Lundbeckfonden Fellowship (under Grant [R272-2017-4345]). MA is supported by the AIAS-COFUND II fellowship programme that is supported by the Marie Skłodowska-Curie actions under the European Union's Horizon 2020 (under Grant [754513]), and the Aarhus University Research Foundation. OKH (née Faul) is supported by a Rutherford Discovery Research Fellowship awarded by the Royal Society of New Zealand, and the Department of Psychology at the University of Otago. FF is supported by a European Research Council Grant (ERC-StG-948838).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.biopsycho.2022.108325](https://doi.org/10.1016/j.biopsycho.2022.108325).

References

- Al, E., Iliopoulos, F., Forschack, N., Nierhaus, T., Grund, M., Motyka, P., ... Villringer, A. (2020). Heart-brain interactions shape somatosensory perception and evoked potentials. *Proceedings of the National Academy of Sciences of the United States of America*, 117(19), 10575–10584. <https://doi.org/10.1073/pnas.1915629117>
- Allen, M. (2020). Unravelling the neurobiology of interoceptive inference. *Trends in Cognitive Sciences*, 24(4), 265–266. <https://doi.org/10.1016/j.tics.2020.02.002>
- Allen, M., Legrand, N., Correa, C. M. C., & Fardo, F. (2020). Thinking through prior bodies: Autonomic uncertainty and interoceptive self-inference. *Behavioral and Brain Sciences*, 43. <https://doi.org/10.1017/S0140525x19002899>

- Prins, N., & Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the palamedes toolbox. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01250>
- Quadt, L., Critchley, H. D., & Garfinkel, S. N. (2018). The neurobiology of interoception in health and disease. *Annals of the New York Academy of Sciences*, 1428(1), 112–128. <https://doi.org/10.1111/nyas.13915>
- Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition*, 19(4), 1069–1078. <https://doi.org/10.1016/j.concog.2009.12.013>
- Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology*, 18(4), 483–488. <https://doi.org/10.1111/j.1469-8986.1981.tb02486.x>
- Schroijen, M., Davenport, P. W., Bergh, O. V. den, & Diest, I. V. (2020). The sensation of breathing. In *Cotes' lung function* (pp. 407–422). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118597309.ch22>
- Sherrington, C. (1952). *The integrative action of the nervous system*. CUP Archive. ([https://books.google.dk/books?hl=en&lr=&id=CRU8AAAAIAAJ&oi=fnd&pg=PR7&dq=Sherrington,+C.++\(1952\).+The+integrative+action+of+the+nervous+system&ots=KBRtrndr41&sig=zQmk7Bfcw364FEPZYmaNZezrvEQ&redir_esc=y#v=onepage&q=Sherrington%2C%20C.%20\(1952\).%20The%20integrative%20action%20of%20the%20nervous%20system&f=false](https://books.google.dk/books?hl=en&lr=&id=CRU8AAAAIAAJ&oi=fnd&pg=PR7&dq=Sherrington,+C.++(1952).+The+integrative+action+of+the+nervous+system&ots=KBRtrndr41&sig=zQmk7Bfcw364FEPZYmaNZezrvEQ&redir_esc=y#v=onepage&q=Sherrington%2C%20C.%20(1952).%20The%20integrative%20action%20of%20the%20nervous%20system&f=false)).
- Stephan, E., Pardo, J. V., Farris, P. L., Hartman, B. K., Kim, S. W., Ivanov, E. H., ... Goodale, R. L. (2003). Functional neuroimaging of gastric distention. *Journal of Gastrointestinal Surgery*, 7(6), 740–749. [https://doi.org/10.1016/S1091-255X\(03\)00071-4](https://doi.org/10.1016/S1091-255X(03)00071-4)
- Takano, N., & Deguchi, H. (1997). Sensation of breathlessness and respiratory oxygen cost during cycle exercise with and without conscious entrainment of the breathing rhythm. *European Journal of Applied Physiology and Occupational Physiology*. <https://doi.org/10.1007/s004210050238>
- Tiller, J., Pain, M., & Biddle, N. (1987). Anxiety disorder and perception of inspiratory resistive loads. *Chest*, 91(4), 547–551. <https://doi.org/10.1378/chest.91.4.547>
- Tsai, H.-W., Chan, P.-Y., von Leupoldt, A., & Davenport, P. W. (2013). The impact of emotion on the perception of graded magnitudes of respiratory resistive loads. *Biological Psychology*, 93(1), 220–224. <https://doi.org/10.1016/j.biopsycho.2013.02.008>
- Tweeddale, P. M., Rowbottom, I., & McHardy, G. J. R. (1994). Breathing retraining: Effect on anxiety and depression scores in behavioural breathlessness. *Journal of Psychosomatic Research*, 38(1), 11–21. [https://doi.org/10.1016/0022-3999\(94\)90004-3](https://doi.org/10.1016/0022-3999(94)90004-3)
- Unal, O., Eren, O. C., Alkan, G., Petzschner, F. H., Yao, Y., & Stephan, K. E. (2021). Inference on homeostatic belief precision. *Biological Psychology*, 165, Article 108190. <https://doi.org/10.1016/j.biopsycho.2021.108190>
- Vaitl, D. (1996). Interoception. *Biological Psychology*, 42(1), 1–27. [https://doi.org/10.1016/0301-0511\(95\)05144-9](https://doi.org/10.1016/0301-0511(95)05144-9)
- van Dyck, Z., Schulz, A., Blechert, J., Herbert, B. M., Lutz, A. P. C., & Vögele, C. (2021). Gastric interoception and gastric myoelectrical activity in bulimia nervosa and binge-eating disorder. *International Journal of Eating Disorders*, 54(7), 1106–1115. <https://doi.org/10.1002/eat.23291>
- van Dyck, Z., Vögele, C., Blechert, J., Lutz, A. P. C., Schulz, A., & Herbert, B. M. (2016). The water load test as a measure of gastric interoception: Development of a two-stage protocol and application to a healthy female population. *PLoS One*, 11(9). <https://doi.org/10.1371/journal.pone.0163574>
- Vickers, K., Jafarpour, S., Mofidi, A., Rafat, B., & Woznica, A. (2012). The 35% carbon dioxide test in stress and panic research: Overview of effects and integration of findings. *Clinical Psychology Review*, 32(3), 153–164. <https://doi.org/10.1016/j.cpr.2011.12.004>
- von Leupoldt, A., Sommer, T., Kegat, S., Baumann, H. J., Klose, H., Dahme, B., & Büchel, C. (2008). The unpleasantness of perceived dyspnea is processed in the anterior insula and amygdala. *American Journal of Respiratory and Critical Care Medicine*, 177(9), 1026–1032. <https://doi.org/10.1164/rccm.200712-1821OC>
- Ward, S. A., Macias, D., & Whipp, B. J. (2001). Is breath-hold time an objective index of exertional dyspnoea in humans? *European Journal of Applied Physiology*, 85(3–4), 272–279. <https://doi.org/10.1007/s004210100457>
- Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113–120. <https://doi.org/10.3758/BF03202828>
- Wiley, R. L., & Zechman, F. W. (1966). Perception of added airflow resistance in humans. *Respiration Physiology*, 2(1), 73–87. [https://doi.org/10.1016/0034-5687\(66\)90039-9](https://doi.org/10.1016/0034-5687(66)90039-9)
- Woznica, A., Vickers, K., Koerner, N., & Fracalanza, K. (2015). Reactivity to 35% carbon dioxide in bulimia nervosa and panic disorder. *Psychiatry Research*, 228(3), 571–575. <https://doi.org/10.1016/j.psychres.2015.05.050>
- Yuan, Y.-Z., Tao, R.-J., Xu, B., Sun, J., Chen, K.-M., Miao, F., ... Xu, J.-Y. (2003). Functional brain imaging in irritable bowel syndrome with rectal balloon-distention by using fMRI. *World Journal of Gastroenterology: WJG*, 9(6), 1356–1360. <https://doi.org/10.3748/wjg.v9.i6.1356>