Received: 18 October 2019 | Revised: 25 October 2019 | Accepted: 12 November 2019

DOI: 10.1096/fba.2019-00085

RESEARCH ARTICLE

FASEB BioAdvances WILEY

# Vascular response to social cognitive performance measured by infrared thermography: A translational study from mouse to man

**Jan Seidel**[1] | **Fabian Bockhop**[1] | **Miso Mitkovski**[2] | **Sabine Martin**[1] | **Anja Ronnenberg**[1] | **Dilja Krueger-Burg**[3] | **Katharina Schneider**[2] | **Heiko Röhse**[2] | **Liane Wüstefeld**[1] | **Filippo Cosi**[4] | **Kai Bröking**[2] | **Annekathrin Schacht**[5] | **Hannelore Ehrenreich**[1]

[1]Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Göttingen, Germany

[2]Light Microscopy Facility, Max Planck Institute of Experimental Medicine, Göttingen, Germany

[3]Department of Molecular Neurobiology, Max Planck Institute of Experimental Medicine, Göttingen, Germany

[4]Biomedical Physics Group, Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany

[5]Department of Affective Neuroscience and Psychophysiology, Georg-Elias-Müller-Institute of Psychology, Georg-August University, Göttingen, Germany

**Correspondence**
Hannelore Ehrenreich, Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Hermann-Rein-Str.3, 37075 Göttingen, Germany.
Email: ehrenreich@em.mpg.de

**Funding information**
Max Planck Society; DFG Research Center for Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB)

## Abstract

To assess complex social recognition in mice, we previously developed the *SocioBox* paradigm. Unexpectedly, 4 weeks after performing in the *SocioBox,* mice displayed robust social avoidance during Y-maze sociability testing. This unique "sociophobia" acquisition could be documented in independent cohorts. We therefore employed infrared thermography as a non-invasive method of stress-monitoring during *SocioBox* testing (presentation of five other mice) versus empty box. A higher *Centralization Index* (body/tail temperature) in the *SocioBox* correlated negatively with social recognition memory and, after 4 weeks, with social preference in the Y-maze. Assuming that social stimuli might be associated with characteristic thermo-responses, we exposed healthy men (N = 103) with a comparably high intelligence level to a standardized test session including two cognitive tests with or without social component (face versus pattern recognition). In some analogy to the *Centralization Index* (within-subject measure) used in mice, the *Reference Index* (ratio nose/malar cheek temperature) was introduced to determine the autonomic facial response/flushing during social recognition testing. Whereas cognitive performance and salivary cortisol were comparable across human subjects and tests, the *Face Recognition Test* was associated with a characteristic *Reference Index* profile. Infrared thermography may have potential for discriminating disturbed social behaviors.

**KEYWORDS**

flushing, IRT, social stimulus, stress, temperature, vasoactivity

**Abbreviations:** *BSI*, Brief Symptom Inventory; *FRT*, Face Recognition Test; *HPA*, Hypothalamic-Pituitary-Adrenal Axis; *IRT*, Infrared Thermography; *LOESS*, Locally Estimated Scatterplot Smoothing; *LPS*, Leistungsprüfsystem; *mK*, Millikelvin; *NEO-FFI*, NEO-Five-Factor Inventory; *NETD*, Noise Equivalent Thermal Difference; *ROI*, Regions Of Interest; *SIAS*, Social Interaction Anxiety Scale; *SPS*, Social Phobia Scale; *STAI*, State-Trait Anxiety Inventory; *WCST*, Wisconsin Card Sorting Test.

Jan Seidel, Fabian Bockhop and Miso Mitkovski are equally contributing first authors.

# 1 | INTRODUCTION

An association of emotions during unaccustomed social interactions with facial flushing in humans has long been recognized.[1-3] The autonomic nervous system response during such social interactions—highly conserved across mammals and perceived like "stress"—leads to altered vasoactivity in peripheral and core body regions. The resulting blood flow changes via vasoconstriction and vasodilation, respectively, affect local body temperature.[4-10] Exposure to an embarrassment task, for instance, led to an increase in facial blood flow in both male and female participants, measured via Laser Doppler Flowmetry.[11] Using this technique, temperature changes upon sympathetic vasoconstriction, occurring with a delay of 5-15 s,[12-14] can be reliably detected. Negative as well as positive social stimuli provoke alterations in surface temperature of various facial areas, with the nose consistently reported as highly reactive to affective and social cues.[12,15-18]

These observations advocate infrared thermography (IRT) as a highly attractive method of contact-free and non-invasive measurement of naturally emitted electromagnetic radiation with a wavelength between 0.75-1000 μm, commonly interpreted as "heat".[19] Modern IRT recording systems are characterized by high spatial and temporal resolutions and require almost no restrictions in movement of test subjects, allowing a more natural/ecological testing environment.[9,15,20] Because of its high accuracy, relative ease of use, and minimal inconvenience for the subjects, IRT has already been implemented in different fields of medical research and practice.[19,21]

While the validity of IRT for assessing surface temperature is generally accepted and has led to several pivotal publications,[4,6,7,16,22-28] its broader applicability in the future will depend on controlling environmental, subject-related, and technological factors[29] as well as improved reliability and reproducibility. So far, no overall accepted, dependable state-of-the-art procedure for IRT testing in social behavior diagnostics has been introduced. Numerous different experimental designs, test stimuli, facial/body target regions, and data extraction/analysis procedures have been reported.[9,21] Often, single or short series of IRT images (before versus after experimental condition) are described, based on rather small and heterogeneous samples, whereas data on thermal dynamics over longer time intervals are scarce. Interpretation of thermal alterations is frequently limited to single directional statements, for instance increase or decrease or unaltered temperature.

In the present translational study, we employ and adapt IRT for more reliable, internally controlled measurement of a social stimulus-related autonomic vaso-response. We start with an unexpected discovery in mice, namely induction of "sociophobia" upon inescapable interaction in a social recognition test, where the C*entralization Index* (ratio body/tail temperature) serves as continuous "whole body stress readout". We then extend these findings to human subjects, exposed to social versus non-social cognitive tasks in a highly standardized fashion. Here, the *Reference Index* (ratio nose/malar cheek temperature) is introduced to determine the autonomic facial response/flushing during social recognition testing. We report a novel non-invasive "sociophobia" model in mice, characterized by a pronounced thermo-reaction during induction and on retrieval, and a typical facial thermo-response in men under cognitive challenge containing a social component.

# 2 | MATERIALS AND METHODS

## 2.1 | Mouse studies

### 2.1.1 | Mice

All experiments were approved by and conducted in accordance with the regulations of the local Animal Care and Use Committee (Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit, LAVES). C57BL/6JRj mice were used as experimental mice, C3H/HeNCrl as stimulus mice (Charles River). Animals were group-housed in standard cages (36.5 × 20.7 × 14 cm, 4-5 mice per cage of the same gender and strain), in rooms separated by gender and strain (to avoid olfactory contact), and kept on a 12 h light-dark cycle (lights off at 7 PM) at 20-22°C. Food and water were provided ad libitum.

### 2.1.2 | SocioBox test for complex social recognition memory and recording

A detailed description of the *SocioBox* as a multiple social recognition task is provided elsewhere[30] (see also Figure 1). Experiments were conducted during light phase of the day (10-15 lux, 23.5°C room temperature), with male or female C57BL/6JRj experimental mice (N = 45 in total) and gender-matched C3H/HeNCrl as stimulus mice (Figure 1A). Male mice were 13-15, female 20-22 weeks old. Prior to test session, experimental and stimulus mice had been habituated separately (in absence of any other mice) to the *SocioBox* for 3 consecutive days. The following test sessions consisted of three phases, namely exposure 1, exposure 2, and recognition test.[30] At beginning of test session, the experimental mouse was placed into the central arena inside a white Plexiglas circular partition, spatially and visually separated from stimulus mice. After 5 minutes of recovery ("Initiation stage"), the circular partition was lifted, and the mouse allowed to freely explore the arena, including the stimulus mice in their inserts, for 5 additional minutes ("Interaction stage"). At the end of exposure 1, the mouse was removed and placed back in its transport cage. The arena was cleaned and the mouse then

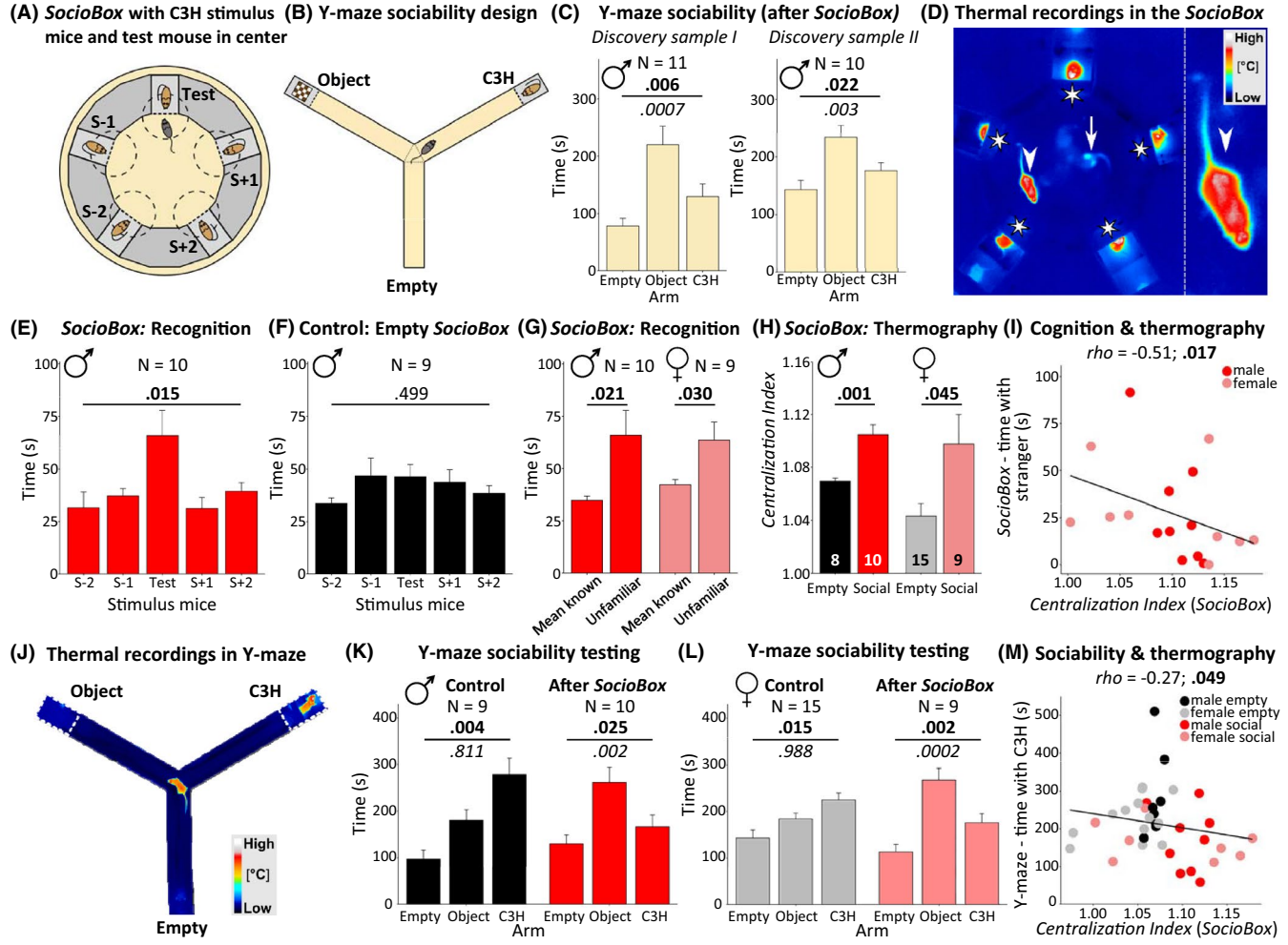


**FIGURE 1** Mouse IRT study: *SocioBox* recognition testing induces lasting social avoidance in mice. A, *SocioBox* arena[30] with experimental mouse in center (gray), unfamiliar "stranger" ("Test") and familiar stimulus mice (all brown). Time spent in zones close to each stimulus mouse (circles) is recorded to determine interaction/recognition. B, Y-maze sociability test:[31] Test mouse starts in center of Y-maze with one arm empty, one containing an object and the third another mouse (C3H). Time spent in each arm is measured. A normal naïve mouse spends most of the time with the other mouse (stair pattern of controls in panel K and L). C, Unexpectedly, 4 weeks after *SocioBox* testing, experimental mice display social avoidance behavior in 2 independent samples: *Discovery I and II*; repeated-measure ANOVA; quadratic-trend analysis (below; italics). D, IRT image of *SocioBox* test with experimental mouse (arrowhead) and stimulus mice (asterisks); white arrow: experimental mouse left trace of urine (evaporation cooling). Magnification on the right illustrates temperature differences in body parts. ***Compare video S1.*** E, Zone preference of male mice during *SocioBox* recognition testing. Mice with normal recognition memory spend most time with the "stranger" (unfamiliar stimulus mouse). F, Control mice tested in empty *SocioBox* do not show appreciable zone preference; repeated-measure ANOVA. G, Average interaction time with all familiar mice versus time with unfamiliar mouse (stranger); paired Student's *t* tests (one-sided). H, Both genders exhibit in *SocioBox* an increase in *Centralization Index* (body/tail temperature), compared to controls in empty box; unpaired Student's *t* tests (two-sided). Note that due to difficulties in tracking tail ROI (≥25% missing values), 3 animals (1 male, 2 female controls) had to be excluded from thermal analyses. I, *Centralization Index* is negatively correlated with social recognition performance; Spearman's *rho* (one-sided). J, Representative IRT image during Y-maze sociability. K-L, Both genders show robust social avoidance 4 weeks after *SocioBox* compared to the expected stair pattern of control animals; repeated-measure ANOVA; quadratic-trend analysis (below; italics). Due to atypical hypoactive behavior during testing, two female control animals were excluded prior to analyses. M, Social aversion priming: Negative correlation between *Centralization Index* in *SocioBox* and time spent with C3H conspecific in Y-maze 4 weeks later; Spearman's *rho* (one-sided)

again placed in the *SocioBox* center. Exposure 2 followed the same procedure. At the end of exposure 2, one of the five stimulus mice was randomly exchanged for a new, unfamiliar conspecific. Next, the mouse was reintroduced and the recognition test conducted accordingly. During the course of the experiment, a black body-calibrated A655sc IRT camera with a 13.1 mm focal length lens was used. The system has a noise equivalent thermal difference (NETD) <30 mK and resolution of 640 × 480 pixels (FLIR ResearchIR Max software v4.40.2.1, TOPA, Hohenpeissenberg, Germany) and was mounted 110 cm above the arena, recording at a frame-rate of 25 Hz. Care was taken that no direct or indirect heat

emissions from external sources affected recordings. The IRT camera was connected to a computer located in a separate room. Readouts were temperature changes of the mouse, duration of interaction with stimulus mice/recognition of the stranger mouse, and distance traveled (video S1). To ensure that sociability changes are not triggered by the *SocioBox* arena itself, the same procedure was conducted with control mice exposed to an empty *SocioBox* (without stimulus mice).

### 2.1.3 | Y-maze sociability testing and recording

Y-maze testing was performed as described with slight modifications.[31] Mice were tested 4 weeks after *SocioBox*/empty box performance on 2 consecutive days at light intensities of 60-70 lux. Day 1 included three habituation trials with an inter-trial-interval of 60 min. The mouse was placed in one of the arms and allowed to explore the empty maze for 10 min per trial. The starting point was rotated through all three arms (dimensions of each arm 46.1 × 8.3 × 13.7 cm). On day 2, an object (6 cm chess piece) and a C3H-stranger mouse, same gender and age, were each presented in an insert, preventing direct access, and positioned at the end of two randomly chosen Y-maze arms while the third arm remained empty (Figure 1B). The mouse was then placed in the empty arm, facing the center, and allowed to explore freely for 10 minutes. All mice underwent the same test procedure. The IRT camera was positioned 130 cm above the maze, readouts were changes in temperature, distance traveled and, to estimate social preference, time spent in each arm (Figure 1J).

### 2.1.4 | Data extraction and preprocessing

Mouse location and stress readouts during *SocioBox* recognition and Y-maze sociability tests were assessed through an image analysis workflow implementing the software packages Ilastik v.1.3.3b2[32] and FIJI,[33] as well as the TrackMate[34] FIJI plugin. Thermal readouts of both, body and tail, were extracted by first using the pixel classification workflow of Ilastik. Pixel groups delineating "background", "body", and "tail" were annotated to train a Random Forest classifier[32] that was used to produce probabilities for the respective classes for each image sequence of the recorded mice. Resulting body and tail probabilities were binarized with FIJI to generate masks, which were then applied to the corresponding, original IRT image sequence as regions of interest (ROI), from which the relative mean body and tail temperatures were obtained.

For the *SocioBox*, five zones were defined in close proximity to each stimulus mouse/empty inset (Figure 1A). Number of frames the respective mouse spent in each zone was summed up to obtain total interaction time with stimulus mice. Interaction with stranger (unfamiliar mouse) served as readout of social recognition.[30] To exclude zone preferences not attributable to experimental setup (eg, room features) during empty *SocioBox* condition, control mouse zone orientations were randomly matched to experimental mouse zones. A similar procedure was employed during Y-maze sociability, counting number of frames in each arm (empty, object, C3H).

After down-sampling (1 Hz) to increase computational speed during following preprocessing steps, frames with missing information (eg, hidden tail) were replaced by the mean of the remaining data points for each mouse. Mice with ≥25% missing values were excluded from respective thermographic analysis. To reduce random noise effects we smoothed data sequences of both body and tail separately, using locally estimated scatterplot smoothing (LOESS). By dividing the relative mean temperature of the body ROI by its corresponding tail ROI relative mean temperature at each time point, we created an intraindividually adjusted measure of endogenous arousal: the ***Centralization Index.*** To evaluate whether potential thermal differences were independent of higher physical activity, we additionally calculated the distance mice traveled in 500 ms intervals for *SocioBox* and Y-maze.

## 2.2 | Statistical analyses

Both male and female experimental (*SocioBox*) versus control (empty box) mice were analyzed. To reduce the impact of extreme values in statistical analyses while avoiding exclusion, data for each group (empty *SocioBox*, *SocioBox* with stimuli) were winsorized: extreme values <5th and >95th percentiles were set to 5th and 95th percentiles, respectively.[35] Total time spent in *SocioBox* zones was analyzed using repeated measure analyses of variance (ANOVA).[30] Additionally, for experimental groups, average time spent in zones with familiar mice was compared with time spent with stranger (unfamiliar zone). Due to expected outcome (more time spent with unfamiliar mouse),[30] one-sided paired Student's *t* tests were calculated. Differences in mean *Centralization Index* and total distance traveled were compared between conditions via two-sided unpaired Student's *t* tests. Exploring the relationship between *Centralization Index* as readout of physiological reactivity (stress) and recognition performance, Spearman's *rho* was calculated for all experimental mice. Because of initial orientation and adaption to the situation with potentially interfering effects on recognition performance, we used only the second half of *SocioBox* test (minutes 4-5). Hypothesizing that a higher *Centralization Index* is associated with worse performance, analysis was one-sided. For Y-maze sociability, differences in time spent in each arm was tested via repeated-measure ANOVA with following linear and quadratic trend analyses using polynomial contrasts. Both differences in

mean *Centralization Index* and total sum of distances traveled were analyzed with two-sided unpaired Student's *t* tests. Hypothesizing that the *Centralization Index* in Y-maze correlates negatively with sociability, one-sided Spearman's *rho* was calculated, including all test mice. Additionally, to investigate the relationship between severity of experience during *SocioBox* (assumed priming of social aversion) and sociability in Y-maze, we calculated Spearman's *rho* for *Centralization Index* in *SocioBox* and time spent in C3H-arm during Y-maze, again with test mice from all conditions. Expecting a negative correlation, a one-sided test was applied. All statistical analyses were performed using R v3.5.2[36] with RStudio v1.1.463 (RStudio Inc, Boston, United States) and significance level of *alpha* = 0.05. Welch-corrected Student's *t* tests were used, and, in cases of violations of sphericity, Greenhouse-Geisser corrections were applied to repeated-measure ANOVA.

## 2.3 | Human studies

### 2.3.1 | Participants

The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the Georg-Elias-Müller-Institute of Psychology, University of Göttingen. Online screening was set up to attract and assess eligibility of potential participants. Besides providing demographic information and answering questions regarding their ability to identify and memorize faces, interested individuals completed the German versions of Brief Symptom Inventory (BSI)[37] and complementary social phobia instruments Social Phobia Scale (SPS) and Social Interaction Anxiety Scale (SIAS)[38]. As additional readout of personality structure, NEO-Five-Factor Inventory (NEO-FFI)[39] was filled out. Those with questionnaire scores within normal limits were invited to the experimental session, aiming to include only mentally healthy individuals without indication of (sub)clinical symptoms. Based upon results of this online screening, a total number of N = 111 subjects were invited. However, due to psychiatric conditions, illicit substance consumption shortly before study onset, or technical difficulties during recording, N = 8 had to be excluded, leaving a final sample of N = 103 participants (see Figure 2A for inclusion process). All subjects were heterosexual, native German men between 18 and 34 years of age with normal or contact lens-corrected vision, no facial piercings or beard, and without history of neuropsychiatric or somatic diseases.

### 2.3.2 | Experimental procedure

To reduce impact of external factors during IRT recordings[9,15,29] participants were asked to avoid alcohol consumption (24 h), physical activity (12 h), and intake of food or activating substances (eg, caffeine, nicotine; 2 h) before test session (Figure 2 and Figure 3). Additionally, they were instructed not to shave or apply facial lotion at testing day. Study participation was compensated with 35€ or course credit. Completing the online screening offered the chance to win 1 of 3 gift cards (10€). All test subjects gave written informed consent and could withdraw participation at any time. Main experiments took place in a 5 × 3 $m^2$-sized testing room without direct sunlight or ventilation and with normal ambient temperature (M = 22.94°C, SD = 1.14) and humidity (M = 59.62%, SD = 7.34). Trained experimenters (JS, FB) ensured standardized test conditions during sessions,[29] which consisted of an initial assessment, habituation, two IRT-recorded computer tests of cognitive abilities, and a closing assessment (Figure 2B). Individual sessions consistently started at either 09:00 AM or 11:00 AM, total test duration did not exceed 120 minutes. During initial assessment, the participant was welcomed and informed about study procedure, followed by an examination of state-trait anxiety (German version of State-Trait Anxiety Inventory, STAI)[40] and, thereafter, general face perception abilities via a short prosopagnosia test. IRT-recorded cognitive testing was performed in a 3 × 2 $m^2$-sized chamber within the experimental room (Figure 3A). One chamber side was not completely closed to allow fresh air supply and communication between subject and experimenters. The participant was seated in a specialized comfortable chair which adapted to his body size and shape, effectively avoiding pressure points (M$^c$Lean REHAtechnik, Duderstadt, Germany). The implemented headrest enabled relaxation of head and neck muscles, while gently minimizing head movements (Figure 3A). The entire setup was highly adaptable to the differing subject shapes and sizes preventing irritation of the vascular system, while ensuring an unobstructed view of the relevant facial features for the IRT camera. Prior to testing, subjects stayed in a relaxed position for 15 minutes to acclimatize to setup (habituation phase).[15,29] Each participant performed computerized tests of both executive functioning and social cognition in counterbalanced order, separated by a 2 min break. The IRT camera was mounted above the monitor and recorded whole-face images of the participant at 25 Hz from approximately 32 cm distance while the subject performed the tasks.

*Executive functioning*

*Executive functioning* as a process of general cognitive abilities was measured via ***Wisconsin Card Sorting Test*, Computer Version 4 (WCST)**.[41] Subjects are required to virtually classify cards regarding different features (symbol, number, color) via button press to 1 of 4 target decks. No additional instructions are provided. Instead, subjects have to infer sorting strategies from a feedback ("correct"/"incorrect") following each sorting decision. After a series of correct answers

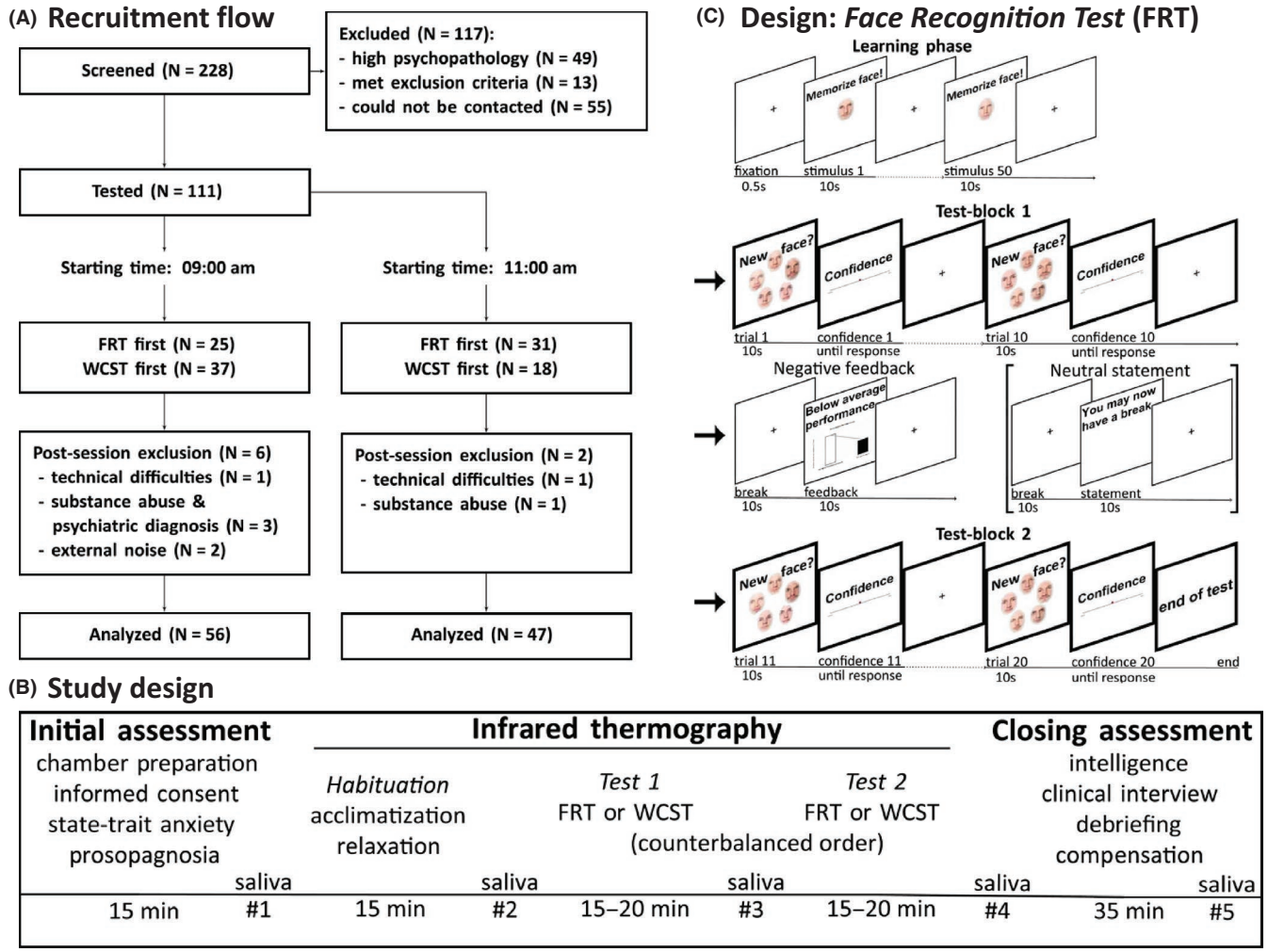


**FIGURE 2** IRT of social recognition versus executive function testing in men: Study logistics. A, Recruitment process: To ensure sample homogeneity, pre-experimental online screenings were implemented. A total of 111 individuals completed experimental session, starting at either 09:00 AM or 11:00 AM. After post-session exclusion, 103 subjects remained as final sample. B, Study design: During initial assessment, participants were welcomed and provided with study information, followed by state-trait anxiety and prosopagnosia examinations. Then, IRT-recorded test phase was conducted, starting with habituation and 2 computerized cognition tests (in counterbalanced order: FRT, WCST). At closing assessment, intelligence was measured, and an interview concerning mental and physical health as well as debriefing (explanation of feedback) and, finally, compensation for participation took place. Over the course of experimental session saliva samples for cortisol analysis were collected at 5 different time points in 15-35 min intervals. C, Sketch of the novel, brief *Face Recognition Test* (FRT): Participants are first asked to memorize 50 male stimulus faces (learning phase). Subsequently, test trials containing 5 stimulus faces each are presented; subjects decide whether all are familiar or not, respectively (test-block 1). After 10 trials either a fabricated, negative feedback or an alternative, neutral statement is presented, before concluding with 10 more trials (test-block 2), analogous to test-block 1

of certain length, sorting criteria shift, prompting subjects to adapt.

### *Social cognition*

*Social cognition* in humans was assessed based on face memory tasks.[42-44] Analogous to the *SocioBox* recognition test in mice, we created the ***Face Recognition Test*** **(FRT)**, a computerized measure of social cognitive abilities (face perception, face memory based on internal features, perception of social evaluation). The FRT consists of a learning phase, followed by two test blocks, each separated by a negative feedback or, alternatively, neutral statement (Figure 2C). The test is presented via PsychoPy v1.85.4[45] for Python v2.7.[46] During the learning phase, 50 male faces from the Göttingen Faces Database[47] with neutral valence, standardized visual features, luminance, and resolution are sequentially presented in random order. Test subjects are instructed to stay calm, focus and memorize the stimulus on screen. Each face is shown once for 10 s at the center of the screen, followed by fixation cross (500 ms). Learning phase lasts 560 s in total. For test-block 1, of the previous 50 stimulus faces, two are randomly replaced with unfamiliar ones. This new set of images is then given in 10 trials. Each trial includes 5 stimuli, randomly presented in a circular order (Figure 2C). Participants indicate via keyboard button press within 20 s if

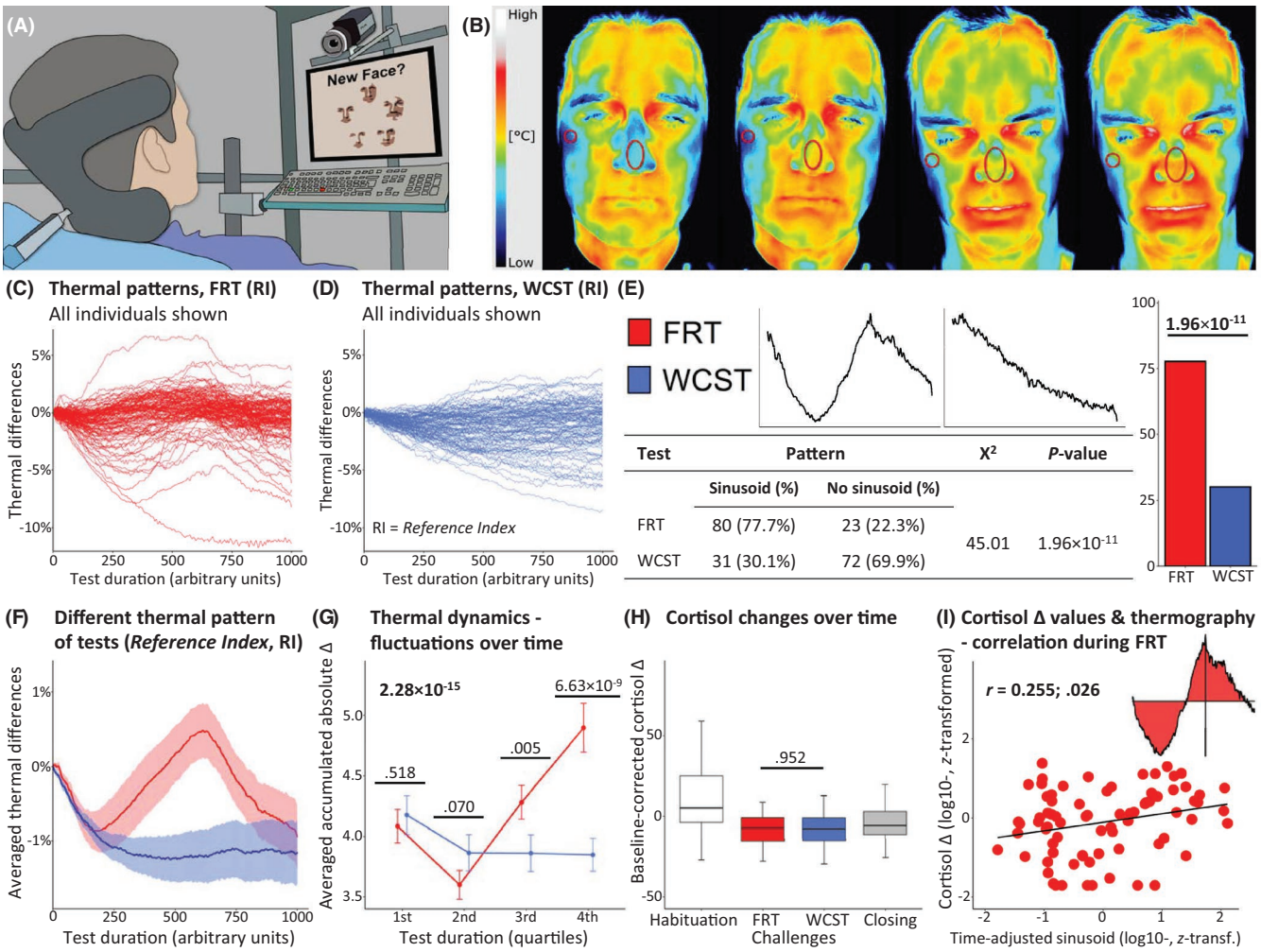


**FIGURE 3** Human IRT study during social recognition versus executive function testing: Social cognition stimulus induces a distinct thermo-pattern. A, Illustration of test setup: Participant sits as still as possible in a comfortable orthopedic armchair, with headrest to minimize head movements, and performs FRT (*Face Recognition Test*) while IRT camera (above screen) records facial regions of interest (ROI). B, IRT images of two sample test subjects with nose and right malar cheek ROI (circled in red) used to calculate *Reference Index* (RI). Images taken from early and late FRT session phase, respectively. ***Compare video S2.*** C-D, Overlays of all participants' normalized *Reference Index* differences from baseline in both tests indicate a sinusoid-shaped thermo-pattern over the course of FRT but not WCST (*Wisconsin Card Sorting Test*). E, Consensus ratings by three examiners (blinded to any test/subject information) revealed that the majority of test subjects exhibited a characteristic *Reference Index* sinusoid pattern during FRT (left example curve) but not WCST (right example curve); $\chi^2$-test. F, Comparison of *Reference Index* course between tests over time using M ± 95% confidence intervals. Groups' thermo-patterns differ significantly where confidence intervals do not overlap. G, Over each test quartile, accumulated absolute changes in *Reference Index* showed differences during second test halves, with higher temperature dynamics in FRT compared to WCST; repeated-measure ANOVA, Bonferroni-adjusted multiple-comparison tests (two-sided). H, Contrary to *Reference Index*, salivary cortisol reactivity was similar between tests, suggesting IRT as a more sensitive tool for measurement of physiological responses in social tasks. Cortisol levels were log10-transformed and normalized to the first sample (baseline); then cortisol *delta* values were calculated between sample time points; paired Student's *t* test (two-sided)**.** I, Pearson correlation coefficient revealed mild-to-moderate positive relationship between log10- and *z*-transformed time-adjusted integrals of *Reference Index* and *z*-transformed salivary cortisol *delta* values during FRT. Only participants with characteristic sinusoid thermal curve and *z*-score ± 2.58 included (N = 78)

all faces are familiar or not. Each stimulus face is shown only once and trials never contain more than one unfamiliar face. Additionally, after each trial, participants rate confidence in their response on a 5-point Likert scale. Then, after a break of 10 s, either an unprompted, fabricated negative feedback or a neutral statement is included, each lasting for 10 s. Negative feedback is shown graphically and as text (*"your performance is below average"*), the non-threatening alternative is displayed as text (*"you may now take a break - the second part of the test will start automatically"*). Immediately thereafter, test-block 2 is initiated with the two unfamiliar faces of test-block 1 again exchanged for two new faces and participants perform the same task with confidence judgments, analogous to test-block 1. Closing assessments covered a structured clinical interview on mental and somatic conditions, a nonverbal intelligence assessment (performance test system subtest-3;

"Leistungsprüfsystem Untertest-3"),[48] monetary compensation and, lastly, a debriefing on the aims of the study. As an additional, biological readout, saliva was collected to measure cortisol levels at five different time points in intervals of around 15-35 min (Figure 2B).

### 2.3.3 | Data extraction & preprocessing

Tracking information of facial ROI in human subjects was obtained with the DeepLabCut software package[49] and implemented into a FIJI-based image analysis workflow, allowing for corrections of small head movements not prevented by the headrest. Labels delineating either nose or right cheek (malar region) coordinates (Figure 3B; video S2) in up to 600 images were used to train the DeepLabCut network. Facial regions were selected regarding reactivity to social stimuli, with nose reacting strongly while malar cheek does not. After $10^6$ iterations, the resulting network had converged sufficiently to be evaluated for accuracy and then applied to human IRT recordings in FIJI in order to track and extract relative mean temperatures of the two facial ROI at their respective position and time. The resulting series of temperature values (25 per second) were subsequently down-sampled (1 Hz) for following processing steps. In order to replace missing values we calculated the sequence (ascending/descending) between the last valid data points before and after the missing. Then both ROI frame sequences were smoothed separately using LOESS fitting.

By dividing mean nose ROI temperature of each frame with its corresponding malar cheek ROI, we calculated the ***Reference Index***, analogous to the ***Centralization Index*** in mice. Due to initial temperature differences in respective ROI, *Reference Index* was normalized as percentage change from the very first frame (baseline), and, due to varying individual test length of participants, duration of both tests was normalized to 1000 arbitrary units, both for FRT (Figure 3C) and WCST (Figure 3D). To investigate differences in thermal curve characteristics, two independent evaluators rated in a blinded manner over both tests whether the normalized *Reference Index* curve was initially decreasing and then increasing (sinusoid curve), or was differently shaped (Cohens' *kappa* = 0.63). In case of dissent, a third evaluator made a final decision on the rating.

### 2.3.4 | Saliva cortisol determination

Saliva was collected at five different time points in intervals of 15-35 minutes (Figure 2B) and stored at −80°C until further use. ELISA was used to detect cortisol levels in saliva samples, according to manufacturer's instruction (Demeditec, Kiel, Germany). To account for circadian cortisol profile differences (ie, cortisol awakening response[50-52]) due to experimental starting points (09.00 AM versus 11.00 AM), we calculated normalized *delta* values (Δ) between sample collection time points: First, all samples were log10-transformed and normalized to percentage alteration from first sample (baseline). Next, differences between consecutive samples were calculated. This way we received adjusted cortisol changes for habituation, FRT, WCST, and closing assessment.

### 2.3.5 | Statistical analyses

Differences in frequency of sinusoid-curve ratings between FRT and WCST were analyzed using Yates'-corrected *chi*-squared test. To display averaged group differences over total test course, normalized *Reference Index* means and 95% confidence intervals were calculated, highlighting significant differences where confidence intervals do not touch. To additionally investigate thermal dynamics over test quarters, mixed-design ANOVA was calculated, with quarter-sums of absolute *Reference Index* changes per arbitrary unit as dependent variable, test quarters as within- and test as between-factor. Bonferroni-adjusted post-hoc multiple comparison tests were calculated to compare differences between individual test quarters. Cortisol reactivity changes between FRT and WCST were analyzed by comparing cortisol-Δ group values via two-sided paired Student's *t* test. All statistical tests were conducted in R using RStudio, with significance levels set to *alpha* = 0.05. Welch-corrected Student's *t* tests were used, and, in cases of violations of sphericity, Huyn-Feldt corrections were applied to repeated-measure ANOVA.

## 3 | RESULTS

### 3.1 | Unexpected discovery: *SocioBox* recognition testing induces lasting social avoidance in mice

To evaluate the social Y-maze test[31,50] for suitability as routine sociability readout in our mouse behavioral test battery, we used mice which had previously undergone *SocioBox*[30] experiments. By serendipity, we this way discovered that 4 weeks after passing through the *SocioBox* paradigm, these mice displayed social avoidance. This unexpected result was fully replicated in a second, independent cohort of former *SocioBox* completers (Figure 1A-C), leading to two first conclusions: (I) By chance, we may have developed a mouse model of social aversion priming/sociophobia, arising from a situation of inescapable social contacts. (II) The *SocioBox* test, even though superior to all other presently available social recognition tests and the first that successfully addresses multiple social contacts in parallel, will have to be treated as a final test in future behavioral test batteries (similar to eg, fear conditioning).

### 3.2 | IRT as non-invasive tool to measure the vascular response to social cognitive performance in a "sociophobia" inducing setup

To further explore the novel "mouse model of sociophobia", we employed IRT as a non-invasive method to continuously approximate experienced stress during these inescapable social contacts in the *SocioBox* (Figure 1D-I; video S1). Conveniently, IRT additionally provides monitoring of spatiotemporal dynamics, and thus location information needed for tracking. As expected, in the *SocioBox* recognition test, male mice spent significantly more time in the zone close to the stranger compared to already acquainted stimulus mice ($F_{(4,36)} = 3.58$; $P = .015$; Figure 1E), while control mice (empty *SocioBox*) did not exhibit any zone preference ($F_{(4,32)} = 0.86$; $P = .499$; Figure 1F). These findings were reproduced in female mice (*SocioBox*: $F_{(4,32)} = 3.03$; $P = .032$; empty box: $F_{(4,64)} = 1.21$; $P = .314$). Comparison of mean time spent with stranger versus all known mice yielded equivalent results (male $t_{(9)} = 2.38$; $P = .021$; female: $t_{(8)} = 2.20$; $P = .030$; Figure 1G). Screening the obtained IRT readouts in a few males first, we observed that mice changed their temperature over time in the *SocioBox* in a typical way, namely displayed an increase in body and a decrease in tail temperature (video S1). We therefore introduced a novel descriptive measure, integrating an internal control (within-subject) aspect, the ***Centralization Index***. This measure, likely approximating the experienced stress during the task, clearly demonstrates an increase in *SocioBox* mice versus controls for both males ($t_{(10.23)} = -4.44$; $P = .001$) and females ($t_{(10.66)} = -2.27$; $P = .045$; Figure 1H). Importantly, enhanced movement and thus physical activity cannot account for this difference since control mice even had a tendency to move more than *SocioBox* performers ($t_{(36.18)} = 1.90$; $P = .066$). Interestingly, the *Centralization Index* correlated negatively with the time spent with the stranger in the *SocioBox* (Spearman's *rho* = −0.51; $P = .017$; Figure 1I), indicating that mice with a higher *Centralization Index* (likely reflecting their stress level) perform worse in this social recognition task.

### 3.3 | Robust induction of social avoidance in the Y-maze sociability test following *SocioBox* recognition testing

Around 4 weeks after *SocioBox* testing, mice were exposed to Y-maze sociability testing, including IRT (Figure 1J). While male control mice exhibited normal social preference ($F_{(2,16)} = 7.88$; $P = .004$; linear trend: $b = 128.32$; $t_{(16)} = 4.85$; $P = .0004$; quadratic trend: $b = 6.44$; $t_{(16)} = 0.24$; $P = .811$), the prior *SocioBox* performers displayed social avoidance behavior, similar to the discovery samples ($F_{(2,18)} = 4.55$; $P = .025$; linear trend: $b = 25.86$; $t_{(18)} = 0.99$; $P = .334$; quadratic trend: $b = -92.48$; $t_{(18)} = -3.55$; $P = .002$; Figure 1K). Comparable effects were found for female mice in both control ($F_{(2,28)} = 4.89$; $P = .015$; linear trend: $b = 57.70$; $t_{(28)} = 3.80$; $P = .0007$; quadratic trend: $b = 0.23$; $t_{(28)} = 0.02$; $P = .988$) and post-*SocioBox* condition ($F_{(2,16)} = 9.56$; $P = .002$; linear trend: $b = 44.67$; $t_{(20)} = 2.16$; $P = .046$; quadratic trend: $b = -100.52$; $t_{(16)} = -4.87$; $P = .0002$; Figure 1L). These results, together with those of the two discovery samples, point to a robust induction of "sociophobia" by the *SocioBox* recognition test.

After Y-maze sociability testing, we controlled for potential differences in basic anxiety-related conduct. As a simple, non-invasive readout, spatial novelty-induced freezing in the fear-conditioning chamber (without shock) was evaluated. Importantly, neither male nor female *SocioBox* performers differed from empty box controls regarding duration of freezing (males: $F_{(2,36)} = 0.09$; $P = .910$; females: $F_{(2,28)} = 2.96$; $P = .068$, tendency in the opposite direction), excluding an "unspecific global fear behavior" underlying their "sociophobia" phenotype.

The next crucial question was whether we would see a correlation between stress, experienced in the *SocioBox*, as measured by the *Centralization Index*, and the degree of sociability evaluated 4 weeks later in the Y-maze. Indeed, a higher *Centralization Index* was associated with lower sociability (Spearman's *rho* = −0.27; $P = .049$) (Figure 1M). Together, these results support our hypothesis that inescapable social encounters can induce sociophobia/social aversion in mice.

The *Centralization Index* during Y-maze sociability testing also tended to be increased in *SocioBox* mice (M = 1.23, SD = 0.03) versus empty box controls (M = 1.21, SD = 0.02; both genders included; two-sided unpaired Student's *t* test, $t_{(34.47)} = -1.88$, $P = .068$). Once again, these higher *Centralization Index* values could not be explained by higher physical activity; control animals traveled more than their *SocioBox* counterparts (both genders included; $t_{(36.07)} = 2.81$; $P = .008$).

### 3.4 | Translational study: IRT as sensitive measure of the vascular response to social cognitive performance in humans

In our mouse experiments, we unexpectedly discovered that stress experienced during an inescapable social encounter (*SocioBox*) likely acts as a "primer" of sociophobia/social aversion. We thus started an IRT study in men, investigating in a translational fashion, whether a simple social component in a cognitive task (face recognition) would already yield thermographic results differing from a "non-social" cognitive test (pattern recognition) (Figure 2A-C; Figure

3A-I; Table 1; video S2). Of a total of N = 228 men screened for participation, N = 111 were tested, and N = 103 finally analyzed. Recruitment flow, study design and *Face Recognition Test* (FRT) are shown in Figure 2A-C.

All 103 subjects displayed high accuracy of face recognition in the prosopagnosia test (part of initial assessment; M = 97.67%, SD = 6.45) as prerequisite to perform the study. To ensure that subjects whose session started with FRT did not differ systematically from those with *Wisconsin Card Sorting Test* (WCST) first, we compared sociodemographic, psychopathological, and cognitive data between the two samples (Table 1). Since none of these variables showed any group differences, we combined both samples for the now following analyses of IRT readouts. First, we screened several facial IRT videos of participating subjects, allowing us to determine our regions of interest (ROI), namely nose (highly variable and seemingly responsive, as also described before)[15,18] and malar region of the cheek (obviously quite stable; compare video S2). We then calculated the *Reference Index*, in some analogy to the *Centralization Index* in mice, by dividing mean nose ROI temperature of each frame with its corresponding malar cheek ROI. Comparing the *Reference Index* course of all individuals during FRT, we noticed a sinusoid pattern predominating in most subjects (initial decrease, followed by increase) (Figure 3C). In contrast, only a minority of participants seemed to show such pattern during WCST (Figure 3D). To consolidate this first visual impression, independent raters estimated in a blinded fashion all individual thermal curves of both FRT and WCST to determine whether they resembled the characteristic sinusoid shape or not, with high interrater reliability (Cohen's *kappa* = 0.63 between first two raters). Comparisons of pattern frequency (sinusoid versus not) yielded highly significant differences, with 77.7% of participants showing the sinusoid curve during FRT but only 30.1% during WCST (*OR* = 7.98; Figure 3C-E). Interestingly, comparing participants with sinusoid-shaped *Reference Index* curve to those without revealed younger age and less time spent in the educational system together with higher scores in *Social Phobia Scale* (SPS) and *Social Interaction Anxiety Scale* (SIAS)[38] as well as lower answer security in FRT (Table 2).

Presentation of the normalized *Reference Index* total course of all subjects for each test (M ± 95% confidence interval) illustrates the sinusoid pattern during FRT versus the

**TABLE 1** Sociodemographic, psychopathological, and cognitive data of test subjects performing FRT or WCST first

| | Total sample N = 103 | FRT first N = 52 | WCST first N = 51 | $t/\chi^2$ | *P* |
|---|---|---|---|---|---|
| Neutral statement in FRT | 23 (22.33%) | 12 (23.08%) | 11 (21.57%) | 0.03 | .854 |
| Years of age | 24.58 (3.34) | 24.76 (3.49) | 24.40 (3.20) | 0.55 | .584 |
| Years of education | 16.83 (2.67) | 17.02 (2.86) | 16.65 (2.51) | 0.65 | .520 |
| BMI | 24.15 (2.97) | 23.75 (1.95) | 24.53 (3.67) | −1.25 | .216 |
| LPS-3 (*T*) | 61.52 (5.53) | 61.58 (6.20) | 61.46 (4.81) | 0.10 | .918 |
| STAI state | 32.41 (5.08) | 32.90 (5.62) | 31.90 (4.45) | 1.00 | .318 |
| STAI trait | 32.85 (6.94) | 33.79 (7.47) | 31.90 (6.27) | 1.39 | .168 |
| SPS sum | 6.22 (4.51) | 5.75 (4.35) | 6.71 (4.67) | −1.08 | .285 |
| SIAS sum | 15.95 (8.05) | 15.38 (7.88) | 16.53 (8.25) | −0.72 | .473 |
| BSI sum (*T*) | 45.81 (9.07) | 46.21 (9.58) | 45.39 (8.59) | 0.46 | .649 |
| NEO-Openness | 32.20 (6.50) | 33.06 (6.68) | 31.33 (6.25) | 1.35 | .179 |
| NEO-Conscientiousness | 32.31 (7.21) | 33.38 (7.09) | 31.22 (7.23) | 1.54 | .127 |
| NEO-Extraversion | 29.21 (6.61) | 29.15 (6.06) | 29.27 (7.19) | −0.09 | .927 |
| NEO-Agreeableness | 31.21 (6.93) | 31.08 (7.44) | 31.35 (6.45) | −0.20 | .841 |
| NEO-Neuroticism | 15.61 (7.45) | 16.12 (8.01) | 15.10 (6.87) | 0.69 | .490 |
| FRT error percentage | 42.28 (12.04) | 41.06 (12.58) | 43.53 (11.46) | −1.04 | .300 |
| FRT duration (s) | 969.12 (64.88) | 975.24 (68.86) | 962.74 (60.48) | 0.98 | .332 |
| FRT confidence | 3.40 (0.52) | 3.44 (0.47) | 3.36 (0.57) | 0.78 | .437 |
| WCST error percentage | 18.35 (9.43) | 17.29 (9.01) | 19.43 (9.82) | −1.15 | .251 |
| WCST duration (s) | 621.15 (161.36) | 612.04 (146.99) | 630.44 (175.80) | −0.58 | .566 |

*Note:* Data represent uncorrected *means* (SD) or N (%). Student's *t* tests and Pearson's *chi*-squared tests for independent comparisons were employed for analyses.

Abbreviations: BMI, Body Mass Index; BSI, Brief Symptom Inventory; FRT, *Face Recognition Test;* LPS-3, Leistungsprüfsystem-3 (performance test system subtest-3); SIAS, Social Interaction Anxiety Scale; SPS, Social Phobia Scale; STAI, State-Trait Anxiety Inventory; WCST, *Wisconsin Card Sorting Test*.

**TABLE 2** Sociodemographic, psychopathological, and cognitive data of subjects with/without sinusoid-shaped thermal curve during FRT

| | Total sample N = 103 | No sinusoid N = 23 | Sinusoid N = 80 | $t/\chi^2$ | *P* |
|---|---|---|---|---|---|
| Neutral statement in FRT | 23 (22.33%) | 5 (21.74%) | 18 (22.50%) | <0.01 | >.999 |
| Presenting FRT first | 52 (50.49%) | 12 (52.17%) | 40 (50.00%) | <0.01 | >.999 |
| Starting at 09:00 AM | 56 (54.37%) | 9 (39.13%) | 47 (58.75%) | 2.04 | .154 |
| Years of age | 24.58 (3.34) | 26.16 (3.90) | 24.13 (3.03) | 2.31 | **.028** |
| Years of education | 16.83 (2.67) | 18.24 (2.99) | 16.43 (2.46) | 2.41 | **.024** |
| BMI | 24.15 (2.97) | 24.92 (4.72) | 23.92 (2.21) | 0.92 | .366 |
| LPS-3 (*T*) | 61.52 (5.53) | 62.41 (4.24) | 61.26 (5.84) | 1.04 | .302 |
| STAI state | 32.41 (5.08) | 31.57 (6.27) | 32.65 (4.70) | −0.77 | .447 |
| STAI trait | 32.85 (6.94) | 30.17 (7.35) | 33.63 (6.66) | −2.03 | .051 |
| SPS sum | 6.22 (4.51) | 4.09 (3.41) | 6.84 (4.62) | −3.13 | **.003** |
| SIAS sum | 15.95 (8.05) | 12.91 (5.85) | 16.83 (8.40) | −2.54 | **.014** |
| BSI sum (*T*) | 45.81 (9.07) | 44.04 (9.48) | 46.31 (8.94) | 1.02 | .313 |
| NEO-Openness | 32.20 (6.50) | 33.00 (5.90) | 31.98 (6.68) | 0.71 | .481 |
| NEO-Conscientiousness | 32.31 (7.21) | 33.35 (8.05) | 32.01 (6.97) | 0.72 | .476 |
| NEO-Extraversion | 29.21 (6.61) | 30.57 (6.40) | 28.83 (6.66) | 1.14 | .262 |
| NEO-Agreeableness | 31.21 (6.93) | 31.13 (6.77) | 31.24 (7.02) | −0.07 | .948 |
| NEO-Neuroticism | 15.61 (7.45) | 14.26 (8.25) | 16.00 (7.21) | −0.92 | .367 |
| FRT error percentage | 42.28 (12.04) | 40.22 (12.20) | 42.88 (12.01) | −0.92 | .362 |
| FRT duration (s) | 969.12 (64.88) | 951.94 (69.46) | 975.79 (62.48) | −1.49 | .147 |
| FRT confidence | 3.40 (0.52) | 3.65 (0.55) | 3.33 (0.49) | 2.48 | **.018** |
| WCST error percentage | 18.35 (9.43) | 16.96 (6.72) | 18.75 (10.08) | −1.00 | .323 |
| WCST duration (s) | 621.15 (161.36) | 594.41 (139.00) | 636.83 (169.32) | −1.23 | .225 |

*Note:* Data represent uncorrected *means* (SD) or N (%). Student's *t* tests or Pearson's *chi*-squared tests for independent comparisons were employed for analyses.

P values < .05 are indicated in bold.

Abbreviations: BMI, Body Mass Index; FRT, *Face Recognition Test*; WCST, *Wisconsin Card Sorting Test*; LPS-3, Leistungsprüfsystem-3 (performance test system subtest-3); STAI, State-Trait Anxiety Inventory; SPS, Social Phobia Scale; SIAS, Social Interaction Anxiety Scale; BSI, Brief Symptom Inventory. Because of the exploratory nature of the study, no *P*-value adjustments were conducted.

continuous decrease followed by a plateau during WCST (Figure 3F). Mean temperature changes over test quartiles as another readout of thermal dynamics were likewise found significantly different, with greater fluctuations in the second half of FRT (quartile × condition-interaction: $F_{(3,612)} = 30.17$; $P = 2.28 \times 10^{-15}$; Figure 3G). Taken together, these data may point to an association of the sinusoid pattern with the emotional perception of the social task component.

Somewhat surprisingly, salivary cortisol alterations as "classical stress measures" did not differ during FRT and WCST ($t_{(101)} = -0.06$; $P = .952$; Figure 3H). This may question the validity of cortisol measurements for determining the specific stress caused by a social test component which can be sensitively detected by IRT. Nevertheless, cortisol reactivity during FRT correlated mildly positively with the time-adjusted integrals of *Reference Index* during FRT (Pearson correlation coefficient; $r = 0.255$; $P = .026$; Figure 3I), indicating at least a slight "typical" stress reaction during FRT.

Contrary to our expectations, no differences were found upon presentation of negative feedback versus neutral statement during FRT, neither in *Reference Index* curve ratings ($\chi^2_{(1)} < 0.01$; $P > .999$), nor average *Reference Index* curve shape, nor post-feedback temperature dynamics (fluctuations during 3rd and 4th test quartile, repeated-measure ANOVA interaction effect: $F_{(1,101)} = 2.38$; $P = .126$), nor any other variable (two-sided unpaired Student's *t* tests; all $P > .05$). This suggests that the negative feedback did not have any relevant impact on these measures.

## 4 | DISCUSSION

The present translational study demonstrates that IRT can be applied as a convenient, easy-to-apply, non-invasive technology to sensitively and reliably assess physiological reactivity ("flushing") in social contexts in humans and mice.

By implementing IRT within the *SocioBox* paradigm, we replicated and extended our previous findings that normal mice are able to easily recognize an unfamiliar stranger out of five stimulus mice. Strikingly, 4 weeks after performing this challenging social cognitive task, mice displayed abnormal social interaction in the Y-maze sociability test, namely distinct social avoidance. This unforeseen "sociophobia" following inescapable exposure to five conspecifics in the relatively narrow environment of the *SocioBox* was robustly reproduced several times in both genders. Control mice in the same narrow box without conspecifics ("empty") did not acquire this phenotype nor show any appreciable change in their basal anxiety behavior, as evaluated by spatial novelty-induced spontaneous freezing. Importantly, performance in the *SocioBox* as inducer of sociophobia was characterized by a higher overall *Centralization Index*, that is, higher temperature in central compared to peripheral body parts (tail), suggesting an increase in the experienced stress. Eye, body, or tail have previously been reported as stress-responsive IRT zones in mice.[4] Considering body and tail temperature of mice in IRT simultaneously, as introduced here with the *Centralization Index,* seems to constitute a promising robust measure of autonomous responses/social stress. Interestingly, the *Centralization Index* during *SocioBox* correlated negatively with the time spent with the stranger mouse in the social recognition task as well as the succeeding sociability test, indicating that the degree of stress influences cognitive performance (*SocioBox*) as well as severity of social avoidance (Y-maze). Hence, the *SocioBox* paradigm may serve not only as a superior test of complex social recognition memory,[30] but also as a reliable inducer of social avoidance, thereby delivering a novel non-invasive animal model of "sociophobia".[51] As a consequence, the *SocioBox* test has to be used as a final test in a behavioral battery, similar to fear conditioning.

This unexpectedly strong relation between social cognition testing and IRT readouts in mice raised the obvious translational question whether the addition of a social component to a cognitive task would yield characteristic IRT data also in human individuals that differ from those obtained during a non-social test. In human IRT, thermo-patterns depend strongly on stimuli used and facial ROI targeted.[9,16,20,29] We focused on the nose because of its high reactivity to social cues.[9,12,15-17,20] Since the introduction of the *Centralization Index* in mice had proven to be a reproducible, sensitive and widely environment-independent measure, we established a similar readout in human subjects. The *Reference Index*, again providing an "internal" (ie, within-subject) control by relating the responsive facial area (nose) to a rather temperature-stable zone (malar cheek) turned out to be a suitable tool to adjust for sources of IRT readout noise (eg, slight differences in ambient temperature, humidity, camera accuracy).[9,29]

Indeed, in the translational human study, we saw a characteristic sinusoid-shaped thermal curve with initial decrease in the majority of test subjects during the social FRT. In contrast, over the course of WCST, as non-social pattern recognition test, this typical curve was widely absent, with the temperature overall decreasing. However, not all participants responded with this characteristic social thermo-pattern, in the following referred to as "non-responders". Contrasting participants that exhibited the typical sinusoid-shaped thermo-pattern in FRT with the "non-responders" revealed interesting and plausible differences: younger age, less time spent in the educational system, higher scores in social phobia questionnaires (though still within the normal range), and less secure feedback-answers regarding their perceived own performance during FRT. Together, these differing items may point to lower experienced stress,[15] that is, to a more "relaxed attitude" toward social test settings.

Whereas IRT analyses revealed differences between social/non-social tests, salivary cortisol levels, an established standard measure of the hypothalamus-pituitary-adrenal (HPA) stress response, did not. To exclude potential bias due to different starting times of our test sessions (9.00 AM or 11.00 AM), falling into the cortisol diurnal profile/awakening response,[52-54] we employed normalized cortisol changes (*delta* values) between time points of sample collection. Comparing subject groups separated by starting time or order of test presentation (FRT or WCST first) did not reveal differences in *delta* cortisol values. These negative cortisol findings are in agreement with previous investigations on the potential of IRT in physical and social stress paradigms, compared to recognized stress markers.[16] While thermal readouts in various facial regions were sensitive to stress-induced mood changes, conventional stress markers, such as cortisol, were not,[16] suggesting a different origin (less HPA-related, more autonomic/catecholaminergic) and time course (fast versus delayed) of the experienced stress. Taken together, IRT seems to have a higher discriminative power for assessment of social cognition-related stress than cortisol.

Somewhat surprising, the fabricated negative feedback within FRT (*"your performance is below average"*), compared to the neutral statement (*"you may now take a break—second part of test will start automatically"*) did not induce any measurable differences in thermograms. This may be due to a lower than expected socially threatening/embarrassing impact, the shortness of presentation (only 10 s each)[13] with test-block 2 following immediately thereafter, or doubts of subjects regarding feedback authenticity. Finally, temperature alterations in connection with such ultra-brief negative feedback may belong to different underlying processes.[20] Neither nose nor malar cheek may be optimal for exploring respective facial thermo-reactions.

In recent years, the potential of IRT as a valid research tool alongside traditional physiological approaches has been increasingly explored.[9,15,17] However, IRT did not instantaneously turn out as a straightforward and easy-to-apply

method. IRT is sensitive to numerous interfering factors, arising from environmental (eg, ambient temperature, humidity, room size, radiation) and individual sources (eg, gender, age, amount of brown adipose tissue, physical activity, food or substance intake). Further inconsistencies and reliability problems were caused by suboptimal study design, such as small or heterogeneous samples, artifacts due to manual ROI definition or quantification, or camera signal noise.[9,29] Many studies did not sufficiently control for these methodological issues, leading to weak internal consistencies.[9,29]

In the present study, considerable effort was made to limit the impact of such interfering factors. As for the animal part, inbred mice were housed under controlled conditions and tested under standardized settings. In the human part, healthy male individuals with highly comparable sociodemographic characteristics were included. Large enough (N = 45 mice, N = 103 humans) test samples and standardized testing and recording procedures under constant ambient conditions were used as suggested by Fernández-Cuevas and colleagues.[29] While various studies analyzed single or short series of images due to technological or memory-storage limitations,[9] we used relatively long IRT video recordings (5-10 minutes for mice; >15 minutes per human participant and test) with high spatial and temporal resolution, and novel methods of data extraction and analysis. *Centralization Index* and *Reference Index* were introduced here as sensitive and widely environment-independent measures, providing internal (within-subject) control in the assessment of thermo-reactions in social contexts. Automated tracking and preprocessing algorithms delivered examiner-independent, objective and clean data extraction and organization, while imputations, smoothing, and winsorizing of data were conducted to reduce the impact of IRT camera inaccuracy, noise, and missing data.

Recently, IRT has also been employed for subjects diagnosed with mild posttraumatic stress disorder, Alzheimer's disease, or schizophrenia,[55-57] underlining that psychological/psychiatric research might profit from the contact-free, non-invasive IRT of freely moving and interacting subjects. In fact, deficits in social interaction/cognition of various origins are frequently seen in neuropsychiatry and often difficult to diagnose cross-sectionally. The current study on healthy individuals may stimulate future standardized social interaction testing using IRT in disease states, thereby opening new avenues for differential diagnostic approaches.

To summarize, based on a unique translational IRT study from mouse to man, we suggest that inclusion of a social component in a cognitive task specifically alters local body or face temperature, indicating a defined vascular response to this particular category of stress. These rather clear-cut findings were only possible on the ground of highly standardized and innovative experimental conditions, including IRT videos over an extended period to long-term monitor temperature alterations, unusually large, homogeneous subject samples, novel measures of internally (within-subject) controlled temperature over time, that is, *Centralization* and *Reference Index* and, finally, novel approaches to data acquisition, preprocessing, and analyses.

## ACKNOWLEDGEMENTS

This study was supported by the Max Planck Society and the DFG Research Center for Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB).

## CONFLICT OF INTEREST

The authors declare no competing financial or other interests.

## AUTHOR CONTRIBUTIONS

Concept, study design and supervision: HE together with MM & AS. Experimental design and interpretation: JS and FB together with HE & AS. Data acquisition and analysis: JS, FB, MM, SM, AR, DKB, KS, HR, LW, FC, KB, AS & HE. Drafting manuscript: HE together with JS, FB, MM & AS. Drafting display items: JS, FB, MM & HE. All authors read and approved the final version of the manuscript.

## DATA AVAILABILITY STATEMENT

Full data availability will be provided. Accession codes will be available before publication.

## REFERENCES

1. Darwin CR. *The expression of the emotions in man and animals*. London: John Murray; 1872.
2. Leary MR, Britt TW, Cutlip WD 2nd, Templeton JL. Social blushing. *Psychol Bull*. 1992;112:446-460.
3. Drummond PD. The effect of adrenergic blockade on blushing and facial flushing. *Psychophysiology*. 1997;34:163-168.
4. Gjendal K, Franco NH, Ottesen JL, Sorensen DB, Olsson IAS. Eye, body or tail? Thermography as a measure of stress in mice. *Physiol Behav*. 2018;196:135-143.
5. Meyer CW, Ootsuka Y, Romanovsky AA. Body temperature measurements for metabolic phenotyping in mice. *Front Physiol*. 2017;8:520.
6. Mufford JT, Paetkau MJ, Flood NJ, Regev-Shoshani G, Miller CC, Church JS. The development of a non-invasive behavioral model of thermal heat stress in laboratory mice (Mus musculus). *J Neurosci Methods*. 2016;268:189-195.
7. Herborn KA, Jerem P, Nager RG, McKeegan DEF, McCafferty DJ. Surface temperature elevated by chronic and intermittent stress. *Physiol Behav*. 2018;191:47-55.
8. Vianna DM, Carrive P. Changes in cutaneous and body temperature during and after conditioned fear to context in the rat. *Eur J Neurosci*. 2005;21:2505-2512.
9. Ioannou S, Gallese V, Merla A. Thermal infrared imaging in psychophysiology: potentialities and limits. *Psychophysiology*. 2014;51:951-963.
10. Drummond PD. Facial flushing during provocation in women. *Psychophysiology*. 1999;36:325-332.
11. Drummond PD, Su D. The relationship between blushing propensity, social anxiety and facial blood flow during embarrassment. *Cogn Emot*. 2012;26:561-567.

12. Esposito G, Nakazawa J, Ogawa S, et al. Baby, you light-up my face: culture-general physiological responses to infants and culture-specific cognitive judgements of adults. *PLoS ONE*. 2014;9:e106705.
13. Kistler A, Mariauzouls C, von Berlepsch K. Fingertip temperature as an indicator for sympathetic responses. *Int J Psychophysiol*. 1998;29:35-41.
14. Yoshihara K, Tanabe HC, Kawamichi H, et al. Neural correlates of fear-induced sympathetic response associated with the peripheral temperature change rate. *NeuroImage*. 2016;134:522-531.
15. Cardone D, Merla A. New Frontiers for Applications of Thermal Infrared Imaging Devices: Computational Psychophysiology in the Neurosciences. *Sensors (Basel)* 2017;17:E1042.
16. Engert V, Merla A, Grant JA, Cardone D, Tusche A, Singer T. Exploring the use of thermal infrared imaging in human stress research. *PLoS ONE*. 2014;9(3):e90782.
17. Kosonogov V, De Zorzi L, Honore J, et al. Facial thermal variations: a new marker of emotional arousal. *PLoS ONE*. 2017;12:e0183592.
18. Ermatinger FA, Brugger RK, Burkart JM. The use of infrared thermography to investigate emotions in common marmosets. *Physiol Behav*. 2019;211:112672.
19. Lahiri BB, Bagavathiappan S, Jayakumar T, Philip J. Medical applications of infrared thermography: a review. *Infrared Phys Technol*. 2012;55:221-235.
20. Ioannou S, Morris PH, Baker M, Reddy V, Gallese V. Seeing a blush on the visible and invisible spectrum: a functional thermal infrared imaging study. *Front Hum Neurosci*. 2017;11:525.
21. Ring EF, Ammer K. Infrared thermal imaging in medicine. *Physiol Meas*. 2012;33:R33- R46.
22. Garbey M, Sun N, Merla A, Pavlidis I. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Trans Biomed Eng*. 2007;54:1418-1426.
23. Murthy R, Pavlidis I. Noncontact measurement of breathing function. *IEEE Eng Med Biol Mag*. 2006;25:57-67.
24. Merla A, Di Donato L, Romani GL, Proietti M, Salsano F. Comparison of thermal infrared and laser doppler imaging in the assessment of cutaneous tissue perfusion in scleroderma patients and healthy controls. *Int J Immunopathol Pharmacol*. 2008;21:679-686.
25. Pavlidis I, Tsiamyrtzis P, Shastri D, et al. Fast by nature - how stress patterns define human experience and performance in dexterous tasks. *Sci Rep*. 2012;2.
26. Krzywicki AT, Berntson GG, O'Kane BL. A non-contact technique for measuring eccrine sweat gland activity using passive thermal imaging. *Int J Psychophysiol*. 2014;94:25-34.
27. Jarlier S, Grandjean D, Delplanque S, et al. Thermal Analysis of Facial Muscles Contractions. *IEEE Transactions on Affective Computing*. 2011;2:2-9.
28. Salazar-Lopez E, Dominguez E, Juarez Ramos V, et al. The mental and subjective skin: Emotion, empathy, feelings and thermography. *Conscious Cogn*. 2015;34:149-162.
29. Fernández-Cuevas I, Bouzas Marins JC, Arnáiz Lastras J, et al. Classification of factors influencing the use of infrared thermography in humans: A review. *Infrared Phys Technol*. 2015;71:28-55.
30. Krueger-Burg D, Winkler D, Mitkovski M, et al. The sociobox: a novel paradigm to assess complex social recognition in male mice. *Front Behav Neurosci*. 2016;10:151.
31. Brimberg L, Mader S, Jeganathan V, et al. Caspr2-reactive antibody cloned from a mother of an ASD child mediates an ASD-like phenotype in mice. *Mol Psychiatry*. 2016;21:1663-1671.
32. Sommer C, Straehle C, Kothe U, Hamprecht FA. Ilastik: Interactive learning and segmentation toolkit. *IEEE*. 2011;230-233.
33. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9:676-682.
34. Tinevez JY, Perry N, Schindelin J, et al. TrackMate: An open and extensible platform for single-particle tracking. *Methods*. 2017;115:80-90.
35. Wilcox RR. Winsorized Robust Measures. 2017; 1–2.
36. R Core Team. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing; 2018.
37. Franke GH. *BSI - Brief Symptom Inventory - deutsche Fassung*. Göttingen: Beltz; 2000.
38. von Consbruch K, Stangier U, Heidenreich T. *Skalen zur Sozialen Angststörung (SOZAS)*. Göttingen: Hogrefe; 2016.
39. Borkenau P, Ostendorf F. *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae*. Göttingen: Hogrefe; 1993.
40. Grimm J. State-Trait Anxiety Inventory nach Spielberger. Deutsche Lang- und Kurzversion: Methodenforum der Universität Wien; 2009.
41. Heaton KH, Chelune GJ, Talley JL, Kay G, Curtiss G. *Wisconsin Card Sorting Test: computer version 4 (WCST: CV4) research edn*. San Antonio, TX: Harcourt Assessment; 2003.
42. Duchaine BC, Weidenfeld A. An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*. 2003;41:713-720.
43. Herzmann G, Danthiir V, Schacht A, Sommer W, Wilhelm O. Toward a comprehensive test battery for face cognition: Assessment of the tasks. *Behav Res Methods*. 2008;40:840-857.
44. Warrington EK. *Recognition memory test: Manual*. Berkshire, UK: NFER-Nelson; 1984.
45. Peirce JW. PsychoPy–Psychophysics software in Python. *J Neurosci Methods*. 2007;162:8-13.
46. Van Rossum G. Python 2.7.10 Language Reference. Samurai Media Limited; 2015.
47. Kulke L, Janßen L, Demel R, Schacht A. Validating the Goettingen Faces Database. Open Science Framework; 2017.
48. Horn W. *L-P-S Leistungsprüfsystem*. Göttingen: Hogrefe; 1983.
49. Mathis A, Mamidanna P, Cury KM, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci*. 2018;21:1281-1289.
50. Lai WS, Johnston RE. Individual recognition after fighting by golden hamsters: a new method. *Physiol Behav*. 2002;76:225-239.
51. Toth I, Neumann ID. Animal models of social avoidance and social fear. *Cell Tissue Res*. 2013;354:107-118.
52. Boehringer A, Tost H, Haddad L, et al. Neural correlates of the cortisol awakening response in humans. *Neuropsychopharmacology*. 2015;40:2278-2285.
53. Fries E, Dettenborn L, Kirschbaum C. The cortisol awakening response (CAR): facts and future directions. *Int J Psychophysiol*. 2009;72:67-73.
54. Clow A, Hucklebridge F, Stalder T, Evans P, Thorn L. The cortisol awakening response: More than a measure of HPA axis function. *Neurosci Biobehav Rev*. 2010;35:97-103.
55. Jian BL, Chen CL, Chu WL, Huang MW. The facial expression of schizophrenic patients applied with infrared thermal facial image sequence. *BMC Psychiatry*. 2017;17:229.
56. Di Giacinto A, Brunetti M, Sepede G, Ferretti A, Merla A. Thermal signature of fear conditioning in mild post traumatic stress disorder. *Neuroscience*. 2014;266:216-223.

57. Perpetuini D, Cardone D, Chiarelli AM, et al. Autonomic impairment in Alzheimer's disease is revealed by complexity analysis of functional thermal imaging signals during cognitive tasks. *Physiol Meas*. 2019;40:034002.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Seidel J, Bockhop F, Mitkovski M, et al. Vascular response to social cognitive performance measured by infrared thermography: A translational study from mouse to man. *FASEB BioAdvances*. 2020;2:18–32. https://doi.org/10.1096/fba.2019-00085