Judgment and Decision Making, Vol. 16, No. 3, May 2021, pp. 766-795

Turning up the heat: The impact of indoor temperature on selected cognitive processes and the validity of self-report

Martijn Stroom*† Nils Kok† Martin Strobel‡ Piet M. A. Eichholtz†

Abstract

Indoor climate interventions are often motivated from a worker comfort and productivity perspective. However, the relationship between indoor climate and human performance remains unclear. We assess the effect of indoor climate factors on human performance, focusing on the impact of indoor temperature on decision processes. Specifically, we expect heat to negatively influence higher cognitive rational processes, forcing people to rely more on intuitive shortcuts. In a laboratory setting, participants (N=257) were exposed to a controlled physical environment with either a hot temperature (28° C) or a neutral temperature (22° C) over a two-hour period, in which a battery of validated tests were conducted. We find that heat exposure did not lead to a difference in decision quality. We did find evidence for a strong gender difference in self-report, such that only men expect that high temperature leads to a significant decline in performance, which does in fact not materialize. These results cast doubt on the validity of self-report as a proxy for performance under different indoor climate conditions.

Keywords: indoor climate, heat, performance, decision quality, heuristics, biases, risk-taking, self-report

*Corresponding Author, Email: m.stroom@maastrichtuniversity.nl, ORCID iD: 0000-0003-3411-4260
†School of Business and Economics, Department of Finance, Maastricht University.
‡School of Business and Economics, Department of Economics, Maastricht University.

Data are available at https://osf.io/5dgeu/.

We thank seminar participants at Maastricht University, Juan Palacios, Wouter van Marken Lichtenbelt, Rick Kramer, the members of the 2019 Social Judgement and Decision Making Conference, as well as participants at the doctoral poster session of the American Real Estate and Urban Economics Association during the 2020 Allied Social Sciences Association Conference. We are especially grateful for the valuable comments of Caroline Goukens and Gordon Pennycook. We thank Sustainably.io for the use of their indoor climate sensor 'Birdnest'. Finally, we thank the editor and referees of JDM for their critical and valuable insights. This paper is financially supported by Maastricht University's Graduate School of Business and Economics Primary Data Collection Seeding Grant 2019.

1 Introduction

Performance at work is influenced by many factors, such as individual characteristics, leadership, work pressure, incentive schemes, and corporate structure (Hermalin & Weisbach, 1991; Perry & Porter, 1982; Wageman & Baker, 1997). The physical climate of the workplace is often overlooked as an important factor influencing performance. And when it is mentioned, the dominant strain of research focuses on comfort, through self-report on physical aspects of the environment and their effect on human performance. This is remarkable, as office buildings have been undergoing rigorous innovations throughout recent decades (for instance, Vermeulen & Hovens, 2006). Developments in the quality of insulation, ventilation, and air-conditioning are effectively changing the indoor environment to which workers are exposed. These innovations are typically motivated by effects on building efficiency and/or worker comfort, but while there is ample research highlighting the effects of increased energy efficiency on building resource consumption (Eichholtz, Holtermans & Kok, 2019; Pérez-Lombard, Ortiz & Pout, 2008), the link between changes in indoor environmental conditions and human performance remains a topic of debate (MacNaughton et al., 2017; Satish et al., 2012; Zhang et al., 2017).

Research regarding the impact of indoor environment on worker performance is hampered by the fact that high-skilled performance measures at work are difficult to obtain directly, and are hard to compare between disciplines. For example, Zivin and Neidell (2012) show that pear-pickers' performance suffers from exposure to bad environmental quality conditions. However, the output of highly skilled workers who face cognitively demanding tasks – such as academics, managers, doctors, or investors – lacks such direct outcome measure. It is exactly this type of high-skilled workers that spends considerable time in confined offices or meeting rooms, subject to specific indoor climate conditions. Parsons (2014) notes that individual factors often dominate performance outcomes, making it even more challenging to compare productivity between workers. Moreover, any output that is measurable is not easily traced back to a quantifiable time period of exposure to the physical indoor climate.

To circumvent the challenge to correctly assess human performance, research has shifted from measuring performance to comfort (Bluyssen, 2013). The implicit expectation is that when the climate is rated as "comfortable", productivity increases. Comfort measures are an attractive proxy for productivity and performance, as they are easily and inexpensively assessed by self-report. Comfort could be treated as a measure of interest on its own (for instance, Nakamura et al., 2008), but whether self-assessed comfort levels are indeed an accurate proxy for performance remains an open question. Psychological research repeatedly suggests self-reported introspection into one's own subjective experience and emotions to be unreliable (Engelbert & Carruthers, 2010).

In this paper, we assess the effect of indoor environmental conditions on human performance, by investigating decision processes. Tversky and Kahneman (1974), amongst others, distinguish decision making as "intuitive" and "rational" processes. Automated, intuitive

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

rules of thumb, or heuristics, are "quick and dirty" and applied without much effort. The rational processes need more time and cognitive resources, are only scarcely applied, and are also associated with high decisional quality. A mainstream application of the interplay between these fast and rational or effortful processes is the default-interventionist approach (Evans, 2007). It stipulates that the effortful processes can intervene in the fast heuristics, when a wrongful application (a bias) in a given context is detected. Thus, whenever the effortful processes are hampered, for instance due to cognitive constraint resulting from environmental factors, increased bias-susceptibility generally lowers overall decisional quality (Gawronski & Bodenhausen, 2006; Muraven & Baumeister, 2000). In other words, we expect that bias detection and correction will (partially) suffer due to cognitive constraint in effortful processes capacity following temperature stress.

1.1 Literature

1.1.1 Temperature and Cognition

Psychological and neurological research has attempted to identify the effects of temperature on cognitive functions. We elaborate on two relevant findings.

The most profound and general finding is that cognitive capacity is lowered by adverse temperature conditions. Wright, Hull and Czeisler (2002) find that changes in the temperature of the body and brain are correlated with changes in performance, such that deviating temperatures from the internal optimal will worsen performance. Shibasaki, Namba, Oshiro, Kakigi and Nakata (2017) show that neurological inhibition processes suffer from heat stress. In decision-making, executive and inhibition processes coordinate which stimuli to act on (execute) and which not (inhibit). Both these biological processes are found to be weaker under heat stress. Van Ooijen, Van Marken Lichtenbelt, Van Steenhoven and Westerterp (2004) suggest that temperature could influence mental performance as a result of fatigue. This view is similar to the theoretical concept of mental depletion, the cognitive model stipulating limited mental "control" resources for self-regulation (Baumeister Bratslavsky, Muraven & Tice, 1998). Mental depletion often results in more instinctive behaviour (such as aggression; Van Lange, Rindery & Bushman, 2017). In general, when external stimuli overstimulate, concentration and performance become more costly (MacLeod, 1991).[1] Indeed, Cheema and Patrick (2012) show that temperature generally lowers cognitive performance, but not for people who were already mentally depleted at the start of the task. Although mental depletion is debated (Carter, Kofler, Forster & Mccullough, 2015; Hagger et al., 2016), the general notion of negative cognitive performance

[1]The distraction due to discomfort and the active act to ignore this distraction can drain additional resources from available mental capacity. However, the majority of the research previously described sees a loss of performance independent of awareness, suggesting that awareness of discomfort alone does not fully explain the decrease in performance. Additionally, the temperature dissatisfaction levels (see Table 1) in our experiment do not reach extreme levels, suggesting against high levels of rumination during the task. Therefore, we argue that the physiological capacity limitations from compensating the effect heat has on the body and its processes is more profound and is our main focus.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

effects after enduring strain on mental capacity seems to be a common denominator in ongoing self-regulation discussions (Cunningham & Baumeister, 2016; Baumeister, Vohs & Tice, 2007; Lin, Saunders, Friese, Evans & Inzlicht, 2020; Hockey, 2013; for an overview, see Inzlicht, Werner, Briskin & Roberts, 2021).

The second key finding of research on temperature and cognition is that not all mental processes are affected equally. Lowered cognitive capacity appears theoretically very close to behavioural fatigue. However, it is important to understand that these two concepts are fundamentally and hierarchically distinct. When discussing behavioural fatigue, we consider a general lowering of behavioural activity (i.e., a "global" effect). Decrease of cognitive capacity does not have a general uniform effect, but is depending on the neurological area that suffers most (i.e., a "local" effect). Lan, Lian, Pan and Ye (2009) found performance to decrease with adverse temperatures, but the effects differ across tasks.

In sum, it is clear that temperature has a general, or global, effect on cognition and cognitive performance, and that some local effects can be identified as well.

1.1.2 Temperature and Intuition

The literature review by Hancock and Vasmatzidis (2003) suggests that high capacity and complex mental processes are more profoundly affected by temperature than automated processes. Automated tasks rely on a strong and fast relation between stimulus and response, making them less susceptible to mental constraints (Kahneman, 1973). Automated tasks are part of system I in Kahneman's cognitive framework – also known as the intuitive system. They rely on intuition and on simple rules of thumb that are learned and are often successfully applied to predictable situations. System II is slow and costly on mental resources, but is generally associated with high-quality decision making.

Cognitive capacity and cognitive control are highly correlated (Engle & Kane, 2003), and the latter has also been found to be affected by temperature. Shibasaki, Namba, Oshiro, Kakigi and Nakata (2017) show that neurological inhibition processes suffer from heat stress. In decision making, inhibition and executive processes coordinate to achieve an optimal solution. As such, the effect of heat on performance can be twofold: not only do higher-order complex tasks suffer more than simple automated tasks (Grether, 1973), but wrongful application of an automated process or application of a wrong automated process might also be less likely to be corrected. In other words, even when the direct effect of heat on simple and automated processes is not evident (as stated by Zhang & de Dear, 2017), the outcome can still suffer in quality due to the lack of high order process intervention. Indeed, Hancock and Vasmatzidis (1998) found that highly skilled operators suffer less from performance decrease under heat stress, and they argued that this is most likely a result of performance depending on automated internalized processes.

The cognitive framework of Tversky and Kahneman leads to relevant predictions when we apply the findings of temperature on task complexity and intuition. The interaction found between temperature and automated tasks and task complexity suggests that system

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

I could be less affected than system II. The default-interventionist approach (Evans, 2007) stated that both systems work parallel to each other, and system II generally attempts to identify mistakes made by system I and intervenes if necessary. Recent advances in this field suggest that logical conclusions also manifest intuitively (De Neys & Pennycook, 2019). In this view, deliberation by system II is activated only when both the heuristic and logic intuition are of similar strength and conflicting. Thus, a correct response on the CRT, for instance, does not need deliberation when the logic intuition is stronger than the heuristic intuitive. For both views, however, the wrongful application of heuristics would be more prevalent when the controlling function of system II would fail as a consequence of the heat stress.[2]

We therefore expect that the distinct effect that heat has on cognition can be (partially) captured by the Kahneman framework. Recent research has investigated the effect on cognitive reflection (Chang & Kajackaite, 2019), but to date, no study has extended this investigation to the specific behavioural biased outcomes stemming from a predisposition to overly adhere to intuitive decision strategies. Although the CRT is highly correlated with specific behavioral biases, we test the effect of heat on bias sensitivity for an array of specific well-known biases directly. To our knowledge, no attempts have been made to distinguish the effects of heat on behaviour and cognition using this approach.

1.1.3 Temperature and Risk

Evidence suggests that temperature has a direct effect on the willingness to take risk. Wang (2017) shows that people making trading decisions will pursue high-risk high-yield options compared to a control condition.

Some indirect evidence on aggression also suggests that risky behaviour could follow from loss of control through the same channel. For instance, solely increasing the temperature makes people subjectively rate other people in the room to be more hostile (Anderson, Anderson, Dorr, DeNeve & Flanagan, 2000). Cao & Wei (2005) hypothesize that aggression leads to increased risk behaviour. Denson, DeWall and Finkel (2012) conclude that it is the loss of self-control that increases aggression. Finally, Frey, Pedroni, Mata, Rieskamp and Hertwig (2017) show self-control to be predictive of various risk behaviour outcomes. Overall, we expect the same channel that increases system I dependency will also increase risk-taking behaviour.

1.1.4 Temperature and Gender

Many individual characteristics mediate the effect heat has cognition, however, the heterogeneous gender-related differences stands out.[3] Biological research (Kingma & Van Marken Lichtenbelt, 2015), metabolic research (Byrne, Hills, Hunter, Weinsier & Schutz,

[2]We discuss the implications of this renewed model in light of our results in the limitations section.

[3]We extensively discuss the potential influence of other individual characteristics in the limitation section.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

2005), and psychological empirical research (Wyon, 1974) shows that hot temperatures have a distinctly different effect on women as compared to men. The most profound example of this distinction and its neglect in the past decade is the temperature comfort level. The 'default' room temperature level of 21° C seems mainly based on male preferences (Kingma & Van Marken Lichtenbelt, 2015). Indeed, anecdotal evidence suggests that women perform better at slightly higher default room temperatures (Chang & Kajackaite, 2019).

As such, finding the effects of adverse temperature on cognition would be incomplete without taking gender-specific preferences into consideration. Without correcting for gender, female preference or tolerance for higher temperatures might influence the overall findings regarding the effect of adverse temperatures on performance. Given that women show a preference for somewhat higher temperatures, women will rate identical absolute temperature increases (subjectively) as less adverse as compared to men. Performance for women might thus also be expected to be less affected by heat.

1.2 This study

We hypothesize that heat exposure will decrease cognitive performance such that biased behaviour will be more prominent, as rational correction will require more effort under heat stress. Heat is a salient factor in the working environment and workers can often elicit control over temperature themselves, making the relevance of our results apparent and immediately applicable. Moreover, by testing detectable temperature differences in each condition, we are able to assess the accuracy and thus relevance of self-reported comfort measures for in future research.

Additionally, we investigate the effect of heat on risk behavior. Through the same channel, we expect that a combination of lack of effortful control and bodily discomfort will increase risk behaviour. This would be in line with aggression studies (for instance, American football players commit more aggressive fouls; Craig, Overbeek, Condon & Rinaldo, 2016). We test both the general self-reported risk attitude, which has generally been claimed to be a rather stable character trait, unaffected by heat (Dohmen et al., 2011), and actual risk behaviour, which we expect to increase following indoor temperature manipulation (see, for example, Wang, 2017).

Our experimental design has several key advantages over current practices in the literature. First, we actively strive to control a variety of factors influencing the physical experience of the environment. That is, we pre-expose all participants to the temperature manipulation for a defined adjustment period of one hour before starting the tasks. All participants are wearing similar clothing provided specifically for the experiment. We further control for the outdoor temperature of the period before testing. Second, we keep all other indoor climate factors constant. For instance, we manipulate the temperature while keeping air ventilation levels unchanged. As a result, CO_2 levels, noise, lighting, and air refreshment are equal between manipulations. Some recent experiments manipulated temperature by opening and closing windows, without controlling for CO_2 and fine particles between

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

groups, and are therefore unable to isolate the effect of just temperature on task performance (Wang, 2017).

2 Method

2.1 Experimental conditions and design

We designed a controlled experiment to measure the effect of heat on decision quality. We employed a stratified random sampling method to recruit a total of 257 participants with an average age of 21.57 (SD = 2.41) years old using the Maastricht University Behavioral Experimental Economics laboratory database. Stratification ensures an equal gender distribution amongst manipulation groups. The final sample allows for a 10% deviation of gender within groups. All participants were proficient in reading and writing of the English language. Participants are randomly distributed to either the control or the experimental condition.[4] This between-subject design used temperature as the main independent variable. Given the clear gender differences in the temperature effect on performance and satisfaction in the literature, gender is the secondary independent variable in our analysis.

Participants were exposed to a controlled physical environment with either a hot temperature (28° C) or a neutral temperature (22° C). The decision for 28° C is derived from the body of literature focused on temperatures below 29° C / 85° F (for an overview, see Hancock & Vasmatzidis, 2003). More specifically, previous research repeatedly showed an effect of hot temperature on performance on neurobehavioural test at 27–28° C (Lan, Lian, Pan & Ye, 2009; Lan & Lian, 2009).[5] In these conditions, a battery of validated tests included cognitive reflection tasks, a heuristics battery, lottery risk tasks, and self-reported risk preferences. Additionally, participants state their personal comfort levels and their subjective estimation as to what extent the environment influences their performance on the battery of tasks. The experiment was programmed using Qualtrics Software (Qualtrics, Provo, UT) and executed at the Behavioral Experimental Economics lab facilities at Maastricht University in the Netherlands. The laboratory is approximately 5 meters wide and 20 meters long. In this room, there are 33 cubicles (approx. 1.0 meter by 1.5 meters), all including a computer and table, which are closed off by shutters. All participants are tested in groups varying between 25 and 30 participants per group. Air quality is controlled using a climate system that holds the air refreshment rate constant.[6] The control condition of 22 ° C is reached running only the climate system. The "hot" condition of 28 ° C is reached

[4]Appendix Table 4 Panel B summarizes individual characteristics per condition.

[5]As we discuss in the limitations section, we acknowledge that higher temperatures could show more profound effects. However, the goal of this paper is to generalize our results to the professional workforce. We argue that the relevance of excessive temperatures upwards of our threshold of 27–28° C will be exponentially decreasing with each increase in degree Celcius. Temperature measurements in real-life settings show repeatedly naturally occurring temperature variations of 28° C within one standard deviation of the mean, but rarely above 29° C (Künn, Palacios & Pestel, 2019; Zivin, Hsiang & Neidell, 2018).

[6]See Appendix Table 4 Panel A for an overview of the average CO_2 and humidity per condition.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

using five 3kW industrial heaters, each with a 115m^3 capacity. During the experiment, four heaters maintain a constant temperature. Manual adjustments to the thermostats of the individual heaters ensures a stable temperature. All heaters also ran without heating during the control condition, such that the noise produced by the heaters is constant between conditions.[7]

All participants were subject to strict clothing prescriptions. These requirements ensure that all participants have a similar physical experience of the heat. For instance, the possibility to remove layers of clothing could increase heterogeneity in the experienced heat within and between conditions. All participants are asked to wear long jeans. To fully ensure homogeneity, we provide all participants with long-sleeved black polyester thermoshirts. Participants are not allowed to wear anything underneath these shirts.[8]

Participants arrived in the laboratory at 11 AM, one hour before the start of the actual experiment. This adaption time ensured that all participants experience the indoor climate similarly, independent of the outdoor temperature or previous activity. During this adaption time, the temperature was kept at the same levels as during the experiment. After one hour, the test battery automatically started. All tasks were completed in English. Each task was presented to each participant only once. We did not impose a time schedule for the different tasks. The average completion time was roughly 45 minutes. Moreover, the outdoor temperature was measured on all testing days and compared between conditions. (Appendix Table 4 Panel A provides an overview of the indoor temperature during task and adaption, as well as the outdoor temperature between conditions.) The tasks were given in the order in which they are presented in Section 2.2. All tasks were presented to each participant only once.

2.2 Dependent measures

2.2.1 Performance measures

Cognitive Reflection Task: The classic Cognitive Reflection Task (CRT) by Frederick (2005) measures participants' propensity to rely on intuition or rational thinking. The test consists of three questions, of which each question has a salient intuitive answer and a correct rational answer. Each of these questions are scored with 1 for a correct response or 0 for an incorrect response. The score for this task is the number of correctly answered questions, such that the score of the CRT lies between 0 (no correct answers) and 3 (all

[7]Although individual preferences and satisfaction regarding illumination and acoustics will differ, the fact that all participants were exposed to the same conditions leads us to conclude that there is no objective reason why we would find a significant difference between the reported satisfaction on either of these variables between the control and manipulation group, on average.

[8]Women are allowed to wear bras underneath. We estimate that the clothing insulation value of all the subjects' ensemble is around 0.65 clo, on average (based on Owen, 2017). However, we note that our main purpose is to minimize variation between subjects and conditions. Therefore, the relative clo value of the clothing between groups is more relevant for the interpretation of the results than the absolute value of the ensemble.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

answers correct). Although this test is often used, Bialek & Pennycook (2017) find that multiple exposure does not reduce its validity.

Cognitive Reflection Task Expansion: To increase the probability of capturing the distinction between intuitive and rational thinking in our sample, we added an expansion of the original CRT. This test (from Toplak, West & Stanovich, 2014) consists of three additional items, following the same structure. It is highly correlated to the original CRT.

Heuristics Battery: The heuristic bias task battery by Toplak, West and Stanovich (2011) includes various questions about well-known economical biases. We select ten questions from this battery concerning casual base rate neglect, sample size problems, sensitivity towards regression to the mean, framing bias, outcome bias, the conjunction fallacy, probability matching, ratio bias, methodological reasoning, and the covariation problem.[9] Each of these questions are scored with 1 for a correct response or 0 for a biased and thus wrong response. The resulting score on this battery is thus between 0 and 10 points (M = 6.32, SD = 2.16), in line with the original authors.

2.2.2 Risk measures

Risk Elicitation Task: The first measure of risk assessment is aimed at inducing or eliciting actual risk behaviour at the time of the experiment. Similar to the original task of Holt and Laury (2002) we showed the participants nine choices between two sets of lotteries. The first lottery is of relatively low risk, where both the high and low payout options diverge only minimally (€6 versus €4.80, respectively). The second lottery can be considered high risk, as there is a strong divergence between the high (€11.55) and low (€0.30) payout option. For each consecutive choice, the probability of the high payout in both lotteries increases with 10%, such that in the first choice the probability of the high payout for each lottery is 10% and in the ninth and final choice this probability has become 90%. Note that the expected payout of the high-risk lottery surpasses the payout of the low-risk lottery from step 5 onwards (since then the expected payout is €5.93 for the high-risk versus €5.40 for the low-risk lottery). Participants are scored on a scale from 1–10, where the score reflects the switching point of the participants. Score 1 indicates a sustained preference for the high-risk lottery, labelling them as "risk-loving". A score of 5 implies risk-neutral behaviour, as participants follow the switching point in which both measures are equivalent. A score of 10 is assigned when participants never switch to the high-risk lottery. We label these participants as "risk averse". Depending on the risk preference, all scores are considered rational, as even in step 1 or 9 there is still a 10% probability of a high win or loss, respectively. This lottery is incentivised, and participants are told that one of the lottery choices will be played at the end of the questionnaire. The outcome of

[9]For an overview of these tasks, see Toplak et al., (2011)

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

their chosen lottery will be added to their total reimbursement. To make this incentive at least 25% of the total reimbursement, the lottery outcomes are multiplied by a factor from the original (Holt & Laury, 2002). Participants who switched their choice of lottery more than once were excluded from the sample; 34 observations were thus excluded (16 male, 18 female).[10]

Risk Attitude Task: In addition to a risk elicitation task, we asked participants how risk-loving they perceive themselves to be, both in general and on specific domains. Participants rated themselves on a 10-point scale, with the lowest score being risk-averse, and the highest score labelled fully prepared to take risk. First, all participants state to what extent they are willing to take risk or avoid taking risk generally as a person. Second, their willingness to take or avoid risk are specified for the following domains: driving, financial matters, leisure and sport, their occupation, health, and faith in other people. This approach has been extensively validated and found to correlate with actual risk behaviour (Dohmen et al., 2011; Falk, Dohmen & Huffman, 2016).

2.2.3 Indoor climate satisfaction

Self-reported Indoor Climate Satisfaction and Hindrance: Self-reported indoor environmental satisfaction was assessed by adapting the occupant indoor environment quality survey developed by Berkeley's Centre for the Built Environment (Huizenga, Abbaszadeh, Zagreus & Arens, 2006). For temperature, air quality, noise, and lighting, all participants are asked to rate their satisfaction level on a scale from 1 to 7. Additionally, for all these factors, participants are asked to what extent they perceive it as hindering or supporting their ability to answer the questions in the questionnaire on a similar 7 point scale. The scores are recoded such that a score of 7 indicates that the factor fully supports their ability, and a score of 1 indicates that the factor fully hinders their ability to answer the questionnaire. We label the totality of these factor-specific measures "satisfaction measures". In the analysis, we control for multiple testing.[11]

2.2.4 Additional checks

CRT multiple exposure check: After the three performance tasks (e.g., original CRT, extended CRT, and the Heuristics battery), all participants were asked to indicate whether they recognize any if these questions and if yes, whether they also remember the correct answer. These questions are scored by 1 – yes, 2 – no, or 3 – unsure.

[10]These participants were only excluded for the risk elicitation analysis. We found no indication that these participants were structural outliers throughout the task battery thus we did not conclude that their inconsistency in the risk elicitation task invalidated their scores for all other tasks.

[11]Multiple testing correction is applied for all 10 conditions using the Benjamini & Hochberg procedure (Benjamini & Hochberg, 1995), see Appendix Table 7. This procedure aims to control the false discovery rate whilst preserving relatively higher power compared to more conservative procedures (e.g., Bonferroni correction; Thissen, Steinberg & Kuan, 2002).

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

Clothing check: All participants were asked to indicate whether they are indeed wearing the thermoshirts provided by the experimenter.[12] On a Likert-scale of 1 (bad) to 7 (good), participants indicate the fit, length, and the comfort of the shirt. Additionally, we ask to what extent the shirt influences the performance on the tasks using the same scale.

Temperature: To be able to check for climate adjustment effects, three questions assessed the current and past climate experienced by the participants as well as their climate preference. Specifically, participants were asked to state in which country they grew up (most time spend until your 18th birthday), in which country they lived for the majority of the last five years, and what their preferred thermostat setting is (in degrees Celsius) in winter.

2.3 Incentives payoff

The payout was determined by adding the outcome of the preferred lottery of the risk elicitation task to the standard endowment of €15. The participants were told that for one of the steps, their chosen lottery will be played, but do not know which step this will be. The Qualtrics Internal Randomizer was used to draw an outcome (50/50 allocation) for the lottery chosen by the participant at step 5. The outcome was displayed at the end of the questionnaire. For the whole sample the average expected payoff of the risk task is 27% of the total payoff (with mean €5.98). No other performance tasks were incentivised, as these specific tasks are found not to be affected by incentives (Brañas-Garza, Kujal & Lenkei, 2019).

2.4 Statistical approach

To investigate statistical significance of the variables of interest, we ran mean comparison tests between the two manipulation conditions. Specifically, we conducted independent samples t-tests using STATA software (StataCorp, 2017). In situations when normality violations are detected (using Shapiro-Wilk normality tests), we tested for significance using Mann-Whitney U (Wilcoxon rank-sum) tests. For all results, we state whether parametric or nonparametric procedures are reported. Additionally, we apply the Benjamini & Hochberg procedure (Benjamini & Hochberg, 1995) as multiple testing correction when required.

[12]One of the participants indicated to be allergic to the fabric of the thermoshirts, and was thus asked to wear a similar (long-sleeved) shirt. All other participants wore the thermoshirts provided by the experimenter.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

3 Results

3.1 Descriptives and Condition Manipulations

The recorded sample consists of 257 students ranging from 17 to 31 years old, of which 53.5% are female (see Appendix Table 3).[13] The recorded indoor and outdoor climate conditions are reported in Appendix Table 4. The average temperature in the control condition was 22.4° C and in the hot condition 28.3° C. Levels of indoor CO_2, outdoor temperature of each test day during the morning, and outdoor temperature of the past three days do not differ significantly between manipulations.

3.2 Satisfaction measures

We first present the climate satisfaction measures in Table 1. Looking at the first column, it is confirmed that temperature (d= 0.77) and air quality (d= 1.53) are significantly less satisfactory in the hot condition. Additionally, both are predicted to hinder the performance on the performance measures. This confirms the notion that the high-temperature manipulation is considered uncomfortable.

Looking at the other indoor factors, and taking male and female participants together, we do not observe lighting satisfaction to be significantly different between conditions. The same holds for the effects of light on perceived performance. Similarly, we find no difference for noise satisfaction between conditions. However, it is reported to improve performance in the hot conditions. Here also, we note that noise was kept constant between conditions. Interestingly, participants actually predict noise to improve performance compared to the control condition. We suggest that in the control condition, when the heaters only produced noise, participants perceive the noise on its own as potentially hindering performance. In the hot conditions the noise of the heaters may be driven to the background by the more salient temperature. Also, in the hot condition there is a justification for the noise. Finally, clothing satisfaction and hindrance do not differ between conditions.

3.3 Gender Differences and Temperature

Following recent studies of gender differences and temperature effects on performance, we examine the satisfaction measures when controlling for gender. Interestingly, the general dissatisfaction and increased hindrance of temperature are reflected in our male sample only. These findings are presented in the middle two columns of Table 1. Our results are in line with Chang and Kajackaite (2019), such that males dislike hot temperatures and report to suffer more from heat as compared to women. This notion is further supported by the observation that temperature experience differs between genders when related factors do not. When we compare air quality satisfaction and its hindrance between the two conditions,

[13]The sample shows a average self-reported math proficiency of 63 on a scale from 0 to 100

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

TABLE 1: Main results of indoor variables: Self-reported indoor variables satisfaction and hindrance.

				Men			Women		
	Control	Hot	p-value	Control	Hot	p-value	Control	Hot	p-value
Temperature Satisfaction	4.66	3.50	.00***	5.13	3.05	.00***	4.25	3.90	.16
	(1.57)	(1.45)		(1.53)	(1.29)		(1.49)	(1.49)	
Air Quality Satisfaction	5.35	3.54	.00***	5.32	3.38	.00***	5.38	3.67	.00*
	(1.18)	(1.41)		(1.23)	(1.39)		(1.15)	(1.44)	
Light Satisfaction	5.33	4.95	.07	5.50	5.57	.56	5.19	4.42	.02*
	(1.46)	(1.64)		(1.55)	(1.03)		(1.39)	(1.88)	
Noise Satisfaction	5.36	5.57	.18	5.42	5.58	.58	5.30	5.55	.18
	(1.43)	(1.42)		(1.51)	(1.39)		(1.36)	(1.46)	
Clothing Satisfaction	5.71	5.55	.14	5.62	5.08	.02*	5.80	5.96	.81
	(1.36)	(1.27)		(1.37)	(1.33)		(1.37)	(5.96)	
Temperature Hindrance	4.68	3.40	.00***	5.27	3.05	.00***	4.17	3.71	.07
	(1.54)	(1.49)		(1.25)	(1.25)		(1.60)	(1.62)	
Air Quality Hindrance	5.07	3.71	.00***	5.03	3.65	.00***	5.10	3.75	.00*
	(1.23)	(1.45)		(1.25)	(1.23)		(1.22)	(1.63)	
Light Hindrance	5.02	4.95	.77	5.12	5.45	.37	4.94	4.52	.20
	(1.55)	(1.58)		(1.57)	(1.23)		(1.54)	(1.72)	
Noise Hindrance	4.94	5.36	.04	5.00	5.22	.52	4.88	5.48	.03*
	(1.69)	(1.59)		(1.77)	(1.65)		(1.64)	(1.53)	
Clothing Hindrance	3.68	3.74	.89	3.93	3.75	.17	3.46	3.74	.30
	(1.29)	(1.25)		(1.23)	(1.19)		(1.31)	(1.30)	
Observations	*129*	*129*		*60*	*60*		*69*	*69*	

Note: all scores are on 1-7 scale, and all scores are recoded such that 1 is bad or low, and 7 is good or high. Significance levels are based on nonparametric analysis. Standard deviation are given in parentheses. * indicates p < .05, ** p <.01, and *** p <.001, after multiple testing correction.

we find that both men and women dislike the hot temperature condition equally compared to the control condition. We note that additional (marginally) significant inconsistencies are seen for rating factors that are stable between conditions such as noise and light. Those discrepancies are correlated with the temperature manipulation (e.g., a potential demand effect; also see limitation section).[14]

Summarizing, we find that, as expected from the manipulations, temperature significantly lowers satisfaction and the perceived performance on the task, but only for the male

[14]The interaction between temperature manipulation and gender for both temperature satisfaction and temperature hinderance are both significant at p<001.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

sample. As such, as the commonly used hypothesis regarding the link between comfort and productivity predicts, we expect to find a decrease in performance on the performance measures for men, but not for women.

3.4 Performance Measures

Panel A of Table 2 shows the non-parametric results for the performance measures. We find no significant difference between control and hot conditions on any of the three performance measurements for the full sample. Only for women do we find a marginally significant difference (T=−1.75, p=0.08; *d*=0.30) between the performance on the CRT original between the control condition (M=1.26, SD=1.09) and the hot condition (M=1.61, SD=1.24).[15] Note that performance is increasing rather than decreasing. We conclude from these first results that the temperature has no direct effect on performance for men and women on our performance measures. If anything, we find weak support in line with Chang and Kajackaite (2019), as women seem to improve rather than decrease their performance on one of the three tasks in the hot temperature condition.[16]

3.5 Risk measures

Risk preference elicitation task. As expected from a strong body of research (for an overview, see Byrnes, Miller & Schafer, 1999), a baseline difference in risk behaviour is observed when comparing the control conditions as can be seen in Table 2, panel B. Based on parametric independent sample t-tests, men (M = 5.70, SD = 1.85) are significantly more risk-taking as compared to women (M = 6.48, SD =1.57; t = −2.42, p < 0.05; *d*=0.45), in line with the literature.

For the risk elicitation measure, participants in general do not differ between conditions. However, when we look at the gender subsamples, the picture changes. First, although men do not differ significantly in risk preference between conditions, women are significantly more risk loving in the hot condition (M = 5.61, SD = 1.89) compared to the control condition (M = 6.48, SD = 1.57; t = 2.75, p < .01 ; *d*= 0.50). As such, for women the risk and heat hypothesis appears to be a valid prediction.[17]

When comparing the risk preferences of women in the hot condition with the control condition of male risk preference, we observe that women do not only become more risk loving in a hot condition, but that their risk preference becomes equal to that of men in a normal control situation.

[15]For post-hoc effect size sensitivity analysis, see appendix Table 10

[16]The results do show a clear and significant difference in CRT performance between genders. These results are in line with earlier findings (Brañas-Garza et al., 2019; Zhang, Highhouse & Rada, 2016) and are suggested to be a result of gender difference in either math proficiency (for the self-reported math proficiency per gender, see Appendix Table 3; Welsh, Burns & Delfabbro, 2013) or math self-efficacy (Brañas-Garza et al., 2019).

[17]The interaction is significant at p<.01.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

TABLE 2: Main Results of Performance and Risk Measures

				Men			Women		
	Control	Hot	p-value	Control	Hot	p-value	Control	Hot	p-value
Panel A. Performance Measures									
CRT original	1.67	1.76	.49	2.13	1.95	.34	1.26	1.61	.08
(scored 0-3)	(1.61)	(1.56)		(1.07)	(1.03)		(1.09)	(1.24)	
CRT Extended	1.53	1.71	.21	1.85	2.03	.33	1.26	1.42	.37
(scored 0-3)	(1.09)	(1.07)		(1.04)	(1.02)		(1.07)	(1.03)	
Heuristics Battery	6.34	6.26	.86	7.33	6.83	.18	5.48	5.83	.32
(scored 0-15)	(2.22)	(2.11)		(2.12)	(1.98)		1.93)	(2.13)	
Observations	*129*	*128*		*60*	*59*		*69*	*69*	
Panel B. Risk Behaviour Elicitation									
Risk Elicitation	6.11	5.90	.45	5.70	6.29	.12	6.48	5.61	.01*
(scored 1-10: 1 = extremely	(1.74)	(1.99)		(1.85)	(2.05)		(1.57)	(1.89)	
risk-loving, 10 = extremely risk averse)									
Observations	*111*	*113*		*53*	*51*		*58*	*62*	
Panel C. Self-reported Risk Attitude									
General Risk Attitude	5.77	5.43	.12	6.08	5.40	.03*	5.49	5.46	.97
(scored 1-10: 1 = risk-averse,	(1.91)	(1.75)		(1.80)	(1.77)		(2.00)	(1.74)	
10 = fully prepared to take risk)									
Observations	*129*	*128*		*60*	*59*		*69*	*69*	

Note: For all panels except C, all significance levels are based on parametric analysis. For panel C, significance levels is based on nonparametric analysis. Standard deviation are given in parentheses. * indicates p < .05, **, p <.01, and *** p <.001

General risk attitude. For the general risk attitude question "Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?" (See Table 2, panel C), men report to be less prepared to take risk when asked in a hot condition (Mdn = 6.5) compared to the control condition (Mdn = 6; z=2.1, p < .05; d=0.38).[18] This is surprising, as we explicitly ask participants to reflect on their general risk attitude. This question has repeatedly shown to be stable over time and context independent, and as such, is supposed to be a stable predictor for risk behaviour. Women do report a stable attitude independent of conditions.[19]

[18]Note that the risk aversion scores are inverse for both measures: In the general attitude measurement, a low score equates risk aversion, whereas in the risk elicitation measure, a high score shows a late (or no) switch to the risky lottery, synonymous for risk averse behaviour according to the authors of the measure.

[19]When verifying the predictive power of the general risk attitude question with the risk behaviour as suggested by Falk, Dohmen and Huffman, (2016), we find that in our sample the general risk attitude is not

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

When looking at the domain-specific risk attitudes, only one differs significantly between conditions: Men predict to be less risky on work-related issues in a hot condition (Mdn=6) compared to the control (Mdn=6.5; z =2.19 p=0.028; d=0.42) condition. [20] For an overview of these results, see Appendix Table 5. This result remains significant when applying the Benjamini-Hochberg rank-dependent multiple testing correction (Benjamini & Hochberg, 1995) on the critical p-value threshold with a Q (false discovery rate) of 15%.[21]

4 Discussion

The increasing frequency of heatwaves, and outside temperatures that used to be exceptional, raises important questions about the impact of temperature on human performance. Of course, outdoor temperature does not need to be harmful given the mitigating effect of buildings, acting as a "shield" against temperature changes and pollution. There is evidence of a positive effect of building quality on human performance and productivity (e.g., Palacios, Eichholtz & Kok, 2020). But research measuring indoor climate also shows negative performance effects resulting from exposure to adverse indoor conditions (e.g., Künn, Palacios & Pestel, 2019; X. Zhang, Wargocki, Lian & Thyregod, 2017). Given that we spend roughly 90% of our time indoors, the effect of these adverse conditions warrants research. Understanding the effects of indoor temperature on human performance is crucial in determining and optimizing the daily indoor environment in work places and beyond.

The focus of this study is twofold: First, we assess the effect of hot temperatures on decision quality, and second, we answer the question whether peoples' stated experiences regarding these temperatures are related to this decision quality. In this study, we assessed the effect of adverse temperature by manipulation of the indoor temperature to 28° C over a two-hour period, compared to a control temperature of 22° C.

From the expectation that rational decision-making would suffer under adverse temperatures, more reliance on intuition would lead to a lower score on the Cognitive Reflection Task and to more biased responses in the Heuristic Battery. However, no significant difference on performance between the hot and control conditions were identified in this study. When looking at risk, a factor often associated with decisional quality and furthermore proposed to be correlated with the intuition-rational trade-off (Leith & Baumeister, 1996), we observe only an increase of risk preference in hot conditions for women.

Comparing these results with self-reported measures show some essential discrepancies. First, in our sample, only men find the hot condition significantly less satisfactory as

correlated with risk behaviour. Moreover, we find a negative correlation in the control condition between self-reported risk attitude and risk behaviour (see Appendix Table 6). These results do not support the validity of the self-reported risk attitude as a proxy for risk behaviour. However, we only find a marginally significant interaction between temperature manipulation and gender for self-reported general risk attitude with p=.08.

[20] The interaction is significant at p<.05.

[21] McDonald (2014) claims that a Q between 10% and 20% would entail relevant results, and underline that Q should not be mistaken for a P-value. For an overview of the critical value for 15% False Discovery Rate (Q) per rank used see Appendix 7.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

compared to the control condition. Women do not seem to make a distinction between conditions. Furthermore, when asking to what extent temperature has an influence on performance, men predict that the hot temperature significantly hinders their performance. Again, women do not make this distinction.

The discrepancy between self-report and actual behaviour is of crucial importance for the literature regarding the effects of indoor climate. Currently, self-reported measures are commonly used as a proxy for performance or productivity, yet this study shows that men are consistently overestimating the effect of adverse temperatures on performance. First, the discrepancy between the actual performance outcomes and the perceived hindrance from adverse temperature for men shows that men would have expected to have performed better in the control condition, which they did not. If policy makers would have assessed this self-perceived hindrance only, they might have spent significant effort and resources to improve indoor temperature conditions. In our study, however, we show that this would not result in an actual increase in performance.

On the domain of risk, we find that men assess their own daily willingness to take risk in general and in work situations to decrease when they are asked about this in the hot condition. This is surprising, since this measure is aimed at assessing the general self-reported risk preference, independent of any manipulation, and would thus be expected to be stable across conditions. For women, no significant difference between conditions is found. As for actual risk behavior, we find no difference between conditions for men.

These results have at least two implications for future indoor temperature (and indoor climate) research. First, we repeatedly find inconsistencies between the self-reported and actual effects of the indoor climate on performance. Specifically, men are overestimating the negative effect the temperature has on their performance. This shows that the use of self-reported measures as a proxy for actual performance is unreliable. Future research should focus on more direct measures of human performance and productivity than self-reported indoor climate satisfaction. Second, our research supports the recent findings of Chang and Kajackaite (2019) that gender plays a moderating part in the effect of temperature on performance. This underlines the conclusion from Kingma & Van Marken Lichtenbelt (2015) that one universal temperature standard does not fit the whole population. Gender differences have to be taken into account in any situation when we include temperature as an influential factor.

4.1 Limitations

Three specific limitations are worth discussing. First, a multitude of factors could mediate our results. We control for many relevant variables, yet we cannot exclude the possibility that some factors confound our results. According to Zhang, De Dear and Hancock (2019), the following factors should be considered regarding the effect of the thermal environment on performance:

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

Environment-related factors include intensity and duration of the indoor environment. We carefully control temperature and keep all other relevant factors constant between conditions. We include an adaption time that extends the total exposure time beyond most comparable studies. However, it is possible that higher temperatures would lead to differences in performance on the (heuristics) tasks battery (Parsons, 2014). For instance, Zhang, De Dear and Hancock (2019) found that reasoning declines from temperatures upwards of 28° C. We justify our decision for the temperature levels based on earlier research and our goal to generalize our finding to a realistic working environment of high skilled workers. By doing so, we inevitably limit the external validity of our results for higher temperatures. Finally, although we measure a multitude of variables between conditions (see Appendix Table 4), unobserved variables could inadvertently influences the results.

Performance-related factors include all individual factors such as age, gender, skill level, acclimation level, and emotional state. We control for individual differences between groups regarding gender, math skill, education level, age, and thermostat preference (see Appendix Table 4). We apply random sampling to counter unobserved variables, such as emotional state, to distort our results. The sample size is limited as the adaption (or acclimation) time required takes more resources than in comparable studies. However, we are confident that addressing the exposure time is a key advantage of our experiment relative to the current literature. Regarding participant age, the sample mainly consists of students around the age of 22 (M = 21.57, SD = 2.41). We attempted to recruit an age category representing an older population (older than 50), but recruitment turned out to be difficult. Moreover, the level of English language skills and task comprehension forced us to exclude a significant part of the successfully recruited "older" sample. The educational background of the majority of our sample (Business and Economics students) increased the likelihood of recognition of the type of tasks we assessed, and previous exposure to these constructs can influence results (we will discuss the results of multiple exposure to the CRT test below). Usage of the relatively unfamiliar extension of the CRT (Toplak et al., 2014) and an unfamiliar heuristic battery (Toplak et al., 2011) at least partially alleviates this concern.

Task-related factors include the complexity and the type of task presented to the participant. Since all participants are performing the same tasks, no confounding effect of task type and complexity is to be expected. However, a new view on the underlying mechanism of the dual process model could explain why we do not find an effect of temperature on cognitive performance using our heuristics battery. De Neys and Pennycook (2019) suggest that the deliberate system is activated only when there is a clear conflict between a heuristic reaction and a logical reaction. It is possible that the nature of our task battery elicits either a intuitive responses or a logical solution, but without a conflict between these two. The lack of conflict, according to De Neys and Pennycook, will not reveal any potential restrictions in the deliberate system because this system is not involved in the response. We deliberately test an extensive battery of well-known heuristic problems which should increase the likelihood of conflicts in which the deliberate system is active. However, we

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

cannot fully excluded the possibility that the lack of conflict (partially) explains why we find no difference between the two groups. We encourage further research to assess both neurological measured deliberate system activation as well as the level to which these tasks present an implicit conflict between logic and intuitive response.

Second, participants likely change behaviour in anticipation of the effect of the manipulation, which is unavoidable in an experiment with temperature manipulation. All participants in the manipulation conditions (e.g., the "hot" temperature condition), are instantly aware of this manipulation when entering the laboratory. To create uniformity between groups and take away emphasis on the temperature, we asked participants in all conditions to wear a provided shirt, and in both conditions the industrial heaters were on. Moreover, the indoor climate quality scale was not limited to temperature, but included other important indoor climate variables, reducing the emphasis on temperature. However, when the participants were asked to state what they thought the experiment was about, they indeed stated (in the manipulation condition) that temperature and task performance was the major aim of the experiment. In the control condition, less than 10% stated temperature to be a decisive factor (popular guesses included the influence of "clothing" or "noise" on performance).

Finally, the choice for our test battery is the outcome of a careful trade-off between practical and theoretical considerations. Research has suggested that the CRT is robust under multiple exposure (Bialek & Pennycook, 2017; Meyer, Zhou & Frederick, 2018) and consistent over time (Stagnaro et al., 2018). Recognition of the original CRT is relatively high (46% recognized at least one question, and 20% recognized all questions) .[22] For the extended CRT questions, however, only 13% recognized one or more questions. The fact that we observe no difference in performance between the classic and extended CRT supports the notion that these levels of recognition and recollection of answers do not affect the results of this study.

Welsh et al. (2013) propose that the CRT merely reflects mathematical skills. In our sample we see that self-reported math skills differ significantly between genders. Women report a proficiency of 59.07 out of 100, whereas males report 67.48 out of 100 ($p < .001$). We indeed find that in the total sample, men outperform women in the CRT. However, this does not affect the result in the sense that we analyse the effect of temperature on performance specifically within gender. We furthermore find no interaction between math proficiency and the effect of temperature on the CRT. Nevertheless, we cannot exclude that the risk assessment is effected by the difference in math proficiency.

5 References

Anderson, C. A., Anderson, K. B., Dorr, N., DeNeve, K. M., & Flanagan, M. (2000). Temperature and aggression. In *Advances in Experimental Social Psychology* (pp. 63–133).

[22]For an overview of CRT and CRT extension recognition and recollection, see Appendix Table 9

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

https://doi.org/10.1016/S0065-2601(00)80004-0.

Baumeister, R., Bratslavsky, E., Muraven, M., & Tice, D. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*(3), 774–789. https://doi.org/10.1037/0022-3514.74.5.1252.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*(6), 351–355. https://doi.org/10.1111/j.1467-8721.2007.00534.x.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bialek, M., & Pennycook, G. (2017). The cognitive reflection test is robust to multiple exposures. *Behavior Research Methods*, 1–7. https://doi.org/10.3758/s13428-017-0963-x.

Bluyssen, P. M. (2013). *The Healthy Indoor Environment: How to Assess Occupants' Wellbeing in Buildings.* Routledge. https://doi.org/10.4324/9781315887296.

Brañas-Garza, P., Kujal, P., & Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics*, 82. https://doi.org/10.1016/j.socec.2019.101455.

Byrne, N. M., Hills, A. P., Hunter, G. R., Weinsier, R. L., & Schutz, Y. (2005). Metabolic equivalent: One size does not fit all. *Journal of Applied Physiology, 99(3)*, 1112–1119. https://doi.org/10.1152/japplphysiol.00023.2004.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 125*(3), 367-383

Cao, M., & Wei, J. (2005). Stock market returns: A note on temperature anomaly. *Journal of Banking and Finance, 29*(6), 1559–1573. https://doi.org/10.1016/j.jbankfin.2004.06.028.

Carter, E. C., Kofler, L. M., Forster, D. E., & Mccullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource *Journal of Experimental Psychology: General, 144*(4), 796–815. https://doi.org/10.1037/xge0000083.supp.

Chang, T. Y., & Kajackaite, A. (2019). Battle for the thermostat: Gender and the effect of temperature on cognitive performance. *PLoS ONE, 14*(5). https://doi.org/10.1371/journal.pone.0216362.

Cheema, A., & Patrick, V. M. (2012). Influence of warm versus cool temperatures on consumer choice: A resource depletion account. *Journal of Marketing Research, 49*(6), 984–995. https://doi.org/10.1509/jmr.08.0205.

Craig, C., Overbeek, R. W., Condon, M. V., & Rinaldo, S. B. (2016). A relationship between temperature and aggression in NFL football penalties. *Journal of Sport and Health Science, 5*(2), 205–210. https://doi.org/10.1016/j.jshs.2015.01.001.

Cunningham, M. R., & Baumeister, R. F. (2016). How to make nothing out of something: Analyses of the impact of study sampling and statistical interpretation in misleading meta-analytic conclusions. *Frontiers in Psychology, 7*(1639). https://doi.org/10.3389/fpsyg.2016.01639.

De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science, 28*(5), 503–509. https://doi.org/10.1177/0963721419855658.

Denson, T. F., DeWall, C. N., & Finkel, E. J. (2012). Self-control and aggression. *Current Directions in Psychological Science, 21*(1), 20–25. https://doi.org/10.1177/0963721411429451.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association, 9*(3), 522–550. https://doi.org/10.1111/j.1542-4774.2011.01015.x.

Eichholtz, P., Holtermans, R., & Kok, N. (2019). Environmental performance of commercial real estate: New insights into energy efficiency improvements. *The Journal of Portfolio Management, 45*(7), 113-129.

Engelbert, M., & Carruthers, P. (2010). Introspection. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(2), 245–253. https://doi.org/10.1002/wcs.4.

Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation - Advances in Research and Theory, 44*, 145–199. https://doi.org/10.1016/S0079-7421(03)44005-X.

Evans, J. S. (2007). On the resolution of conflict in dualprocess theories of reasoning. *Thinking and Reasoning, 13*(4), 321–329.

Falk, A., Dohmen, T., & Huffman, D. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *IZA Discussion Paper*, No. 9674(9674). http://ftp.iza.org/dp9674.pdf.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42. https://doi.org/10.1257/089533005775196732.

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances, 3*(10). https://doi.org/10.1126/sciadv.1701381.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692.

Grether, W. F. (1973). Human performance at elevated environmental temperatures. *Aerospace Medicine, 44*(7), 747–755. http://www.ncbi.nlm.nih.gov/pubmed/4715089.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W.

K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., . . . Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*(4), 546–573. https://doi.org/10.1177/1745691616652873.

Hancock, P. A., & Vasmatzidis, I. (1998). Human occupational and performance limits under stress: The thermal environment as a prototypical example. *Ergonomics, 41*(8), 1169–1191. https://doi.org/10.1080/001401398186469.

Hancock, P. A., & Vasmatzidis, I. (2003). Effects of heat stress on cognitive performance: The current state of knowledge. *International Journal of Hyperthermia, 19*(3), 355–372. https://doi.org/10.1080/0265673021000054630.

Hermalin, B. E., & Weisbach, M. S. (1991). The effects of board composition and direct incentives on firm performance. *Financial Management, 20*(4), 101-112. https://doi.org/10.2307/3665716.

Hockey, R. (2013). *The psychology of fatigue: Work, effort and control.*Cambridge University Press.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644–1655. https://doi.org/10.1257/000282802762024700.

Huizenga, C., Abbaszadeh, S., Zagreus, L., & Arens, E. (2006). Air quality and thermal comfort in office buildings: Results of a large indoor environmental quality survey. *Proceedings of Healthy Buildings, 3*, 393–397. https://doi.org/10.12659/PJR.894050.

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating Models of Self-Regulation. *Annual Review of Psychology, 72*, 319–345. https://doi.org/10.1146/annurev-psych-061020-105721.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs.

Kingma, B., & Van Marken Lichtenbelt, W. (2015). Energy consumption in buildings and female thermal demand. *Nature Climate Change, 5*(12), 1054–1056. https://doi.org/10.1038/nclimate2741.

Künn, S., Palacios, J., & Pestel, N. (2019). The impact of indoor climate on human cognition: Evidence from chess tournaments. *IZA Discussion Paper*, 12632.

Lan, L., Lian, Z., Pan, L., & Ye, Q. (2009). Neurobehavioral approach for evaluation of office workers' productivity: The effects of room temperature. *Building and Environment, 44*(8), 1578–1588. https://doi.org/10.1016/j.buildenv.2008.10.004.

Leith, K. P., & Baumeister, R. F. (1996). Why do bad moods increase self-defeating behavior? Emotion, risk taking, and self-regulation. *Journal of Personality and Social Psychology, 71*(6), 1250–1267. https://doi.org/10.1037/0022-3514.71.6.1250.

Lin, H., Saunders, B., Friese, M., Evans, N. J., & Inzlicht, M. (2020). Strong effort manipulations reduce response caution: A preregistered reinvention of the ego-depletion paradigm. *Psychological Science, 31*(5), 531–547. https://doi.org/10.1177/0956797620904990.

MacLeod, C. M. (1991). Half a century of reseach on the stroop effect: An integrative

review. *Psychological Bulletin, 109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163.

MacNaughton, P., Satish, U., Laurent, J. G. C., Flanigan, S., Vallarino, J., Coull, B., Spengler, J. D., & Allen, J. G. (2017). The impact of working in a green certified building on cognitive function and health. *Building and Environment, 114*, 178–186. https://doi.org/10.1016/j.buildenv.2016.11.041.

McDonald, J. H. (2014). *Handbook of Biological Statistics (3rd ed.).* Sparky House Publishing.

Meyer, A., Zhou, E., & Frederick, S. (2018). The non-effects of repeated exposure to the cognitive reflection test. *Judgment and Decision Making, 13*(3), 246–259.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*(2), 247–259. https://doi.org/10.1037/0033-2909.126.2.247.

Nakamura, M, Yoda, T, Crawshaw L.I., Yasuhara S., Saito Y., Kasuga M., Nagashima K., & Kanosue K. (2008). Regional differences in temperature sensation and thermal comfort in humans. *Journal of Applied Physiology 105*(6), 1897-1906.

Owen, M. (Ed.). (2017). *ASHRAE Handbook - Fundamentals.* American Society of Heating, Refrigerating and Air-Conditioning Engineers Inc.

Palacios, J., Eichholtz, P., & Kok, N. (2020). Moving to productivity: The benefits of healthy buildings. *PLoS ONE, 15*, https://doi.org/10.1371/journal.pone.0236029.

Parsons, K. (2014). *Human thermal environments: The effects of hot, moderate, andcold environments on human health,comfort, and performance.* CRC Press https://doi.org/10.1201/b16750.

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and Buildings, 40*(3), 394–398. https://doi.org/10.1016/j.enbuild.2007.03.007.

Perry, J. L., & Porter, L. W. (1982). Factors affecting the context for motivation in public organizations. *Academy of Management Review, 7*(1), 89–98. https://doi.org/10.5465/amr.1982.4285475.

Satish, U., Mendell, M. J., Shekhar, K., Hotchi, T., Sullivan, D., Streufert, S., & Fisk, W. J. (2012). Is CO2 an indoor pollutant? direct effects of low-to-moderate CO2 concentrations on human decision-making performance. *Environmental Health Perspectives, 120*(12), 1671–1677. https://doi.org/10.1289/ehp.1104789.

Shibasaki, M., Namba, M., Oshiro, M., Kakigi, R., & Nakata, H. (2017). Suppression of cognitive function in hyperthermia; From the viewpoint of executive and inhibitive cognitive processing. *Scientific Reports, 7*(1), 1-8https://doi.org/10.1038/srep43528.

Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). Performance on the cognitive reflection test is stable across time. *Judgment and Decision Making, 13*(3). https://doi.org/10.2139/ssrn.3115809.

StataCorp. (2017) Stata Statistical Software: Release 15. *College Station, TX: StataCorp*

LLC.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics, 27*(1), 77–83. https://doi.org/10.3102/10769986027001077.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory and Cognition, 39*(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking and Reasoning, 20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124.

Van Lange, P. A. M., Rinderu, M. I., & Bushman, B. J. (2017). Aggression and violence around the world: A model of CLimate, Aggression, and Self-control in Humans (CLASH). *Behavioral and Brain Sciences, 40*, e75. https://doi.org/10.1017/S0140525X16000406.

Van Ooijen, A. M. J., Van Marken Lichtenbelt, W. D., Van Steenhoven, A. A., & Westerterp, K. R. (2004). Seasonal changes in metabolic and temperature responses to cold air in humans. *Physiology and Behavior, 82*(2–3), 545–553. https://doi.org/10.1016/j.physbeh.2004.05.001.

Vermeulen, W. J. V., & Hovens, J. (2006). Competing explanations for adopting energy innovations for new office buildings. *Energy Policy, 34*(17), 2719–2735. https://doi.org/10.1016/j.enpol.2005.04.009.

Wageman, R., & Baker, G. (1997). Incentives and cooperation: The joint effects of task and reward interdependence on group performance. *Journal of Organizational Behavior, 18*(2), 139–158. https://doi.org/10.1002/(SICI)1099-1379(199703)18:2<139::AID-JOB791>3.0.CO;2-R.

Wang, X. (2017). An empirical study of the impacts of ambient temperature on risk taking. *Psychology, 08*(07), 1053–1062. https://doi.org/10.4236/psych.2017.87069.

Welsh, M. B., Burns, N. R., & Delfabbro, P. H. (2013). The Cognitive Reflection Test: How much more than Numerical Ability? *35th Annual Conference of the Cognitive Science Society, 35*(35), 1587–1592.

Wright, K. P., Hull, J. T., & Czeisler, C. A. (2002). Relationship between alertness, performance, and body temperature in humans. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 283*(6), R1370–R1377. https://doi.org/10.1152/ajpregu.00205.2002.-Body.

Wyon, D. P. (1974). The effects of moderate heat stress on typewriting performance. *Ergonomics, 17*(3), 309–317. https://doi.org/10.1080/00140137408931356.

Zhang, D. C., Highhouse, S., & Rada, T. B. (2016). Explaining sex differences on the

Cognitive Reflection Test. *Personality and Individual Differences, 101*, 425–427. https://doi.org/10.1016/j.paid.2016.06.034.

Zhang, F., & De Dear, R. (2017). University students' cognitive performance under temperature cycles induced by direct load control events. *Indoor Air, 27*(1), 78–93. https://doi.org/10.1111/ina.12296.

Zhang, F., De Dear, R., & Hancock, P. (2019). Effects of moderate thermal environments on cognitive performance: A multidisciplinary review. *Applied Energy, 236*, 760–777. https://doi.org/10.1016/j.apenergy.2018.12.005.

Zhang, F., Haddad, S., Nakisa, B., Rastgoo, M. N., Candido, C., Tjondronegoro, D., & de Dear, R. (2017). The effects of higher temperature setpoints during summer on office workers' cognitive load and thermal comfort. *Building and Environment, 123*, 176–188. https://doi.org/10.1016/j.buildenv.2017.06.048.

Zhang, X., Wargocki, P., Lian, Z., & Thyregod, C. (2017). Effects of exposure to carbon dioxide and bioeffluents on perceived air quality, self-assessed acute health symptoms, and cognitive performance. *Indoor Air, 27*(1), 47–64. https://doi.org/10.1111/ina.12284.

Zivin, J. G., Hsiang, S. M., & Neidell, M. (2018). Temperature and human capital in the short and long run. *Journal of the Association of Environmental and Resource Economists, 5*(1), 77–105. https://doi.org/10.1086/694177.

Zivin, J. G., & Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review, 102*(7), 3652–3673. https://doi.org/10.1257/aer.102.7.3652.

Appendix

TABLE 3: Sample Descriptive Statistics

					Male (43%)				Female (57%)				
	Mean	Min	Max	N	Mean	Min	Max	N	Mean	Min	Max	N	p-value
Age	21.57 (2.41)	17	31	257	21.70 (2.36)	19	31	119	21.46 (2.45)	17	31	138	0.34
Math Proficiency	62.97 (17.9)	1	100	257	67.48 (16.42)	2	100	119	59.07 (18.26)	1	88	138	0.00***
Thermostat Preference	21.91 (2.65)	12	28	235	20.94 (2.74)	12	28	106	21.58 (2.55)	12	27	129	0.02*

Note. Statistics presented are mean values and standard deviation are presented in parentheses. Math Proficiency is on a 0–100 scale. Thermostat Preferece is in °C, in winter. Extreme thermostat preferences were excluded (below zero degrees and above 30 degrees). p-values results from nonparametric independent sample t-tests. * indicates p-vale < .05, ** a p-value <.01, and *** a p-value <.001.

TABLE 4: Descriptive statistics per condition.

	Control	Hot	p-value
Panel A. Indoor and Outdoor Conditions			
Indoor Temperature during Task (°C)	22.44	28.65	*.00****
Indoor Temperature during Adaption (°C)	22.07	28.01	*.00****
Indoor Temperature Average (°C)	22.21	28.25	*.00****
Indoor CO2 (ppm)	692.12	726.93	*.72*
Indoor humidity (%)	48.87	39.06	*.00***
Outdoor (°C) temperature at start of the experiment	13.88	14.65	*.66*
Average outdoor (°C) temperature (three days average)	14.44	13.84	*.79*
Panel B. Individual Characteristics			
Age	21.43	21.71	*.36*
Math Proficiency (1-100 scale)	63.49	62.44	*.64*
Thermostat Preference (°C, in winter)	21.32	21.27	*.89*
Education Level (0-5 scale)	2.92	3.01	*.58*

Note. Statistics presented are mean values and standard deviation are presented in parentheses. Panel A describes the indoor and outdoor climate conditions. ppm stands for particles per million. Panel B describes the individual characteristics per condition. Thermostat Preference stated is in winter conditions. Education level in on a 0 to 5 scale, where 0 is without high school diploma, and 5 is completed masters diploma.p-values results from parametric independent sample t-tests. * indicates p < .05,**, p <.01, and *** p <.001.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

TABLE 5: Multiple testing correction Panel A and Panel C for 15% false discovery rate level

				Men			Women		
	Control	Hot	p-value	Control	Hot	p-value	Control	Hot	p-value
Panel A. Self-reported Risk Attitude									
General	5.77	5.43	.12	6.08	5.40	.03*	5.49	5.46	.97
	(1.91)	(1.75)		(1.80)	(1.77)		(2.00)	(1.74)	
Driving	3.39	3.16	.35	4.20)	3.48	.07	2.68	2.87	.70
	(2.24)	(2.38)		(2.50)	(2.47)		(2.13)	(2.29)	
Financial Matters	5.31)	5.11	.47	5.80	5.73	.88	4.88	4.57	.31
	(2.24)	(2.18)		(2.38)	(2.23)		(2.03)	(2.00)	
Sports and Leisure	7.85	7.65	.58	7.93	7.52	.37	7.77	7.77	.95
	(2.13)	(2.38)		(1.95)	(2.35)		(2.28)	(2.41)	
Work	6.72	6.45	.20	7.03	6.18	.02*	6.45	6.68	.67
	(2.18)	(2.16)		(1.99)	(2.06)		(2.32)	(2.23)	
Health	4.64	4.17	.18	4.73	4.28	.23	4.57	4.07	.49
	(2.75)	(2.67)		(2.41)	(2.72)		(3.03)	(2.65)	
Others (social)	6.55	6.58	.98	6.43	5.85	.23	6.65	7.22	.30
	(2.53)	(2.56)		(2.27)	(2.52)		(2.75)	(2.44)	
Observations	*129*	*128*		*60*	*59*		*69*	*69*	

Note: All scores are on 1-10 likert scale, and all scores are recoded such that 1 is risk averse, and 10 is risk loving. Significance levels are based on nonparametric analysis. Standard deviation are given in parentheses. * indicates p < .05, ** p <.01, and *** p <.001.

TABLE 6: Correlation Matrix between the risk attitude measure and the risk behaviour measure.

	Full sample			Control			Hot		
	M	SD	1	M	SD	1	M	SD	1
1. Risk Elicitation Task (Holt & Laury, 2002)	6.05	1.91		6.11	1.76		6	2.05	–
2. General Risk Attitude (Dohmen et al., 2011)	5.65	1.83	−.12 [−.25, .01]	5.78	1.90	−.05 [−.24,.13]	5.52	1.77	−.20* [−.37,−.01]
Observations	*224*			*111*			*113*		

Note. The Risk Elicitation task has missing values, the summary statistics excluded all risk attitude cases that are matched to missing values for the risk task. Correlation coefficient presented is the Spearman's rho and 95% confidence interval in brackets. * indicates p < .05, ** p <.01, and *** p <.001.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

TABLE 7: Multiple testing correction Panel A and Panel C for 15% false discovery rate level.

			Men		Women	
	p-value	Q = 15%	p-value	Q = 15%	p-value	Q = 15%
Panel A. Self-reported Indoor Variables Satisfaction and Hindrance						
Temperature Satisfaction	.00	Sig	.00	Sig	.16	
Air Quality Satisfaction	.00	Sig	.00	Sig	.00	Sig
Light Satisfaction	.07	Sig	.56		.02	Sig
Noise Satisfaction	.18		.58		.18	
Clothing Satisfaction	.14		.02	Sig	.81	
Temperature Hindrance	.00	Sig	.00	Sig	.07	Sig
Air Quality Hindrance	.00	Sig	.00	Sig	.00	Sig
Light Hindrance	.77		.37		.20	
Noise Hindrance	.04	Sig	.52		.03	Sig
Clothing Hindrance	.89		.17		.30	
Panel C. Self-reported Risk Attitude						
Driving	.35		.07		.70	
Financial Matters	.47		.88		.31	
Sports and Leisure	.58		.37		.95	
Work	.20		.02	Sig	.67	
Health	.18		.23		.49	
Others (social)	.98		.23		.30	

Note. The p-value are the result of nonparametric ranksum tests as shown in table 3. The chosen levels of False Discovery Rates (Q) are chosen given that Q=15% implies less than 1 FDR per 7 tests. Q=5% is the most conservative FDR rate, with the highest risk of False Negatives (McDonald, 2014). Applying the FDR formula (False Discovery Rate = Expected (False Positive / (False Positive + True Positive))) to the risk domain entails that the change of two significant findings amongst 7 domains would be 28.6%. We find two significant findings (in the male sample) if we correct for a FDR as low as 15%. The significance of the general risk attitude in the male sample is robust against a FDR of 12%.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

TABLE 8: Critical value for 15% false discovery rate (Q) per rank used for multiple testing correction. The critical p-value thresholds according to the Benjamini & Hochberg (1995) are dependent on the total amount of multiple tests. According to their rank, each level of significance will be compared to their rank critical value as stated in this table. The 7 items critical value are applied to the Self-Reported Risk Attitude (table 5, panel C), the 10 items critical values are applied to the Self-reported Indoor Variables Satisfaction and Hinder (table 5, panel A).

Rank	7 items	10 items
1	0,025	0,015
2	0,050	0,030
3	0,075	0,045
4	0,100	0,060
5	0,125	0,075
6	0,150	0,090
7		0,105
8		0,120
9		0,135
10		0,150

TABLE 9: Overview of percent recognition and answer remembering for the CRT Classic and CRT Extention. (N=257).

		Recognize Question	Remember the Answer *		
		Yes	Yes	No	Unsure
CRT Original	Lily pads	45.52	42.54	44.03	13.43
	Widget problem	26.85	26.04	59.38	14.58
	Bat and ball	40.86	45.30	38.46	16.24
CRT Extended	Class ranking	2.72	6.38	78.72	14.89
	Stock market	5.45	58.33	16.67	25.00
	Barrel of water	10.89	7.02	77.19	15.79

Note. *The percentage in the remembering column is conditional on recognition. For example: For the Lilypads, of the 45.52% that recognizes the questions, 44.03 % does not remember the answer.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press

TABLE 10: Post-hoc sensitivity analysis.

		Sample Size		Effect Size d	
		Control	Hot	Non-Parametric Mann-Whitney	Parametric T-Test
Majority Measures	Full Sample	129	128	0.42	0.41
	Men	60	59	0.62	0.61
	Women	69	69	0.58	0.56
Risk Elicitation Task	Full Sample	111	113	0.45	0.44
	Men	53	51	0.66	0.65
	Women	58	62	0.62	0.60

Note. Effect size sensitivity is reported per groupsize. The first rows apply to the majority of all presented results in the paper. Only for the risk elicitation task, the latter rows applies, due to some exclusion cases in that sample. We present for each sample-size sensitivity estimates for parametric as well as non-parametric tests.

https://doi.org/10.1017/S1930297500007828 Published online by Cambridge University Press