

## Stereotypes Concerns and Discreet Existence of Differences between Men and Women in Risk-Aversion – a Replication Study



Magdalena Adamus<sup>1</sup> , Matúš Grežo<sup>2</sup> , Katarína Dudeková<sup>2</sup> 

<sup>1</sup> Department of Public Economics, Faculty of Economics and Administration, Masaryk University in Brno, Czech Republic

<sup>2</sup> Institute of Experimental Psychology, Centre of Social and Psychological Sciences, Slovak Academy of Sciences, Bratislava, Slovak Republic

The present research conceptually replicates and extends the results of a study on the relation between individuals' sex, their risk attitudes and stereotype threat (Carr & Steele, 2010). The authors reported that differences between men and women in risk aversion emerged only after activating negative stereotypes about women's performance in mathematics. A total of 321 Slovaks, randomly assigned to control or experimental treatments, answered questions on their risk aversion, anxiety, analytical reasoning and gender self-concept. We expected to observe differences between men and women only after activating stereotypes. Aware of the issues with the consistency of different risk aversion measures, we investigated whether the effect of stereotype threat on risk aversion differs across three different risk aversion measures. Additionally, we explored whether this effect depends on how the stereotype threat is activated (explicit vs. implicit activation). Finally, to explain the mechanism through which stereotypes foster women's risk aversion, we explored the moderating effect of gender self-concept and mediating effects of anxiety and analytical reasoning on the relationship between stereotype threat and risk aversion. In general, the study found no differences between men and women in risk aversion and did not replicate the original effect of stereotype threat on risk aversion.

*Key words:* risk aversion, sex/gender differences, stereotype threat, replication study

### Introduction

For decades, risk taking was believed to be an element of masculinity (Byrnes, Miller, & Schafer, 1999), where masculinity is tightly

related to manliness or – in other words – to being a man. Wilson and Daly (1985) claimed that favourable risk attitudes are part of the young male syndrome strengthened by the competition between peers for scarce, yet prestigious goods. In line with sociobiological

Correspondence concerning this article should be addressed to Magdalena Adamus, Department of Public Economics, Faculty of Economics and Administration, Masaryk University in Brno, Lipová 41a, 602 00 Brno-střed, Czech Republic. E-mail: lena.adamus@econ.muni.cz

Received June 5, 2019



and evolutionary theories, economic research presents women as systematically and consistently less inclined to engage in activities involving risk-taking (Charness & Gneezy, 2012; Eckel & Grossman, 2002; 2008). Croson and Gneezy (2009, p. 20) even concluded that there are “fundamental differences between men and women” in risk attitudes. However, when Hyde (2005) investigated meta-analyses of studies related to sex differences in cognitive abilities, communication, social and personality variables, wellbeing and happiness, motor skills and others, she found that most traits show considerable similarity between men and women. Specifically, 78% of investigated effect sizes were close to zero or small, indicating that similarities between men and women are so great that the differences are virtually meaningless and provide no valuable information at the individual level. In line with these results, recent economic literature shows that the differences between men and women in their risk attitudes are negligible (Nelson, 2012a, 2018). And yet, people tend to exaggerate the differences, believing that women are not only more risk averse than men but also that women are more risk averse than they actually are (Eckel & Grossman, 2002). As Lemaster and Strough (2014, p. 149) observed, “research that focuses solely on biological differences between men and women is even more pronounced than differences” actually are.

Although the idea that women are more risk averse than men has “became accepted as a truism” (Nelson, 2018, p. 3), little research is done to explain or interpret sources of the alleged differences. Even fewer studies investigate their actual scope. Apart from evolutionary or sociobiological theories, differences in risk preferences are sometimes explained with the content of testosterone in saliva or prenatal testosterone exposition (Apicella et al., 2008; Sapienza, Zingales, & Maestriperi,

2009). However, the link between hormones or brain structures and behaviour is not yet well-recognised and established. Therefore, as Maney (2016) noticed, interpolating these observations on behavioural differences is premature and unjustified. Finally, Nelson (2018) observed that if differences between men and women in their risk preferences were essential or natural, they should be also constant across cultures. Rare economic research that focuses on cultural aspects indicates that culture can interfere with women’s and girl’s risk preferences. For example, Booth and Nolen (2009) showed that girls in single-sex schools are equally willing to risk as boys from both single-sex and co-ed schools. Henrich and McElreath (2002) found no gender differences in any of the non-WEIRD cultures they studied. Gneezy, Leonard, and List (2009) compared behaviour in a risk-related task in a matrilineal Khasi culture with a patriarchal Maasai and found no differences between men and women. Some researchers claim that there are no differences between men’s and women’s risk preferences and there is only a “white male effect” (Finucane, Slovic, Mertz, Flynn, & Satterfield, 2000; Kahn, Braman, Gastil, Slovic, & Mertz, 2013). The authors suggest that white men clearly stand out from other social and ethnic groups.

The present study builds on the results reported by Carr and Steele (2010, Study 2a). The authors tried to explain gender differences in risk and loss attitudes by examining the effect of stereotype threat activation. They reported that differences in risk attitudes between men and women became apparent only after activating gender stereotypes. Specifically, activating stereotypes in a women sample increased their risk and loss aversion, resulting in significant differences between men and women. The main aim of the present study is to investigate the effect of stereotype threat on individual’s risk aversion, conceptu-

ally replicating the protocol used by Carr and Steele (2010). Additionally, although much has been written about the role of stereotype threat in various aspects of life (Steele, Spencer, & Aronson, 2002), it is still unclear how and under what conditions stereotypes interfere with reasoning and decision-making. In order to understand the mechanisms through which stereotypes affect risk aversion, we extend the study by Carr and Steele (2010) by focusing on the investigation of situational explanatory factors in four ways.

First, we extend previous research by introducing three different measures of risk aversion. Carr and Steele (2010) used a method proposed by Porcelli and Delgado (2009) to elicit risk preferences and a well-known measure of loss aversion adapted from Gächter, Johnson, and Hermann (2007). However, previous studies (Pedroni et al., 2017) showed that research on financial risk attitudes can be distorted by the application of specific methods which often lead to inconsistent results. It means that not only the size of an individual's risk-aversion differs but also his/her relative position to others within the same sample changes when a different risk elicitation method is applied. In this study, thus, we aim to use three risk aversion measures (hypothetical investment task, questionnaire and hypothetical lotteries) and then explore whether the risk aversion measured by these methods is affected by stereotype threat in a consistent manner.

Second, Seibt and Forster (2004) observed that under specific conditions, particularly in tasks unrelated to academic performance, the more explicit stereotype priming – i.e., pointing to differences between groups – could motivate threatened individuals to disconfirm negative stereotypes. Consequently, to explore whether the effect of stereotype threat on risk aversion differs across these different conditions, our study included im-

PLICIT and explicit modes of stereotype threat activation.

Third, it has been speculated that stereotypes can make individuals more risk averse by activating prevention focus through increasing anxiety (making individuals more careful and analytical) or leading to ego depletion (suppressing analytical thinking and inducing intuitive reasoning) (Steele, Spencer, & Aronson, 2002). Carr and Steele (2010) showed that the effect of stereotypes was mediated by ego depletion. Alternatively, Seibt and Forster (2004) indicated that negative stereotypes activate prevention focus making individuals more risk averse but also more analytical. Extending this research, we aim to investigate whether analytical reasoning and anxiety mediate the effect of stereotype threat on risk aversion.

Finally fourth, driven by the results reported by Meier-Pesti and Penz (2008), the present study investigates the role of psychological traits stereotypically associated with masculinity or femininity in risk preferences. The study contributes to the literature by exploring not only the relation between biological sex and risk-taking but also the question of how self-ascribed gender self-concept shapes the effect of stereotype threat on individual's risk aversion.

## **Theoretical Background and Hypotheses**

### **Research on Women's Risk-Taking**

The tradition of research on risk taking goes back to the 1950s and it belongs to the most important phenomena studied by economists. In one of his most influential papers, Arrow (1951) claimed that the presence of risk is an essential element of capitalistic economy. With the development of behavioural and experimental economics providing innovative and interdisciplinary tools for studying the

phenomenon, the research flourished. Risk aversion along with the differences between men and women in their risk preferences became one of the prime topics. Soon after studies on risk aversion had become widespread, numerous researchers started to report differences between men and women. Byrnes, Miller, and Schafer (1999) analysed 150 studies studying various types of risk. The oldest studies they included in the meta-analysis dated back to the 1960s. The authors cautiously concluded that their results generally support the view that – in line with sociobiological explanations – women are more risk averse than men are. However, they also observed that only 48% of studied effects were larger than 0.20 and the differences depended largely on the context of the study and age of participants. With increasing age, men and women tended to become more similar. Later, however, meta-analysis was used – unjustly – to support more definite claims that women are consistently more risk averse than men across various context and task types (cf. Croson & Gneezy, 2009; Nelson, 2018).

The results serve also to explain differences in outcomes in various decisions related to economics and finances. Numerous studies use risk aversion as an explanation of women's less optimal choices in areas such as investing in stocks, education, health, remuneration schemes, and finally choices related to self-employment or starting a business. Bonin, Dohmen, Falk, Huffman, and Sunde (2007) confirmed that individuals with low willingness to take risk have lower-paid occupations. Dohmen et al. (2011) investigated choices of various payment schemes in relation to productivity and risk aversion, and found that women preferred fixed-pay instead of more competitive and performance-related incentive schemes. The authors claimed that women's lower risk tolerance can be, thus, responsible for the wage gap. More generally, Eckel

and Grossman (2002) suggest that although women's higher risk aversion might have been adaptive in the past, in the modern society it makes them disadvantaged. Specifically, since risk is considered ubiquitous in managerial decisions, women tend to be considered less adequate candidates for such posts (Adams & Funk, 2012; Eckel & Grossman, 2002). Furthermore, women, by default, receive more conservative investment recommendations with lower return rates resulting in lower values of their asset portfolios (Agnew, Anderson, Gerlach, & Szykman, 2008; Nelson, 2018). Consequently, since women tend to, on average, live longer than men, they face greater risk of poverty during their retirement (Lemaster & Strough, 2014).

Although psychology sees risk-taking as an ambivalent feature – it may be adaptive but excessive risk taking may also involve harmful consequences – economists have unambiguous views on the phenomenon. A specific feature of economic research is the belief that lower risk aversion is generally a positive trait (Nelson, 2018, p. 118). Consequently, mainstream economists encourage women to adopt more masculine risk preferences to adjust to demands of the modern economy (Eckel & Grossman, 2002). After 2008, some scholars identified excessive risk-taking as one of the causes of the global financial crisis and suggested that the presence of women in executive boards would have prevented this crash (Kristof, 2009; Nelson, 2012b; Nigel, 2009). Women, with their more cautious and risk averse attitudes, would invest in options that yield lower but more secure and stable profits. However, Nelson (2012a, 2018) criticised these suggestions and indicated that previous research provides at best mixed or inconclusive findings on the support of such claims.

In her seminal paper, Nelson (2012a) introduced a similarity index indicating that men and women have generally similar risk-pref-

erences. She found that similarity between men's and women's risk preferences – understood as a similarity of the distributions of risk aversion – ranged from at least 60 to over 90%. The figure means that when guessing whether a randomly selected man takes more risk than a randomly selected woman – i.e., using sex a proxy for risk preferences – we would be correct only 55% of times. Hardly a progress from tossing a coin. Nelson's meta-analysis can be considered a turning point, after which an increasing number of studies end with a conclusion that there are no significant differences between men and women in their attitudes toward risk. Drawing on the current knowledge about women's risk preferences and risk aversion measures, the present study applies 3 different risk aversion measures to verify findings of Nelson (2012a). Specifically, we expected that without activating negative stereotypes, women and men would express similar risk aversion in decision-making (H1, see Table 1 for hypotheses and research questions overview).

### **Stereotype Threat and Decisions Involving Risk**

The extant literature attributes differences between men and women in risk aversion – if significant – either to biological, social, or situational causes. The third type of factors, situational, receives less attention. Apart from the measure, there are other situational threats that can distort an individual's performance in risk-related tasks, including, particularly, stereotype threat. Steele and Aronson (1995) defined stereotype threat as a condition when individuals, aware of the existence of negative stereotypes about a group they belong to, are afraid to confirm the stereotype with their own behaviour. The authors of the concept claimed that under specific conditions, negative stereotypes may have dis-

ruptive effects on individuals' performance. Individuals need not to believe in the stereotypes, it is sufficient they are aware that such negative stereotypes exist. Although most studies on stereotype threat are related to academic performance, Carr and Steele (2010) extended the approach to financial decision making. In two studies they investigated how activation of negative stereotypes of women's mathematic skills would affect their decisions in risk- and loss-related tasks.

In both studies, stereotypes were activated by informing participants that the test was diagnostic of their mathematic, logical and rational reasoning competence. Immediately after priming, participants specified their biological sex. In all studies, the authors found evidence that stereotype threat increased women's loss and risk aversion. In the control conditions, there were no differences between men and women. Consequently, driven by these results, we hypothesized that women experiencing stereotype threat would be more risk averse compared to men (H2) and also compared to women not threatened by stereotypes (H3).

The extant literature provides substantial information about risk aversion measures inconsistency (e.g., Pedroni et al., 2017). However, little is known about the interaction between specific methods and stereotype threat. It seems likely that risk preferences elicited with various measures might be susceptible to stereotype threat to varying degrees. Consequently, we put here forward two research questions: Without activating negative stereotypes, is there any difference in risk aversion of women and men between different types of risk aversion measures? (RQ1), and Does the effect of stereotype threat on risk aversion vary across different types of risk aversion measures? (RQ2) We intend to investigate these questions in the exploratory analysis.

Kray, Thompson, and Galinsky (2001) demonstrated that explicitly activated negative stereotypes about women's performance in negotiations lead women to adopt strategies aimed at disconfirming the stereotypes (*stereotype reactance*). Explicit stereotype activation methods are those which openly state that, in a given task type, the stereotyped group – for example women – performed worse. Conversely, implicit methods describe a stereotyped task – for example mathematical – but refrain from indicating which group of individuals is expected to perform better. Implicit methods, however, may also indicate that individuals who tend to be successful in a given type of tasks are those who possess certain characteristics, for example, being rational and assertive. The characteristics are stereotypically related to members of the positively stereotyped group. Seibt and Forster (2004) observed that the reactive response to stereotypes may emerge in tasks, where individuals have cognitive resources to overcome the alleged deficits. In academic performance tests, this is rather unlikely. However, in tasks related to attitudes, explicit activation of stereotypes may lead to behaviour directly opposing the stereotypes. Carr and Steele (2010, p. 1412) used an implicit method, informing participants that the tasks measure their “mathematical, logical, and rational reasoning abilities”.

The present study investigates potential differences between the two methods – implicit and explicit – as well as the relation between the activation methods and specific risk aversion measures. However, since previous results are inconclusive, we formed no hypotheses and intended to delve into the issue in the exploratory analysis. Consequently, we aimed to answer the question of whether the effect of stereotype threat on risk aversion depends on the stereotype threat condition (implicit or explicit) (RQ3).

### **Mediators of the Relationship between Stereotype Threat and Risk Aversion**

Although several studies confirmed the effect of stereotype threat on academic achievement, the knowledge about the precise mechanisms through which stereotypes impair this achievement remains vague. Steele and Aronson (1995) speculated that the stereotype threat effect might be mediated by various factors including distraction, narrowed attention, anxiety, self-consciousness, withdrawal of effort, and even excessive effort. So far, the strongest evidence was related to the mediating role of anxiety (Spencer, Steele, & Quinn, 1999; Steele, Spencer, & Aronson, 2002) but not easy but even this effect proved to be weak.

Carr and Steele (2010) showed that the effect of stereotype threat on risk aversion was mediated by ego depletion as measured by a Stroop task. According to dual process theories, each individual has two separate systems responsible for decision making processes. System 1 involves automatic, heuristic responses, while System 2 – although slower – is associated with more analytic reasoning in decision making. Ego depletion is a state, in which cognitive resources are limited and an individual is more likely to rely on fast heuristics instead of analytic reasoning (System 1 rather than System 2). On the contrary, Seibt and Forster (2004) indicated that negative stereotypes activate prevention focus. Individuals affected by negative stereotypes become more risk averse and focus on avoiding errors. This implies that stereotype threat activates analytic reasoning through prevention focus (i.e., activating System 2).

Both papers refer to the dual nature of the mind, however, explanations the papers offer are contradictory – under stereotype threat, either System 1 or System 2 is pronounced. In order to have a deeper insight into the mechanisms driving the stereotype threat influence on

risk aversion, we investigated mediating effects of state anxiety (RQ4) and analytical reasoning (RQ5). Following the study by Spencer et al. (1999), we first investigated whether activation of negative stereotypes increases anxiety among threatened individuals and whether state anxiety could be reliably pinned as a mediator of the effect. Secondly, if Seibt and Forster (2004) are correct, then activation of negative stereotypes would prompt individuals to process tasks more analytically. If threatened individuals' responses are more analytic than non-threatened individuals, that would speak in favour of System 2 being activated and indicate that prevention focus might be responsible for channelling the effect of stereotype threat on risk aversion.

### Sex and Gender in Research on Risk Aversion

In line with the Hyde's (2005) findings, Reis and Carothers (2014) observed that in most cases, psychological research does not support the taxonic view of differences between men and women. The taxonic view implies that when we know only that an individual belongs to a specific taxon (group), we are able to infer other properties of the individual fairly precisely. On the other hand, the differences between men and women are dimensional (Carothers & Reis, 2013; Reis & Carothers, 2014). This means that men and women do not form two distinct taxa. Although men and women may differ, the magnitude of the differences vary across traits and individuals. Consequently, predictions about individual's characteristics tend to be inaccurate if based on biological sex.

In contrast to the vast literature indicating differences between men and women's risk preferences (Byrnes, Miller, & Schafer, 1999; Charness & Gneezy, 2012; Croson & Gneezy, 2009; Eckel & Grossman, 2002; 2008) and a considerable number of studies on biological determinants of these differences (Apicella et al., 2008; Coates, Gurnell, & Rustichini, 2009;

Dreber & Hoffman, 2010; Schipper, 2012), economic research on psychological and social determinants of risk aversion is rare. The economic literature points to *gender differences*, including differences in risk aversion, but tends to use terms *sex* and *gender* interchangeably. This approach suggests that an individual's sex is a good approximation of their risk aversion. However, the practice has some deficiencies. Firstly, it assumes that the differences are large, which we now know is not accurate. Secondly, it supports the belief that men and women differ "naturally" – by the virtue of sex (Nelson, 2018), which is also incorrect. Not only does this approach disregard the impact of confounding variables, such as prescriptive norms, it also obscures the extent to which men and women are similar, as well as the nature and the true size of the differences.

Consequently, instead of being understood as descriptive, differences between men and women are tamed as self-explanatory in economics. Apart from statistic tools recommended by Maney (2016) and Nelson (2018), one of the methods to avoid confusion when reporting results on differences between men and women is to clearly distinguish between *sex* and *gender*. Despite the calls to replace biological sex with aspects of gender identity as potential behaviour predictors (Unger, 1979), such research is scarce or rather virtually absent in economics. Lemaster and Strough (2014) were able to identify only two studies exploiting the perspective of gender as a social construct in risk-related research (Demaree, DeDonno, Burns, Feldman, & Everhart, 2009; Meier-Pesti & Penz, 2008)<sup>1</sup>.

<sup>1</sup> The authors actually mention three studies. However, the third one (Kastlunger, Dressler, Kirchler, Mittone, & Voracek, 2010) investigated the relation between masculinity and femininity, and tax compliance in experimental settings. Although it involved measurement of gender-role orientation by Personal Attributes Questionnaire (PAQ), the study did not report any results related to risk preferences.

Meier-Pesti and Penz (2008) investigated actual investment behaviour, decisions in a hypothetical investment task, responses to Waerneryd's scale assessing investment behaviour and a questionnaire eliciting general risk attitudes (Waerneryd, 1996). Except for the hypothetical task, men reported greater willingness to take risk than women. However, the differences disappeared, when the masculinity subscale of Bem Sex Role Inventory (Bem, 1974) was introduced as a covariate in the analysis. Although, masculinity mediated the effect of sex on the willingness to take risk, the femininity subscale proved to be unrelated to risk preferences. In their second study, males primed with typically masculine roles were more risk prone than males in the control treatment and those primed with stereotypically feminine roles. The priming had no effect on women's risk preferences. Interestingly, although masculinity was related to both men and women risk preferences, there were no sex differences in the masculinity subscale of the Bem Sex Role Inventory. Demaree et al. (2009) established that trait dominance – believed to be related to masculinity – predicted risk-taking behaviour in a

hypothetical task involving choices between sure options and gambles. In accordance with recent literature on gender, Lemaster and Strough (2014) defined gender as a multidimensional phenomenon. Consequently, they applied a series of instruments measuring gender identification, gender typicality and gender-role orientation. As a measure of risk preferences, they applied a survey intended to measure risk tolerance of potential investors. They found that individuals who evaluated themselves as more masculine were more risk tolerant. The effect, however, was stronger for men than for women. Inspired by the results reported by Meier-Pesti and Penz (2008) and corroborated more recently by Lemaster and Strough (2014), we aimed to answer the question of whether the gender self-concept moderates the effect of stereotype threat on risk aversion (RQ6).

## Methods

### Participants and Data Collection

To ensure a sufficient statistical power of this study for both replicating prior findings of

Table 1 *Hypotheses and research questions overview*

H1	Without activating negative stereotypes, women and men express similar risk aversion in decision-making.
H2	Women experiencing stereotype threat are more risk averse compared to men.
H3	Women experiencing stereotype threat are more risk averse compared to women not threatened by stereotypes.
RQ1	Without activating negative stereotypes, is there any difference in risk aversion of women and men between different types of risk aversion measures?
RQ2	Does the effect of stereotype threat on risk aversion vary across different types of risk aversion measures?
RQ3	Does the effect of stereotype threat on risk aversion depend on the stereotype threat condition (implicit or explicit)?
RQ4	Does anxiety mediate the effect of stereotype threat on risk aversion?
RQ5	Does analytical reasoning mediate the effect of stereotype threat on risk aversion?
RQ6	Does gender self-concept moderate the effect of stereotype threat on risk aversion?



Carr and Steele (2010) and conducting our exploratory analyses, we have performed an a priori power analysis. The analysis was based on alpha level of .05, a power of .95 and lower bound of a medium effect size ( $f^2 = .06$ ; see Cohen, 1988, further details for power analysis can be found in supplementary materials at Open Science Framework – OSF).<sup>2</sup> Based on the a priori power analysis result +10%, we collected data from 321 Slovaks ( $n_{\text{control group}} = 108$ ;  $n_{\text{experimental group 1}} = 104$ ;  $n_{\text{experimental group 2}} = 109$ ). The age of participants ranged between 18 to 81 ( $M = 44.86$ ;  $SD = 16.26$ ). The research sample representatively involved individuals from all Slovak regions and of all education levels (see further details in OSF supplementary materials).

The data were collected by an external agency through an online survey hosted on Qualtrics. Application of the computer-based survey allowed for the elimination of the experimenter effect. The data collection was governed by the ESOMAR code.<sup>3</sup> The rules guarantee that participants are randomly selected in accordance with predetermined criteria from the agency's database and do not participate in a research more than twice (or in specific conditions three times) a month. Participants were not deceived at any point. Furthermore, the study was carried out in accordance with ethical principles introduced by the American Psychological Association. Participants were informed about their right to remain anonymous and to withdraw from the study at any time. The design was parsimonious and included only the data necessary to verify the hypotheses we put forward. All data were stored with due diligence and used only for the purposes directly related to the present study. All materials are available at the Open Science Framework (OSF). The

samples were balanced in terms of biological sex and age. The agency provided sufficient incentives for participants consistent with local market conditions (either cash or vouchers). Furthermore, the agency was responsible for preliminary checks of the data quality (response times and completeness).

All materials have been translated into Slovak by natives and back translated to check for translation accuracy. The research protocol was constructed so that each of the tasks related to measured variables was compulsory. Participants were not allowed to proceed with the study unless they provided a response. Consequently, we obtained no incomplete data. The questionnaire contained two control questions (attention checks) such as "If you read this sentence, press 4". Individuals who failed to select correct answers, were considered as potentially contaminating the data set and were not included in the analysis.

### Study Design and Procedure

The study involved a between-subject design: 3 (implicit stereotype activation, explicit stereotype activation, and a control condition without any stereotype threat) x 2 (biological sex). Similarly as in the original study (Carr & Steele, 2010), in the control condition participants were informed that they are participating in a psychological study on decision making involving solving some simple puzzles. In the implicit stereotype activation condition, we repeated the manipulation used by Carr and Steele (2010) – informing participants that the tasks measure their "mathematical, logical, and rational reasoning abilities". In the explicit stereotype activation condition, we additionally mentioned that the tasks used in the study showed gender differences in the past with men performing, on average, better than women. Participants were not informed that the study investigated differ-

<sup>2</sup>[https://osf.io/pvke7/?view\\_only=cd4a55ceb0174fd8bf3037ede424d81b](https://osf.io/pvke7/?view_only=cd4a55ceb0174fd8bf3037ede424d81b)

<sup>3</sup><https://www.esomar.org/what-we-do/code-guidelines>

ences between men and women. Participants were randomly assigned to conditions, while specific measures were taken to ensure that the samples are gender- and age-balanced (through randomization and filters available in Qualtrics). In the control condition, instructions were followed by the measures of State Anxiety Inventory, risk aversion measures (hypothetical investment task, questionnaire and hypothetical lotteries), analytic reasoning (Cognitive Reflection Test), measure of gender self-concept and socio-demographic questions. In experimental conditions, the question about a participant's biological sex was placed between the stereotype-invoking priming and the anxiety inventory to strengthen the impact of the priming. Other than this, the conditions followed the same pattern (see Figure 1). The order of risk aversion measures as well as CRT tasks was randomised in Qualtrics.

## Variables and Measures

### Dependent Variables

*Risk aversion* was measured by three risk elicitation methods (REM). First, we used a single *hypothetical investment task* (REM1) adapted from Dohmen, Falk, Huffman, and Sunde (2012). Participants were asked to imagine that they won €100,000 in a lottery. Immediately after picking up the prize, a renowned bank offered them the opportunity to double the sum in two years. However, there was a possibility that they would lose half of the invested sum. Participants indicated how much of the €100,000 they would invest. Options range from 1 (0 €) to 6 (100,000 €). The higher the score, the lower the individual's risk aversion. We have chosen this measure because of its simplicity. The task is easily understood even by participants not trained in economics, which is particularly important since more complex methods – such as gambles – can

pose considerable difficulties and, consequently, participants are likely to provide random answers.

Second, we measured *investment risk attitudes* (REM2) using the Waerneryd's (Waerneryd, 1996) scale composed of 6 items (e.g., "I think it is more important to have safe investment and guaranteed returns, than to take a risk to have a chance to get the highest possible returns" or "If I think an investment will be profitable, I am prepared to borrow money to make this investment"). The items are rated on a 7-point Likert scale from 1 (*Strongly disagree*) to 7 (*Strongly agree*). Three items are reverse-coded. The scale score ranges from 6 to 42. The McDonald's omega test showed a poor reliability of this scale ( $\omega = .66$ ;  $SE = .03$ ; 95% CI [.59, .72]). Since omega-if-item-deleted data showed that there could be only negligible increase in reliability ( $\omega = .67$ ), we have decided to use all items for our analyses. However, such a low reliability suggests that one should be very cautious when interpreting results brought by using this scale.

Third, the present study replicates the method used in the original study, i.e. *hypothetical lotteries* (REM3) adapted from Porcelli and Delgado (2009). Participants were presented with a series of 14 choices between paired lotteries in the gain domain, involving two options with equal expected value but different probability of winning (e.g., "You can choose between two fair lotteries. One of the lotteries offers a 20% chance of winning 4 euros. The other an 80% chance of winning 1 euro. Which of the two lotteries do you chose to participate in?"). The risk aversion score was calculated as the number of lower-risk options chosen by a participant.

### Independent Variables and Controls

*Anxiety* (ANX) was measured by a short state version of the Spielberger State-Trait Anxiety

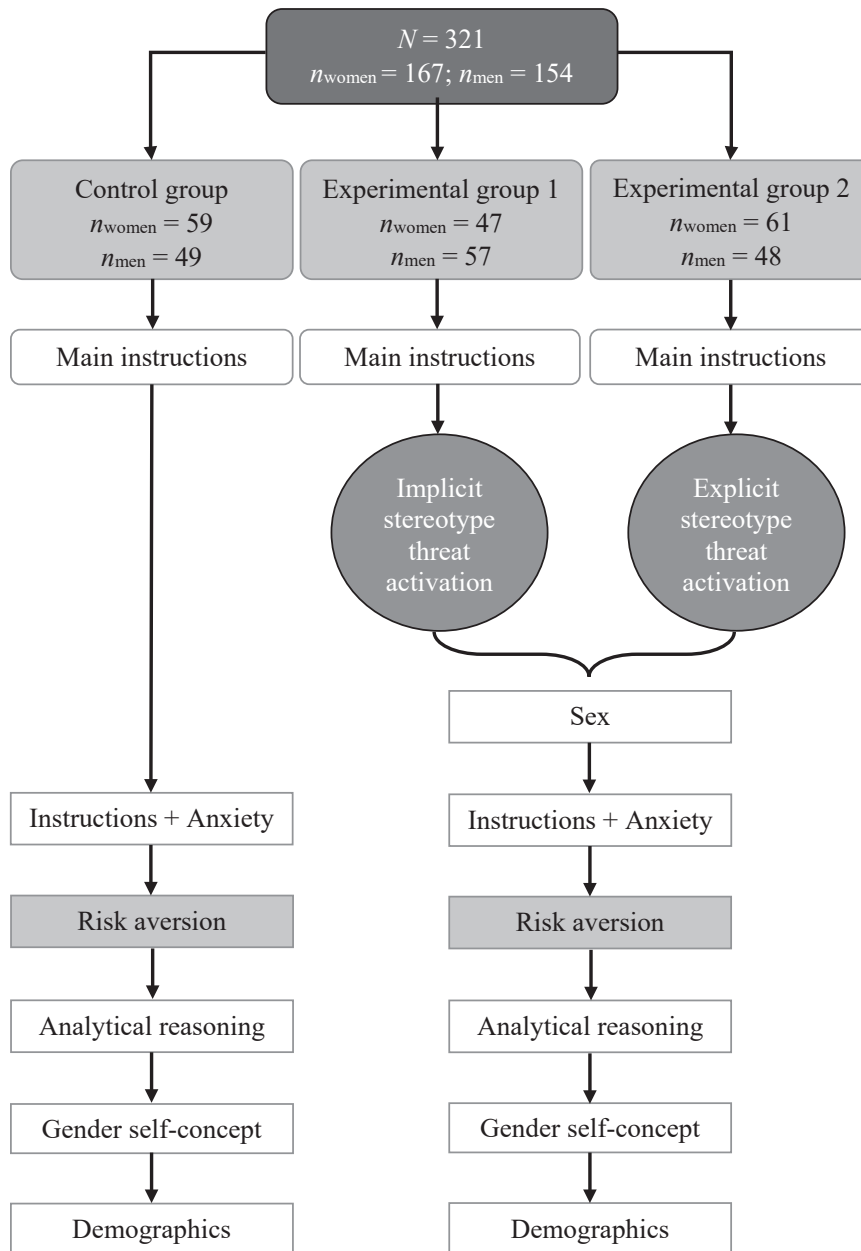


Figure 1 Logical flow of the study

Inventory (STAI<sub>short</sub>) adapted from Marteau and Bekker (1992). The measure is composed of six items rated on a 4-point scale from 1 (*not at all*) to 4 (*very much*). Since we are interested in examining the current state of experienced anxiety during testing, we used only state version of this inventory. The instructions asked participants to describe their current feelings indicating to which extend they feel calm, tense, upset, relaxed, content and worried. Questions 1, 4, and 5 are anxiety-absent items and are reversed in the analysis. The McDonald's omega test showed a good reliability of this scale ( $\omega = .87$ ;  $SE = .01$ ; 95% CI [.84, .89]).

*Analytical reasoning* was measured using an extended version of the Cognitive Reflection Test (CRT) adapted from Sirota et al. (2018). This test consists of three mathematical and five verbal open-ended questions. Exemplary items are: "Coffee and milk cost 1.20€. The coffee costs 1 euro more than the milk. How much costs the milk?" (mathematical reasoning) or "Mary's father has 5 daughters but no sons – Nana, Nene, Nini, Nono. What is the fifth daughter's name probably?" (verbal reasoning). The score was calculated as a number of correct answers. The higher the score, the more analytical the participants' reasoning.

*Gender self-concept* was measured using a short version of the Bem Sex Role Inventory (BSRI<sub>short</sub>, Bem, 1974), which is a 12-item version of the original questionnaire with six items related to traditionally feminine traits (affectionate, sympathetic, sensitive to needs of others, warm, tender, gentle) and six items reflecting traditionally masculine traits (defends own beliefs, has a strong personality, has leadership abilities, makes decisions easily, dominant, acts as a leader). Participants indicate how well each trait describes them on a 7-point scale from 1 (*never*) to 7 (*always*). The masculinity and femininity scores (MASC

and FEM, respectively) were separately calculated as an average of masculine and feminine items, respectively. The higher the score, the greater the identification of an individual with the masculine or feminine gender-role. The McDonald's omega test showed a good reliability of both masculinity and femininity subscales (Masculinity –  $\omega = .86$ ;  $SE = .01$ ; 95% CI [.83, .88]; Femininity –  $\omega = .91$ ;  $SE = .01$ ; 95% CI [.88, .92]).

## Results

### Descriptive Statistics

For the purpose of future replications or meta-analytical studies, Table 2 reports descriptive statistics and correlation matrix for our study variables. We report the data for the whole study sample as well as for separate samples of women who did not experience a stereotype threat (control group) and women who did experience stereotype threat (women from experimental group one and two together). The correlational matrix showed that, in the groups of women, our third measure of risk aversion (REM3) did not significantly correlate with other two risk aversion measures.

### The Comparison of Women and Men in Risk Aversion without Activated Stereotype Threat

We hypothesized that without activating negative stereotypes, women and men would express similar risk aversion in decision-making (H1). Additionally, we were interested whether this similarity was present across different risk aversion measures (RQ1). Since our hypothesis had a null formulation, and our REM3 risk aversion measure did not correlate with the other two measures, we performed three Bayesian independent samples tests for each risk aver-

Table 2 Descriptive statistics

The whole study sample ( $N = 321$ )									
Variable	<i>M</i>	<i>SD</i>	ANX	CRT	MASC	FEM	REM1	REM2	REM3
AGE	44.85	16.26							
ANX	11.00	3.85	—						
CRT	2.63	2.10	-.14*	—					
MASC	27.67	6.25	-.17*	.02	—				
FEM	32.56	5.77	-.15*	-.09	.31***	—			
REM1	4.80	1.11	-.03	.08	.03	.16**	—		
REM2	29.19	5.60	-.04	-.10	.01	.09	.46***	—	
REM3	8.30	4.18	.11	-.03	-.05	-.07	.03	.19***	—
Women without stereotype threat activated ( $n = 59$ )									
Variable	<i>M</i>	<i>SD</i>	ANX	CRT	MASC	FEM	REM1	REM2	REM3
AGE	44.42	17.37							
ANX	11.90	4.31	—						
CRT	2.24	2.02	.04	—					
MASC	27.63	6.48	-.01	-.12	—				
FEM	33.53	5.13	-.03	-.12	.54***	—			
REM1	4.78	1.1	.13	.19	-.04	.14	—		
REM2	28.81	5.98	.03	.03	<-.01	.17	.53***	—	
REM3	7.17	4.05	.05	.04	.14	.02	<.01	.09	—
Women with stereotype threat activated ( $n = 108$ )									
Variable	<i>M</i>	<i>SD</i>	ANX	CRT	MASC	FEM	REM1	REM2	REM3
AGE	45.23	15.88							
ANX	11.28	3.91	—						
CRT	2.19	1.92	-.07	—					
MASC	27.03	6.61	-.13	<.01	—				
FEM	33.76	5.37	-.21*	.02	.27**	—			
REM1	4.94	1.07	.02	.06	.06	.25**	—		
REM2	30.13	5.10	-.18	-.04	.10	.12	.39***	—	
REM3	8.35	4.19	.15	-.10	-.11	.20	-.04	.10	—

Note. ANX – anxiety score; CRT – cognitive reflection test score; MASC – masculinity score; FEM – femininity score; REM1 – risk elicitation method 1 score; REM2 – risk elicitation method 2 score; REM3 – risk elicitation method 3 score.

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

sion measure separately. Table 3 shows that, for the first two risk aversion measures (REM1 and REM2), average scores showed to be very similar, with women expressing negligible higher risk aversion. However, this was not the case of the third risk aversion measure, where women expressed lower risk aversion.

Bayesian independent samples tests (see Figure 2) for first two risk aversion measures supported our hypothesis. Specifically, Bayes factors indicated moderate evidence for  $H_0$ ,  $BF_{01} = 3.39$ ; and  $BF_{01} = 4.87$  respectively, which means that the data were approximately 3.4 and 4.9 times more likely under  $H_0$  than un-

der H+ (see classification scheme proposed by Wagenmakers, Love, Marsman, Jamil, Ly, ... & Morey, 2018). However, the third risk aversion measure showed anecdotal evidence that there is a difference in risk aversion between men and women. The Bayes factor ( $BF_{10} = 1.46$ ) suggested that the data are approximately 1.5 times more likely under H+ than under H0. An interesting finding to re-

peat here is that women showed lower risk aversion than men.

In addition to Bayesian statistics, we performed equivalence testing using the TOSTER module in Jamovi proposed by Lakens (2017). As suggested by the author, groups were considered equivalent when both equivalence bounds were rejected. *P*-values for TOST Upper and Lower in Table 4 show that this was

Table 3 Descriptive statistics for three risk aversion measures of control group

Measure	Group	<i>n</i>	<i>M</i>	<i>SD</i>	95% CI [LL, UL]
REM1	Men	49	4.57	1.29	[4.20, 4.94]
	Women	59	4.78	1.10	[4.49, 5.07]
REM2	Men	49	28.71	5.47	[27.14, 30.29]
	Women	59	28.81	5.98	[27.26, 30.37]
REM3	Men	49	8.82	4.02	[7.66, 9.97]
	Women	59	7.17	4.05	[6.11, 8.23]

Note. REM1 – risk elicitation method 1 score; REM2 – risk elicitation method 2 score; REM3 – risk elicitation method 3 score; LL and UL indicate the lower and upper limits of a 95% credible interval, respectively.

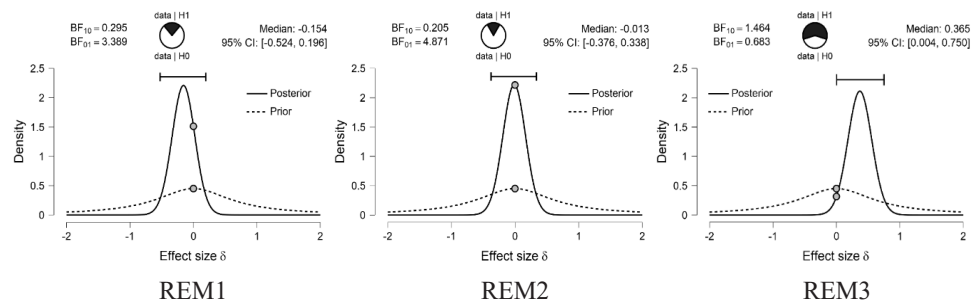


Figure 2 Comparisons of risk aversion of men and women without activated stereotype threat

Note. Figure shows three separate Bayesian two-sample t-tests for the examining the differences in three risk elicitation measures between men and women without activated stereotype threat. The probability wheel on top visualizes the Bayes factor evidence for supporting H0 and H+. The two gray dots indicate the prior and posterior density at the test value. Finally, the figure reports median and the 95% central credible interval of the posterior distribution for all three analyses.

REM1 – risk elicitation method 1 score, REM2 – risk elicitation method 2 score, REM3 – risk elicitation method 3 score.

satisfied for REM1 and REM2 risk aversion measures, but not for REM3 measure. In general, these results are in line with the Bayesian statistics, indicating that risk aversion of women and men in control group were equal for REM1 and REM2, but not for REM3.

### The Comparison of Women and Men in Risk Aversion with Activated Stereotype Threat

We hypothesized that women experiencing stereotype threat would be more risk averse compared to men (H2). As for the H1, we performed three Bayesian independent samples tests to test for the difference between women and men in experimental groups 1 and 2 for each risk aversion measure separately. Table 5 shows average means of three risk aversion measures for men and women in experimental groups 1 and 2.

For experimental group 1, Bayesian independent samples tests showed an anecdotal

evidence for supporting our hypothesis for REM1 measure ( $BF_{10} = 1.43$ ), suggesting that the data were approximately 1.5 times more likely under H+ than under H0. However, this was not the case for REM2 ( $BF_{01} = 1.45$ ) and REM3 ( $BF_{01} = 3.56$ ) measures. For these two risk aversion measures, the Bayes factors showed that the data were approximately 1.5 and 3.5 times more likely under H0 than under H+, providing anecdotal evidence of moderate support that women and men showed similar risk aversion (see Figure 3).

For experimental group 2, Bayesian independent samples tests showed even weaker support for our hypothesis. We found moderate evidence for supporting a null hypothesis using REM1 ( $BF_{01} = 4.53$ ) and REM3 ( $BF_{01} = 4.88$ ) measures, suggesting that the data were approximately 4.5 and 4.9 times more likely under H0 than under H+. For the REM2, the resulting Bayes factor indicated an anecdotal evidence for supporting H0 ( $BF_{01} =$

Table 4 Comparisons of risk aversion of women and men without activated stereotype threat - equivalence testing

Risk aversion measure		<i>t</i>	<i>df</i>	<i>p</i>
REM1	t-test	-.89	94.79	.37
	TOST Upper	-3.46	94.79	<.001
	TOST Lower	1.68	94.79	.05
REM2	t-test	-.09	104.95	.93
	TOST Upper	-2.69	104.95	<.01
	TOST Lower	2.51	104.95	<.01
REM3	t-test	2.11	102.64	.04
	TOST Upper	-.48	102.64	.32
	TOST Lower	4.70	102.64	<.001

Note. REM1 – risk elicitation method 1 score; REM2 – risk elicitation method 2 score; REM3 – risk elicitation method 3 score.

Table 5 Descriptive statistics for three risk aversion measures of experimental groups 1 and 2

Experimental group 1					
Measure	Group	<i>n</i>	<i>M</i>	<i>SD</i>	95% CI [LL, UL]
REM1	Men	57	4.72	1.18	[4.4, 5.03]
	Women	47	5.17	.99	[4.88, 5.46]
REM2	Men	57	29.00	6.23	[27.35, 30.65]
	Women	47	30.85	5.03	[29.38, 32.33]
REM3	Men	57	8.77	4.26	[7.64, 9.90]
	Women	47	8.11	3.96	[6.94, 9.27]
Experimental group 2					
REM1	Men	48	4.85	.95	[4.58, 5.13]
	Women	61	4.77	1.10	[4.49, 5.05]
REM2	Men	48	28.27	5.48	[26.68, 29.86]
	Women	61	29.57	5.13	[28.26, 30.89]
REM3	Men	48	8.48	4.26	[7.24, 9.72]
	Women	61	8.54	4.38	[7.42, 9.66]

Note. REM1 – risk elicitation method 1 score; REM2 – risk elicitation method 2 score; REM3 – risk elicitation method 3 score; LL and UL indicate the lower and upper limits of a 95% credible interval, respectively.

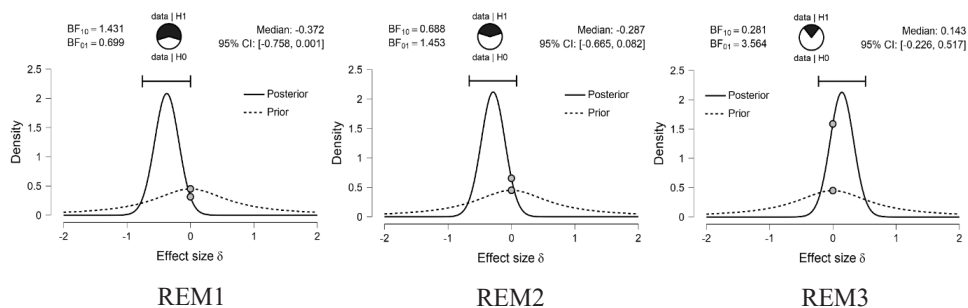


Figure 3 Comparisons of risk aversion of men and women with activated implicit stereotype threat

Note. Figure shows three separate Bayesian two-sample *t*-tests for the examining the differences in three risk elicitation measures between men and women with activated implicit stereotype threat. The probability wheel on top visualizes the Bayes factor evidence for supporting H0 and H+. The two gray dots indicate the prior and posterior density at the test value. Finally, the figure reports median and the 95% central credible interval of the posterior distribution for all three analyses.

REM1 – risk elicitation method 1 score, REM2 – risk elicitation method 2 score, REM3 – risk elicitation method 3 score.



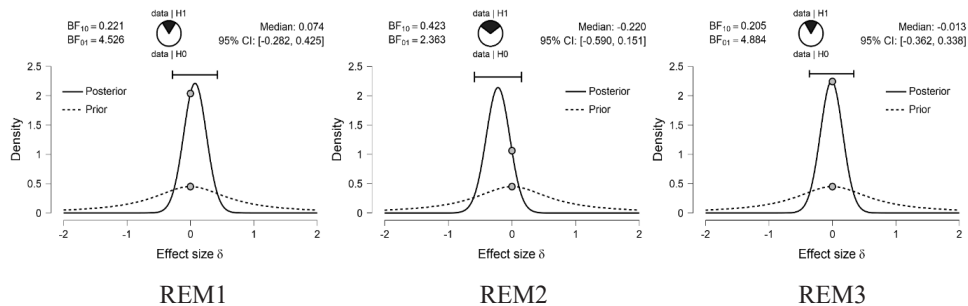


Figure 4 Comparisons of risk aversion of women and men with activated explicit stereotype threat

*Note.* Figure shows three separate Bayesian two-sample  $t$ -tests for the examining the differences in three risk elicitation measures between men and women with activated explicit stereotype threat. The probability wheel on top visualizes the Bayes factor evidence for supporting  $H_0$  and  $H_+$ . The two gray dots indicate the prior and posterior density at the test value. Finally, the figure reports median and the 95% central credible interval of the posterior distribution for all three analyses.

REM1 – risk elicitation method 1 score, REM2 – risk elicitation method 2 score, REM3 – risk elicitation method 3 score.

2.36); the data were approximately 2.4 times more likely under  $H_0$  than under  $H_+$ . Overall, all three analyses suggested that women and men in experimental group 2 did not differ in risk aversion (see Figure 4).

### The Effect of Stereotype Threat on Women's Risk Aversion

In order to replicate the findings of Carr and Steele (2010), we hypothesized that women experiencing stereotype threat are more risk averse compared to women not threatened by stereotypes ( $H_3$ ). In addition, we were interested whether the effect of stereotype threat on risk aversion varies across different types of risk aversion measures (RQ2) and whether the effect of stereotype threat on risk aversion depends on the stereotype threat condition (RQ3). In order to test all these three aims together, we aimed to perform a multi-

ple analysis of variance test (MANOVA) with three women groups (control, experimental 1, and experimental 2) and three dependent risk aversion measures. However, the assumptions testing for MANOVA showed a violation of multivariate normality for the control group (Henze-Zirkler = 1.08;  $p = .02$ ) and experimental group 1 (Henze-Zirkler = 1.46;  $p < .001$ ). We have checked for multivariate outliers for our three dependent variables using the Mahalanobis distance. The analysis showed that there were no multivariate outliers, which would significantly deviate from the research sample and should be removed from our hypotheses testing analyses. Finally, since multivariate normality test showed that our data were not normally distributed, we used Levene's test for equality of variances for each dependent variable instead of using the Box's M Test. This testing showed that the variances were equal for all three dependent

variables ( $F_{rem1} = 1.33; p = .27; F_{rem2} = 1.79; p = .17; F_{rem3} = .41; p = .66$ ).

Due to the non-normal distribution of our data, we have decided to perform a semi-parametric modified multivariate ANOVA-type test statistics (MATS) using the MANOVA\_RM package for R version 3.6.2 (R Core Team, 2020). This modified statistic, proposed by Friedrich and Pauly (2018), can be used for non-normal distributed data with heteroscedastic variances and different groups' sample sizes. In addition to multivariate testing, this statistic is also applicable for univariate comparisons with using Bonferroni correction for adjusting the  $p$ -value level according to multiple testing.

In this analysis, we compared three groups of women, specifically women not threatened by stereotype threat (control group), women implicitly threatened by stereotypes (experimental group 1), and women explicitly threatened by stereotypes (experimental group 2). The groups were compared in risk aversion measured by three risk aversion measures together (see average scores for these measures in Table 6). The results of the semi-parametric one-way MANOVA test with 10,000 parametric bootstrap runs showed a non-significant multivariate difference in risk aversion between three groups (MATS  $Q_n = 12.21; p = .06$ ), indicating that the stereotype threat did not significantly affect risk aversion.

Although the multivariate testing did not reject a null hypothesis, we have decided to perform univariate analyses with each risk aversion measure separately. This was done due to the non-significant correlation of our

REM3 measure with our other two risk aversion measures. Additionally, since Carr and Steele (2010) used only the REM3 method for measuring risk aversion, these separate analyses allowed us to better see whether the findings of Carr and Steele (2010) were replicated.

As in the previous analysis, we performed a semi-parametric modified multivariate ANOVA-type test for three separate univariate analyses. As suggested by Friedrich, Konietzschke, and Pauly (2019), we used a Bonferroni's correction to adjust the significance level to  $\alpha = .017$ . The results of this univariate testing are in line with the multivariate findings, showing that the three women groups did not significantly differ in any of the three risk aversion measures (REM1 - MATS  $Q_n = 5.08; p = .07; \eta^2 = .03$ ; REM2 - MATS  $Q_n = 3.78; p = .17; \eta^2 = .02$ ; REM3 - MATS  $Q_n = 3.35; p = .18; \eta^2 = .02$ ).

Finally, for the REM3 risk aversion measure, we performed pairwise post hoc comparison of women from the control group and experimental group 1 experiencing implicit stereotype threat. This was done in order to evaluate the replicability of the exact same comparison that was performed by Carr and Steele (2010). Compared averaged scores for this analysis can be found in Table 6. It was showed that these two groups did not significantly differ ( $t = -1.2; df = 104; p = .24; d = -.23; 95\% \text{ CI for Cohen's } d [-.62, .15]$ ) and our results were far from being similar to those in Study 2a brought by Carr and Steele (2010). Although the average score in risk aversion

Table 6 Average scores in three risk aversion measures in groups of women

Group	$n$	REM1	REM2	REM3
Control	59	4.78	28.81	7.17
Experimental group 1	47	5.17	30.85	8.11
Experimental group 2	61	4.77	29.57	8.54

Note. REM1 – risk elicitation method 1 score; REM2 – risk elicitation method 2 score; REM3 – risk elicitation method 3 score.

showed to be higher for women in the stereotype condition than for women from the control group (see Table 6), this difference was not significant. Moreover, the effect size was significantly lower, with confidence interval not containing the values found by Carr and Steele (2010).

#### Anxiety and Analytical Reasoning as Mediators of the Relationship between Stereotype Threat and Risk Aversion

We aimed to investigate the role of anxiety (RQ4) and analytical reasoning (RQ5) as mediators of the relationship between stereotype threat and risk aversion in a sample of women. For the investigation of indirect mediating effects, we used PROCESS macro for SPSS proposed by Preacher and Hayes (2004) with 5,000 bootstrap samples and 95% bias-corrected CIs. To test the mediating effect of anxiety and analytical reasoning jointly, we used multiple mediation parallel Model 4 (see Figure 5). Since the stereotype threat – our independent variable – was a multicategorical variable, we used the indicator coding approach (Hayes & Preacher, 2014). Two dummy variables were created to represent group

membership for two stereotype threat conditions (explicit and implicit), while control group was a reference. Due to this, instead of computing the indirect mediating effect using common strategy with single parameter  $a$  and  $b$ , this analysis used a set of two parameter estimates for path  $a$  (see path diagram in Figure 5). These two parameters corresponded to the mean differences in anxiety/analytical reasoning between implicit stereotype threat group relative to the control group ( $a_1$ ), and explicit stereotype threat group relative to the control group ( $a_2$ ). Consequently, the indirect mediating effect was quantified by multiplying  $a_1$  and  $a_2$  by  $b$ . If at least one of these two relative indirect effects was different from zero, according to a percentile-based bias-corrected bootstrap confidence intervals (CIs), we concluded that the mediating effect was present (see Hayes & Preacher, 2014).

Table 7 reports unstandardized regression weights and 95% CIs for the  $a_1b$  and  $a_2b$  indirect mediating effects of anxiety and analytical reasoning on the relationship between stereotype threat and three REM measures. The results showed that, for all three REMs, 95% CIs for all indirect mediating effects contained a zero value, suggesting that neither

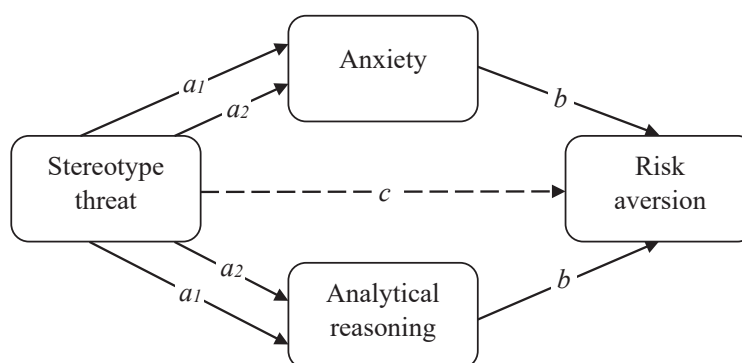


Figure 5 Investigated parallel multiple mediation model with two mediators.

Table 7 Anxiety and analytical reasoning as mediators of the relationship between stereotype threat and risk aversion

Criterion	Mediator	Path	<i>b</i>	<i>SE</i>	95% CI [LL, UL]
REM1	Anxiety	$a_1b$	-.02	.03	[-.12, .01]
		$a_2b$	-.01	.02	[-.09, .02]
	Analytical reasoning	$a_1b$	.01	.02	[-.02, .09]
		$a_2b$	-.01	.03	[-.10, .02]
REM2	Anxiety	$a_1b$	.10	.17	[-.10, .68]
		$a_2b$	.05	.14	[-.10, .51]
	Analytical reasoning	$a_1b$	-.01	.11	[-.37, .13]
		$a_2b$	.02	.11	[-.13, .40]
REM3	Anxiety	$a_1b$	-.10	.14	[-.56, .05]
		$a_2b$	-.05	.12	[-.43, .09]
	Analytical reasoning	$a_1b$	-.02	.07	[-.29, .06]
		$a_2b$	.02	.07	[-.06, .27]

Note.  $n = 167$ ; for all analyses the predictor variable was stereotype threat;  $b$  represents unstandardized regression coefficients; LL and UL indicate the lower and upper limits of a 95% confidence interval, respectively.

anxiety nor analytical reasoning significantly mediated the relationship between stereotype threat and risk aversion.

#### Masculinity and Femininity as Moderators of the Relationship between Stereotype Threat and Risk Aversion

Finally, we investigated the moderating effect of masculinity and femininity on the relationship between stereotype threat and risk aversion (RQ6). Six separate two-step hierarchical multiple regressions were conducted with masculinity and femininity as moderators and three REMs as dependent variables. Before conducting these analyses, two dummy variables were created to represent group membership for two stereotype threat conditions (explicit and implicit), while control group was a reference. Subsequently, two separate product terms were created to present the stereotype threat-by-masculinity (femininity) interaction. Since both moderators were continuous variables, we performed a mean centering for a more meaningful interpretation of

the effect of predictor on dependent variable.

In the hierarchical multiple regression analyses, two stereotype threat dummies and masculinity (femininity) were entered in the first step, and two interactions were added in the second step. In order to determine a moderating effect, the interaction terms had to show a statistically significant amount of variance explained for the risk aversion, with 95% confidence interval not containing a zero (Hayes & Rockwood, 2017).

The results showed that the relationship between stereotype threat and risk aversion was not significantly moderated by masculinity. Specifically, for all three REMs, the hierarchical multiple regression revealed that stereotype threat dummies and masculinity did not significantly contribute to the regression model at first step (see significance of  $F$ -tests in Table 8), accounting for only negligible variation in risk aversion. Adding the interactions in the second step did not result in significant changes in the models. For all three REMs, the effects of interactions were non-significant and introducing these interactions did not result in significant changes in  $R^2$ .

Table 8 Masculinity as moderator of the relationship between stereotype threat and risk aversion

Step	Criterion	Predictor	b	SE	$\beta$	p	95% CI [LL, UL]	R <sup>2</sup>	R <sup>2</sup> change	F	p
1	REM1	ST1	.39	.21	.16	.07	[-.02, .81]	.03	-	1.53	.21
		ST2	-.01	.20	-.00	.96	[-.40, .38]				
		MSC	.00	.08	.00	.99	[-.17, .17]				
2	REM1	ST1	.38	.21	.16	.08	[-.04, .80]	.03	.00	.93	.46
		ST2	-.00	.20	-.00	.97	[-.40, .39]				
		MSC	-.04	.14	-.04	.78	[-.32, .24]				
		INT1	.07	.22	.04	.73	[-.35, .50]				
		INT2	.06	.20	.03	.78	[-.34, .45]				
1	REM2	ST1	2.00	1.06	.17	.06	[-.10, 4.10]	.02	-	1.36	.26
		ST2	.83	1.00	.07	.41	[-1.14, 2.80]				
		MSC	.26	.43	.05	.54	[-.58, 1.10]				
2	REM2	ST1	1.95	1.08	.16	.07	[-.19, 4.08]	.03	.00	.85	.51
		ST2	.83	1.01	.07	.41	[-1.16, 2.82]				
		MSC	-.01	.72	-.00	.99	[-1.44, 1.42]				
		INT1	.48	1.09	.05	.66	[-1.67, 2.62]				
		INT2	.37	1.00	.04	.72	[-1.63, 2.37]				
1	REM3	ST1	.95	.82	.10	.25	[-.66, 2.57]	.02	-	1.15	.33
		ST2	1.35	.77	.16	.08	[-.16, 2.86]				
		MSC	-.09	.33	-.02	.78	[-.74, .55]				
2	REM3	ST1	.94	.82	.10	.26	[-.68, 2.55]	.05	.02	1.51	.19
		ST2	1.21	.76	.14	.12	[-.30, 2.72]				
		MSC	.56	.55	.14	.31	[-.52, 1.65]				
		INT1	-.38	.82	-.05	.64	[-2.01, 1.24]				
		INT2	-1.49	.77	-.22	.05	[-3.01, .02]				

Note. n = 167, ST1 – stereotype threat dummy variable for the mean differences in risk aversion between implicit stereotype threat group and control group; ST2 – stereotype threat dummy variable for the mean differences in risk aversion between explicit stereotype threat group relative to the control group; MSC – masculinity; b represents unstandardized regression weights, beta indicates the standardized regression weights; LL and UL indicate the lower and upper limits of a 95% confidence interval for unstandardized regression coefficients.

Similar to masculinity, we did not find a significant moderating effect of femininity on the relationship between stereotype threat and risk aversion (see Table 9). For REM1, we found that the regression model including stereotype threat dummies and femininity was significant. In this model the femininity alone significantly contributed to the explanation of variation in risk aversion. However, when we added interactions between stereotype threat dummies and femininity in the second step, there was hardly any change in  $R^2$  and the interaction effects were non-significant. For REM2 and REM3 measures, stereotype threat dummies and femininity did not significantly contribute to the regression model at first step (Table 9). Adding the interactions of stereotype threat dummies and femininity in the second step did not result in significant changes in  $R^2$ . Additionally, the effects of interactions were non-significant in both models.

### Discussion

The main aim of the study was to conceptually replicate research reported by Carr and Steele (2010, Study 2a) showing that risk attitudes are affected by the presence of adverse stereotypes about women's lower competences in mathematical, logical and rational reasoning. Beyond the attempt to replicate the original effects, we extended the study by Carr and Steele (2010). The paper investigated the impact of situational threats by introducing three distinct risk elicitation methods and two stereotype conditions. Additionally, we extended the limited knowledge about the explanatory variables of the effect of stereotype threat on risk aversion by investigating the mediating effects of anxiety and analytical reasoning. Finally, we aimed to provide insights into the issue of gender differences in risk aversion and the relation between per-

sonality traits – associated with masculinity and femininity – and risk preferences.

In general, our data support our H1 that there are no differences between women and men in risk aversion. Moreover, the risk aversion did not significantly differ between women threatened and not threatened by stereotypes, which lead us to reject our H2 and H3. The exploratory analyses did not seem to shed much light into explaining the differences of women threatened and not threatened by stereotypes. Specifically, we found no mediating effect of anxiety and analytical reasoning and no moderating effect of masculinity and femininity for the relationship between stereotype threat and risk aversion. Below, we discuss these findings in more detail.

### Discussion of the Main Findings

In line with recent literature about gender differences in risk aversion (Nelson, 2012a, 2018), when women were not threatened by stereotypes, the present study found no gender differences for REM1 and REM2 (hypothetical investment task and questionnaire, respectively). Contrary to the original replicated study by Carr and Steele (2010), there were minor gender differences in risk aversion measured by REM3 hypothetical lotteries. However, since our analyses showed only anecdotal evidence, these slight differences should be interpreted very carefully. The direction of the differences, though, was opposite to the one suggested by previous studies using similar measures (Csermely & Rabas, 2016; Filippin & Crosetto, 2016), i.e., in our study women showed lower risk aversion compared to men. This result, although unexpected, is not surprising. In fact, several studies with specific samples as well as meta-analyses suggested that it is not uncommon for women being less risk averse than men. For instance, Beckmann, and Menkoff (2008) compared risk attitudes of over 600 fund managers in Italy, Germany,

Table 9 Femininity as moderator of the relationship between stereotype threat and risk aversion

Step	Criterion	Predictor	b	SE	$\beta$	p	95% CI [LL, UL]	R <sup>2</sup>	R <sup>2</sup> change	F	p
1	REM1	ST1	.30	.21	.16	.06	[-.02, .79]	.07	-	4.21	.01
		ST2	-.02	.19	-.01	.91	[-.40, .36]				
		FEM	.23	.08	.21	.01	[.07, .39]				
2	REM1	ST1	.39	.21	.16	.06	[-.02, .79]	.08	.00	2.66	.02
		ST2	-.02	.19	-.01	.92	[-.40, .36]				
		FEM	.16	.14	.14	.28	[-.13, .44]				
		INT1	.19	.21	.09	.39	[-.24, .61]				
		INT2	.06	.19	.04	.76	[-.32, .44]				
1	REM2	ST1	2.02	1.05	.17	.06	[-.06, 4.09]	.04	-	2.36	.07
		ST2	.72	.98	.06	.47	[-1.22, 2.66]				
		FEM	.76	.42	.14	.07	[-.07, 1.58]				
2	REM2	ST1	2.01	1.06	.17	.06	[-.08, 4.10]	.04	.00	1.47	.20
		ST2	.71	.99	.06	.47	[-1.24, 2.66]				
		FEM	1.02	.73	.19	.17	[-.42, 2.46]				
		INT1	-.64	1.09	-.06	.56	[-2.80, 1.52]				
		INT2	-.22	.98	-.03	.83	[-2.16, 1.72]				
1	REM3	ST1	.94	.82	.10	.25	[-.67, 2.55]	.02	-	1.14	.33
		ST2	1.38	.76	.16	.07	[-.13, 2.88]				
		FEM	-.08	.32	-.02	.80	[-.72, .56]				
2	REM3	ST1	.93	.82	.10	.26	[-.68, 2.55]	.02	.00	.73	.60
		ST2	1.38	.77	.16	.07	[-.13, 2.89]				
		FEM	.10	.56	.02	.86	[-1.02, 1.21]				
		INT1	-.12	.85	-.02	.89	[-1.80, 1.55]				
		INT2	-.37	.76	-.06	.63	[-1.87, 1.13]				

Note. n = 167, ST1 – stereotype threat dummy variable for the mean differences in risk aversion between implicit stereotype threat group and control group; ST2 – stereotype threat dummy variable for the mean differences in risk aversion between explicit stereotype threat group relative to the control group; FEM – femininity; b represents unstandardized regression weights; beta indicates the standardized regression weights; LL and UL indicate the lower and upper limits of a 95% confidence interval for unstandardized regression coefficients.

US, and Thailand. Only in Italy did they find evidence of women's greater risk aversion. In Germany and US, women were less risk averse, although the effect was statistically insignificant. In line with these findings, Adams and Funk (2012) established that Swedish managers were not only less risk averse than women in the general Swedish population but also than male managers. Generally, according to Nelson (2012a) women showed greater willingness to take risk in 4 out of 24 studies she included in her meta-analysis. Therefore, we can conclude that our results speak in favour of no gender differences and, specifically, do not support the view of women's greater risk aversion. The results indicate that the claim about women's greater risk aversion can hardly be generalised outside the study context. Although an individual's risk aversion may be related to various forms of behaviour, including educational and occupational choices or choices of compensation schemes (Bertrand, 2018; Bonin et al., 2007; Dohmen et al., 2011; Francesconi & Parey, 2018), the belief that women are characterised by greater risk aversion than men found little justification in recent research and should be seen rather as a stereotype than as a robust fact.

Following the study by Carr and Steele (2010) we also hypothesized that women threatened by stereotypes would be more risk averse compared to men (H2) as well as unthreatened women (H3). Only in the group with activated implicit stereotype threat and only for the case of REM1 (the hypothetical investment task), was there an anecdotal evidence that women had higher risk aversion than men. Considering other comparisons showing no differences between men and women, our data are far from being supportive of our H2. Moreover, the comparison of three women groups showed no significant differences in risk aversion scores across conditions. These findings are in line with consid-

erable literature claiming that the stereotype theory provides mixed or even unreliable results. Most notably, in a large registered report, Flore, Mulder, and Wicherts (2018) failed to replicate the effects and to identify any moderators of the possible impact of stereotypes on mathematical ability, including gender and field identification as well as test difficulty. Similarly, Ganley et al. (2013) failed to replicate the effect in any of their three studies with any of the priming types they used (either explicit or implicit). Overall, Stoet, and Geary (2012) found that only 30 percent of replications corroborated the effect of stereotype threat on women's mathematical performance, concluding that the enthusiasm for the theory is, likely, exaggerated. Although previous meta-analyses supported the view that stereotype threat could affect performance of threatened groups, the overall effects were small and differed depending on contextual factors, such as the priming type (Flore & Wicherts, 2015; Picho, Rodriguez, & Finnie, 2012). Specifically, Flore and Wicherts (2015) identified a large heterogeneity of effects in primary studies ranging from small to medium, concluding that publication bias in the field of stereotype threat theory may be responsible for the overrepresentation of studies confirming impact of stereotype threat on the threatened groups' performance. Overall, despite our broad design including two stereotype threat conditions and three measures of risk aversion, and the use of power analysis to establish the sample size, the data do not support the claim that women threatened by negative stereotypes about their cognitive abilities would express distinct risk preferences than men and women in neutral conditions. Consequently, we conclude that our data do not support hypotheses 2 and 3 and thus fail to replicate the original results reported by Carr and Steele (2010).



As Spencer, Steele, and Quinn (1999) speculated, one reason why stereotype threat manipulations could be ineffective is that individuals perceived the tasks as relatively easy (cognitively undemanding) and thus they would not feel the threat of disconfirming their abilities. In line with this explanation, it is possible that risk elicitation measures are not seen as diagnostic and thus the stereotype threat effect is mitigated or even eliminated. Although it is possible that the REMs we used were perceived as non-diagnostic, this cannot explain why the method used in the original study (Carr & Steele, 2010) now failed to inhibit the individuals' performance. Another reason, however, could be the sample composition. Unlike Carr and Steele (2010), we used a general population sample. Spencer, Steele, and Quinn (1999) observed that one of the conditions necessary to capture the effect of stereotype threat on performance is the identification with the domain. If an individual has already disidentified with mathematics, the effect is unlikely to occur. It is possible, thus, that outside educational or academic context, mathematical skills are less relevant for individuals' self-concept and they may feel less threatened by the stereotypes. Consequently, the manipulation had no effect on their state anxiety and cognitive capacity, posing no threat to behaviour as their performance expectations were already low or irrelevant for self-image. Lastly, some authors warned that studies on stereotype threat were, so far, performed in a limited set of countries and nearly two-thirds of the published results come from the US (Flore, Mulder, & Wicherts, 2018; Stoet & Geary, 2012). Thus, it is possible that the effects are less pronounced or different types of threats come to the foreground in other cultural settings. Stoet and Geary (2012) speculated that the effects could be stronger in regions where egalitarian views of gender norms re-

ceive lower support. However, our results do not support this claim. As the Eurobarometer shows (see Cukrowska-Torzewska & Lovasz, 2020), Slovakia belongs to the most conservative countries in terms of stereotypical gender roles division, yet our results do not support the hypothesis that stereotype threat has any effect on women's risk aversion.

### Discussion of Registered Exploratory Analyses

Previous research showed evidence that risk preferences elicited with different measures may significantly differ and that the measures provide results that are not robust across the contexts (Crosetto & Filippin, 2016; Csermely & Rabas, 2016; Pedroni et al., 2017). In line with these findings, we explored the issue of risk measures consistency in our study. We found that the three risk elicitation methods did not correlate or correlated weakly. Interestingly, we found correlations between REM1 and REM2 (hypothetical investment task and questionnaire) while REM3 – i.e., the very same method used in the original study – correlated weakly only with the questionnaire (REM2) but not in the women samples. Although this area requires further investigation, we can conclude that it is likely that risk preferences are sensitive to the methods used and that responses to specific REMs may differ by gender (Filippin & Crosetto, 2016; Pedroni et al., 2017). Consequently, risk preferences elicited with various methods cannot be easily compared and generalised. Particularly, using paired lotteries may pose greater difficulty, resulting in noisy and inconsistent results obtained (Csermely & Rabas, 2016). In fact, the weak relationships between three risk aversion measures used in our study may be caused by the methodological differences between the measures. Nevertheless, these findings contribute to the debate about rep-

lication of risk aversion studies. Crandall and Sherman (2016) claimed that, in general, conceptual replications have a considerable advantage over simpler protocols. However, in studies on risk aversion, one should be cautious when choosing risk elicitation methods different than those used in the original study, as the alleged failure or success may be related rather to the method chosen and not the validity of the investigated theory itself. Therefore, it would be recommended to use either the same method or, as indicated by Pedroni et al. (2017), to employ more than one method in order to obtain a reliable and consistent measure of risk aversion.

Driven by results by Seibt and Forster (2004) we explored the idea that different stereotype threat conditions may have distinct effects on threatened individuals' behaviour. Specifically, the authors (Seibt & Forster, 2004) indicated that, contrary to tacit activation, making the stereotypes salient may incline individuals to disconfirm the stereotypes. Our results do not support this claim. Not only did the stereotypes not induce women to be more risk averse but there were also no major differences between the explicit and implicit conditions, i.e., the difference in reaction to both of these stereotype threats was negligible. Unlike previous studies (Seibt & Forster, 2004; Spencer, Steele, & Quinn, 1999), we were also unable to corroborate the mediating effect of state anxiety and analytical reasoning on behaviour under stereotype threat. Overall, the results do not support a significant effect of stereotypes on women's risk preferences, either direct or mediated by anxiety.

Following Seibt and Forster (2004), we also considered analytical reasoning as a mediator of the effect of stereotype threat on risk preferences. The authors claimed that activation of negative stereotypes should make threatened individuals more cautious and accurate in analytical tasks. However, this was not true

for our study. Activation of negative stereotypes showed to have no effect on the threatened individuals' ability to reason analytically and, generally, analytical reasoning was unrelated to risk preferences as measured by any of the three instruments. Although literature discusses two competing explanations of the effect related to the individuals' cognitive ability – activation of prevention focus (Seibt & Forster, 2004) and ego depletion (Carr & Steele, 2010) – we found no support for any of the explanations. In neither of the stereotype threat conditions, did the manipulation affect cognitive skills – either positively or negatively. The findings are in line with recent studies indicating that the ego depletion theory itself is under crisis and the findings are now considered inconclusive and unconvincing (Carter & McCullough, 2014; Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2019).

Finally, based on research on the relation between masculinity and femininity on risk preferences we explored the impact of gender self-concept on risk aversion. Previous research provided support for the hypothesis that masculinity is related to lower risk aversion, i.e., more masculine individuals are willing to take more risks, regardless of their biological sex (Demaree et al., 2009; Lemaster & Strough, 2014; Meier-Pesti & Penz, 2008). However, we did not corroborate this effect. Although our results show that biological sex is not predictive of risk aversion, unlike previous studies, we found that femininity alone weakly correlated with one of the risk aversion measures, while masculinity was unrelated to any of the measures. Interestingly, we identified the effect only for REM1 (hypothetical investment task) – not for REM2 or REM3. The correlation was positive, indicating that the greater an individual's femininity, the greater risk aversion a participant expressed. Our findings suggest that the relation between gender self-concept and risk

preferences requires further investigation as the findings, so far, are mixed. Despite posing interesting research questions, the studies on gender self-concept and risk preferences are still scarce and should be, thus, interpreted with caution. Although personality traits are associated with risk preferences (Demaree et al., 2009; Lemaster & Strough, 2014), it is possible that traditional associations of instrumental and communal traits investigated in the 1970s are not valid any more (Kachel, Steffens, & Niedlich, 2016). Indeed, research indicates that greater social desirability of agentic traits increases women's identification with masculinity (Twenge, 1999). Consequently, it is likely that agentic traits will cease to be predictive of gender-typed forms of behaviour such as risk taking. Previous studies showed that masculinity and femininity measured with the Bem Sex Role Inventory are two relatively unrelated factors. In our study we found a moderate positive correlation between these two dimensions (Bem, 1974). The finding supports the view that gender self-concept should not be viewed as a unidimensional, bipolar scale. Rather, we should see gender self-concept as having at least two separate dimensions.

### Conclusions

In general, our study failed to replicate findings concerning the effect of stereotype threat on risk aversion reported by Carr and Steele (2010). However, it is important to remember that a failure to replicate previous findings does not necessarily mean the original results were unreliable. Even done with most care, replications can fail for various reasons including systematic and random errors, unintentional differences between samples and conditions, false positive and false negative effects (Freese & Peterson, 2017). Replication studies, instead of being judge-

mental, should help us identify good practices, control research quality, increase transparency with the ultimate goal of producing knowledge and raising public trust in science and scientists (Nature, 2014). We need to keep in mind that replications are never meant to be a witch-hunt but rather a quest for better, robust, more reliable and possibly more context-conscious science. If, after all, some axioms fall down when brought into the limelight, it is only for the sake of scientific progress itself. Consequently, despite the replication failure, our study provides an important voice in the debate about replicability of previous research and the usefulness of registered reports to avoid distorting knowledge on gender differences in risk aversion. Nelson (2012a, 2018) claimed that the overrepresentation of results confirming women's greater risk aversion – compared to men – may be related to two biases: confirmation and publication biases. The former indicates that researchers may believe that there are differences and thus assess results that disconfirm the belief as faulty. The latter is related to a systematic preference of editors to publish studies that provide significant results in an expected direction. Given the practical significance of risk aversion, the present study indicates that registered protocols may contribute to showing an accurate picture of women's risk preferences. Consequently, we call for more replications and more registered reports as sources of unbiased and reliable knowledge, particularly in areas subjected to heated debates such as gender differences. Beyond the impact on the condition of science itself, changing the optics and acknowledging that differences between men and women in risk-related decision are rarer and less pronounced than older studies suggest, will likely mitigate the stereotypical perception of women, particularly in the labour market.

### Acknowledgment

The research was supported by the Operational Programme Research, Development and Education MSCA-IF grant Gender and risk-taking: The role of social and situational antecedents on performance in risk-related tasks (MSCAfellow3@MUNI; CZ.02.2.69/0.0/0.0/19\_074/0012727; Magdalena Adamus) and the VEGA grant: Cognitive and personality predictors of trust building (VEGA no. 2/0035/20; Matúš Grežo).

### Authors' ORCID

Magdalena Adamus

<https://orcid.org/0000-0001-5317-0415>

Matúš Grežo

<https://orcid.org/0000-0002-3537-2862>

Katarína Dudeková

<https://orcid.org/0000-0002-0310-3844>

### References

- Adams, R. B., & Funk, P. (2012). Beyond the glass ceiling: Does gender matter? *Management Science*, 58(2), 219–235. <https://doi.org/10.1287/mnsc.1110.1452>
- Agnew, J. R., Anderson, L. R., Gerlach, J. R., & Szykman, L. R. (2008). Who chooses annuities? An experimental investigation of the role of gender, framing, and defaults. *American Economic Review*, 98(2), 418–422. <https://doi.org/10.1257/aer.98.2.418>
- Apicella, C. L., Dreber, A., Campbell, B., Gray, P. B., Hoffman, M., & Little, A. C. (2008). Testosterone and financial risk preferences. *Evolution and Human Behavior*, 29(6), 384–390. <https://doi.org/10.1016/j.evolhumbehav.2008.07.001>
- Arrow, K. J. (1951). Alternative approaches to the theory of choice in risk-taking situations. *Econometrica*, 19(4), 404–437.
- Beckmann, D., & Menkhoff, L. (2008). Will women be women? Analyzing the gender difference among financial experts. *Kyklos*, 61(3), 364–384.
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42(2), 155–162. <https://doi.org/10.1037/h0036215>
- Bertrand, M. (2018). Coase lecture – the glass ceiling. *Economica*, 85, 205–231. <https://doi.org/10.1111/ecca.12264>
- Bonin, H., Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics*, 14(6), 926–937. <https://doi.org/10.1016/j.labeco.2007.06.007>
- Booth, A. L., & Nolen, P. (2009). Gender differences in risk behaviour: Does nurture matter? *IZA Discussion Paper Series*, 122(558). <https://doi.org/10.1111/j.1468-0297.2011.02480.x>
- Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 125(3), 367–383. <https://doi.org/10.1037/0033-2909.125.3.367>
- Carothers, B. J., & Reis, H. T. (2013). Men and women are from Earth: Examining the latent structure of gender. *Journal of Personality and Social Psychology*, 104(2), 385–407. <https://doi.org/10.1037/a0030437>
- Carr, P. B., & Steele, C. M. (2010). Stereotype threat affects financial decision making. *Psychological Science*, 21(10), 1411–1416. <https://doi.org/10.1177/0956797610384146>
- Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology*, 5(July), 1–11. <https://doi.org/10.3389/fpsyg.2014.00823>
- Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior and Organization*, 83(1), 50–58. <https://doi.org/10.1016/j.jebo.2011.06.007>
- Coates, J. M., Gurnell, M., & Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proceedings of the National Academy of Sciences of the United States of America*, 106(2), 623–628. <https://doi.org/10.1073/pnas.0810907106>
- Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, 66, 93–99. <https://doi.org/10.1016/j.jesp.2015.10.002>
- Crosetto, P., & Filippin, A. (2016). A theoretical and experimental appraisal of four risk elicita-

- tion methods. *Experimental Economics*, 19(3), 613–641. <https://doi.org/10.1007/s10683-015-9457-9>
- Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 1–27. <https://doi.org/10.1257/jel.47.2.448>
- Csermely, T., & Rabas, A. (2016). How to reveal people's preferences: Comparing time consistency and predictive power of multiple price list risk elicitation methods. *Journal of Risk and Uncertainty*, 53(2–3), 107–136. <https://doi.org/10.1007/s11166-016-9247-6>
- Cukrowska-Torzewska, E., & Lovasz, A. (2020). The role of parenthood in shaping the gender wage gap – A comparative analysis of 26 European countries. *Social Science Research*, 85. <https://doi.org/10.1016/j.ssresearch.2019.102355>
- Demaree, H. A., DeDonno, M. A., Burns, K. J., Feldman, P., & Everhart, D. E. (2009). Trait dominance predicts risk-taking. *Personality and Individual Differences*, 47(5), 419–422. <https://doi.org/10.1016/j.paid.2009.04.013>
- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes. *Review of Economic Studies*, 79(2), 645–677. <https://doi.org/10.1093/restud/rdr027>
- 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>
- Dreber, A., & Hoffman, M. (2010). Biological basis of sex differences in risk aversion and competitiveness. *Working Paper*, 1–34.
- Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4), 281–295. [https://doi.org/10.1016/S1090-5138\(02\)00097-1](https://doi.org/10.1016/S1090-5138(02)00097-1)
- Eckel, C. C., & Grossman, P. J. (2008). Chapter 113 Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results*, 1(C), 1061–1073. [https://doi.org/10.1016/S1574-0722\(07\)00113-8](https://doi.org/10.1016/S1574-0722(07)00113-8)
- Filippin, A., & Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Management Science*, 62(11), 3138–3160. <https://doi.org/10.1287/mnsc.2015.2294>
- Finucane, M. L., Slovic, P., Mertz, C. K., Flynn, J., & Satterfield, T. A. (2000). Gender, race, and perceived risk: The “white male” effect. *Health, Risk and Society*, 2(2), 170–172.
- Flore, P. C., Mulder, J., & Wicherts, J. M. (2018). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: A registered report. *Comprehensive Results in Social Psychology*, 3(2), 140–174. <https://doi.org/10.1080/23743603.2018.1559647>
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, 53, 25–44.
- Francesconi, M., & Parey, M. (2018). Early gender gaps among University graduates. *European Economic Review*, 109, 63–82. <https://doi.org/10.1016/j.euroecorev.2018.02.004>
- Freese, J., & Peterson, D. (2017). Replication in social science. *Annual Review of Sociology*, 43, 147–165.
- Friedrich, S., Konietzschke, F., & Pauly, M. (2019). Resampling-based analysis of multivariate data and repeated measures designs with the R Package MANOVA.RM. *The R Journal*, 11(2), 380–400.
- Friedrich, S., & Pauly, M. (2018). MATS: Inference for potentially singular and heteroscedastic MANOVA. *Journal of Multivariate Analysis*, 165, 166–179.
- Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2019). Is ego depletion real? An analysis of arguments. *Personality and Social Psychology Review*, 23(2), 107–131. <https://doi.org/10.1177/1088868318762183>
- Gächter, S., Johnson, E. J., & Herrmann, A. (2007). Individual-level loss aversion in riskless and risky choices. *IZA Discussion Paper Series Discussion Paper*, (2961).
- Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology*, 49(10), 1886–1897. <https://doi.org/10.1037/a0031412>
- Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5), 1637–1664. <https://doi.org/10.3982/ecta6690>

- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67, 451–470. <https://doi.org/10.1111/bmsp.12028>
- Hayes, A. F., & Rockwood, N. J. (2017). Regression-based statistical mediation, moderation, and conditional process analysis: Observations, recommendations, and implementation. *Behaviour Research and Therapy*, 98, 39–57.
- Henrich, J., & McElreath, R. (2002). Are peasants risk-averse decision makers? *Current Anthropology*, 43(1), 172–181. <https://doi.org/10.1086/338291>
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581–592. <https://doi.org/10.1037/0003-066X.60.6.581>
- Kachel, S., Steffens, M. C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Frontiers in Psychology*, 7(JUL), 1–19. <https://doi.org/10.3389/fpsyg.2016.00956>
- Kahan, D. M., Braman, D., Gastil, J., Slovic, P., & Mertz, C. K. (2013). Culture and identity-protective cognition: Explaining the white-male effect in risk perception. *The Feeling of Risk: New Perspectives on Risk Perception*, 4(3), 163–182. <https://doi.org/10.4324/9781849776677>
- Kastlunger, B., Dressler, S. G., Kirchler, E., Mitton, L., & Voracek, M. (2010). Sex differences in tax compliance: Differentiating between demographic sex, gender-role orientation, and prenatal masculinization (2D: 4D). *Journal of Economic Psychology*, 31, 542–552. <https://doi.org/10.1016/j.joep.2010.03.015>
- Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology*, 80(6), 942–958. <https://doi.org/10.1037/0022-3514.80.6.942>
- Kristof, N. (2009). Mistresses of the universe. *New York Times Magazine*.
- Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. *Social Psychological & Personality Science*, 8, 355–362.
- Lemaster, P., & Strough, J. N. (2014). Beyond Mars and Venus: Understanding gender differences in financial risk tolerance. *Journal of Economic Psychology*, 42, 148–160. <https://doi.org/10.1016/j.joep.2013.11.001>
- Maney, D. L. (2016). Perils and pitfalls of reporting sex differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1688). <https://doi.org/10.1098/rstb.2015.0119>
- Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, 31, 301–306.
- Meier-Pesti, K., & Penz, E. (2008). Sex or gender? Expanding the sex-based view by introducing masculinity and femininity as predictors of financial risk taking. *Journal of Economic Psychology*, 29(2), 180–196. <https://doi.org/10.1016/j.joep.2007.05.002>
- Nature. (2014). Journals unite for reproducibility. *Nature*, 515(7).
- Nelson, J. A. (2012a). Are women really more risk-averse than men? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2158950>
- Nelson, J. A. (2012b). Would women leaders have prevented the global financial crisis? Implications for teaching about gender and economics. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1961722>
- Nelson, J. A. (2018). *Gender and risk-taking: Economics, evidence, and why the answer matters*. London: Routledge.
- Nigel, M. (2009). Harriet Harman: "If only it had been Lehman Sisters." *The Independent*. London.
- Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour*, 1(11), 803–809. <https://doi.org/10.1038/s41562-017-0219-x>
- Picho, K., Rodriguez, A., & Finnie, L. (2012). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of Social Psychology*, 153(3), 37–41. <https://doi.org/10.1080/00224545.2012.737380>
- Porcelli, A. J., & Delgado, M. R. (2009). Acute stress modulates risk taking in financial decision making. *Psychological Science*, 20(3), 278–283.
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers*, 36, 717–731.
- Reis, H. T., & Carothers, B. J. (2014). Black and white or shades of gray: Are gender differences

- categorical or dimensional? *Current Directions in Psychological Science*, 23(1), 19–26. <https://doi.org/10.1177/0963721413504105>
- Sapienza, P., Zingales, L., & Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences of the United States of America*, 106(36), 15268–15273. <https://doi.org/10.1073/pnas.0907352106>
- Schipper, B. C. (2012). Sex hormones and choice under risk. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2046324>
- Seibt, B., & Forster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology*, 87(1), 38–56. <https://doi.org/10.1037/0022-3514.87.1.38>
- Sirota, M., Dewberry, C., Juanchich, M., Kostovičová, L., & Marshall, A. (2018). *Measuring cognitive reflection without maths: Development and validation of the verbal cognitive reflection test*. Retrieved from <https://psyarxiv.com/pfe79/>
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28. <https://doi.org/10.1006/jesp.1998.1373>
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811. <https://doi.org/10.1037/0022-3514.69.5.797>
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, 34, 379–440.
- Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology*, 16(1), 93–102. <https://doi.org/10.1037/a0026617>
- Twenge, J. M. (1999). Mapping gender. The multifactorial approach and the organization of gender-related attributes. *Psychology of Women Quarterly*, 23, 485–502.
- Unger, R. K. (1979). Toward a redefinition of sex and gender. *American Psychologist*, 34(11), 1085–1094. <https://doi.org/10.1037/0003-066X.34.11.1085>
- Waerneryd, K.-E. (1996). Risk attitudes and risky behavior. *Journal of Economic Psychology*, 17, 749–770.
- Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., ... & Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25, 58–76. <https://doi.org/10.3758/s13423-017-1323-7>
- Wilson, M., & Daly, M. (1985). Competitiveness, risk taking, and violence: The young male syndrome. *Ethology and Sociobiology*, 6(1), 59–73. [https://doi.org/10.1016/0162-3095\(85\)90041-X](https://doi.org/10.1016/0162-3095(85)90041-X)