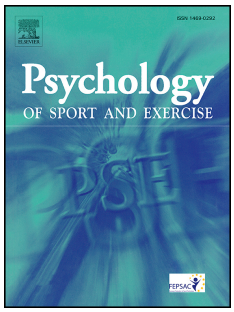
Accepted Manuscript



'You play like a Woman!' Effects of gender stereotype threat on Women's performance in physical and sport activities: A meta-analysis

Ambra Gentile, Stefano Boca, Isabella Giammusso

PII: S1469-0292(17)30508-3

DOI: 10.1016/j.psychsport.2018.07.013

Reference: PSYSPO 1399

To appear in: *Psychology of Sport & Exercise*

Received Date: 26 July 2017

Revised Date: 25 June 2018

Accepted Date: 27 July 2018

Please cite this article as: Gentile, A., Boca, S., Giammusso, I., 'You play like a Woman!' Effects of gender stereotype threat on Women's performance in physical and sport activities: A meta-analysis, *Psychology of Sport & Exercise* (2018), doi: 10.1016/j.psychsport.2018.07.013.

‘You Play Like a Woman!’

Effects of Gender Stereotype Threat on Women’s Performance in Physical and Sport Activities: A Meta-Analysis

Ambra Gentile, Stefano Boca, Isabella Giammusso

University of Palermo

Objectives: The purpose of this quantitative review was to provide an estimation of the effect of stereotype threat on women's performance in sport.
Design: This review employed a meta-analytic technique.
Method: a meta-analysis with random effects model was performed on 24 effects. Publication bias was tested through funnel plots and Egger's regression test.
Results: Findings show a symmetric distribution of effects, making it possible to conclude that no file-drawer problem affected the collected sample of effects. Aggregating the results of the reviewed studies, a medium effect of stereotype threat manipulation on women's sport performances emerged (d = 0.33). Collected studies were coded for stereotypicality of threatened exercise. The effect of stereotype threat was significantly higher for sports activities perceived as masculine.
Conclusions: This meta-analysis reveals that gender stereotype affects the sport activities of women and that this is particularly true for sports typically considered suited to males.

‘You Play Like a Woman!’

Effects of Gender Stereotype Threat on Women’s Performance in Physical and Sport Activities: A Meta-Analysis

Introduction (3)

Apart from research regarding people holding stereotypes, there is also a body of research concerning people who belong to stereotyped groups (Abbate, Boca, & Bocchiaro, 2004; Crocker & Major, 1989; Pinel, 1999; Wout, Shih, Jackson, & Sellers, 2009). According to Steele and Aronson (1995), stereotype threat is a self-evaluative threat that appears when an individual is at risk of confirming a negative stereotype about him- or herself. Since their seminal work appeared, many studies about this effect have been published, and many domains have been investigated, such as aging and memory tasks(Chasteen, Bhattacharyya, Horhota, Tam, & Hasher, 2005; Hess, Auman, Colcombe, & Rahhal, 2003), race and intelligence tests (Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002) and gender and the math domain (O'Brien & Crandall, 2003; Spencer, Steele, & Quinn, 1999). A growing body of research focuses on the physical activity domain, where negative stereotypes about women persist and could reduce their performance (Chalabaev, Sarrazin, Stone, & Cury, 2008; Stone & McWhinnie, 2008). The present meta-analytic review examined stereotype threat on women’s performance in sport or motor activities and hypothesized a number of moderators that might explain why this effect is stronger or weaker in some situations than in others. This particular domain also provides an occasion to question whether one of the main explanations of stereotype threat, the so-called ‘cognitive load hypothesis’, may also apply in a non-academic domain. Finally, the paper discusses if the present meta-analytic review is comparable to other meta-analyses conducted on stereotype threat.

Overview of Stereotype Threat

Stereotype threat theory covers a range of two decades of research. Steele and Aronson (1995) found that Black participants underperformed compared to White participants if a difficult verbal test was presented as a measure of their intelligence, but this effect did not occur if the same test was presented as non-diagnostic. Two years later Steele (1997) hypothesized this effect to be a situational threat, able to affect members of any group about whom a negative stereotype exists. Thus, stereotype threat may be defined as the ofthe fear of stigmatized individuals to be judged or treated stereotypically (Steele, et al., 2002). This effect usually consists of a suboptimal performance in a task related to a judgment dimension in which that particular group is 'known to be weak'. This commonly happens when a salient element of a stereotype relates to something that people are able or not able to do, and it also comes into play whenever people become aware of a negative stereotype about themselves. According to Schmader, Johns, and Forbes (2008), stereotype threat stems from a situation in which three core elements are activated in memory: the concept of one's ingroup, the concept of the ability domain in question, and the self concept. When a propositional relation (Gawronski & Bodenhausen, 2006) links these concept in a such a way that an individual becomes aware of an imbalance among them (Heider, 2013) he/she struggles to resolve the incongruence. The activation of a negative stereotype could harm a person's self-worth. When this happens, the individual may show some acute reactions, for example, domain avoidance, self-handicapping, counterstereotypic behaviour and disengagement (Davies, Spencer, Quinn, & Gerhardstein, 2002; Pyszczynski & Greenberg, 1983; Steele & Aronson, 1995).

According to Schmader, et al. (2008) can affect three types of performance outcomes: cognitive, social and sensorimotor threat. The authorsclaimthat cognitive overload is the principal reason why people under stereotype threat perform worse than those not affected by this threat in cognitive tasks. To test if stereotype threat was systematically associated with working memory load, Schmader and Johns (2003) used the so-called 'operation span task' developed by Turner and Engle (1989). In the experimental procedure, participants were required to solve math operations and keep words in their memory at the same time, and the number of words correctly recalled was assumed to be a measure of working memory load. A sample of women and men had to complete this task under threat and no-threat conditions. The results of the experiments confirmed the hypothesis that stereotype threat undermines an individual's performance on the cognitive test since it reduces the individual's working memory capacity. In particular, Schmader, et al. (2008)maintained that stereotype threat would impair the ability to regulate the attention during complex tasks, where it is necessary to inhibit thoughts, behaviours and feelings that are not related or are negatively related to the current task.

The second kind of threat refers to social situations in which stigmatized people show non-verbal anxiety (Bosson, Haymovitz, & Pinel, 2004; Richeson & Shelton, 2003). Drawing on the stereotype that gay and bisexual people have a negative influence on children's behaviour, Bosson, et al. (2004) asked a group of heterosexual, gay and bisexual people to complete a sociodemographic questionnaire. Half of them were submitted to the threat condition in which sexual orientation was primed through a question related to their sexual orientation. In the no-threat condition, the questionnaire did not contain any reference to sexual orientation. Participants were videotaped while interacting with children for 5 minutes. The results confirmed that gay/bisexual participants in the threat condition showed

more non-verbal anxiety than gay/bisexual people in the no-threat condition as well as more such anxiety than heterosexual participants in both conditions.

The third kind of threat concerns physical activities whose performances are harmed when explicit attention is directed to automatic processes that are not usually controlled by working memory (e.g., Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Gray, 2004). Beilock, et al. (2006) conducted a series of experiments asking a group of men who were expert golfers to complete the 'golf putting task' (Beilock & Carr, 2001). Participants in the threat condition were told that women tended to perform that task better than men, while those in the control condition were told that the task investigated individual differences in golf performance. As expected, participants in the threat condition performed significantly worse than those in the control group. Furthermore, the Authors found that giving to them a secondary task eliminates the effect of stereotype threat, since distracting the attention from thoughts regarding the stereotype. According to the authors, this suboptimal performance occurred because mentioning a negative stereotype led participants to focus attention on processes that commonly are automatic in expert golfers.

The explicit monitoring hypothesis (Beilock & Carr, 2001; Schmader, et al., 2008) claims that stereotype threat could harm proceduralized skills when people pay attention to processes that generally work better without conscious control. Thus people under stereotype threat might underperform because their attention is focused on actions that are usually automatic. If this theory is proven correct, the effect of stereotype threat should be greater for well learned gestures compared to new ones.

Gender Differences and Stereotypes in Physical and Sport Activity

In many fields of daily life (e.g., school, sports club), the differences between males and females are commonly considered more 'real' than they effectively are. In agreement with gender stereotypes that describe women as 'the weaker sex' physically, mentally and emotionally, the sport domain has been considered for centuries an exclusively masculine affair, not suited to the delicacy of a woman. It was only in the late 1800s that women were given the opportunity to participate in sports, confined to situations in which the activity did not allow physical contact or strain (Sherrow, 1996; Woolum, 1998). Although the number of women taking part in traditionally male sports (even the extreme ones) has been growing (Zimmerman, 2011), gender stereotypes persist in physical activities.

Nowadays, sports are perceived as masculine, feminine or neutral, and this perception depends on the degree of masculinity and femininity of their characteristics (Chalabaev, Sarrazin, Fontayne, Boiché, & Clément-Guillotin, 2013). Masculine sports possess particular characteristics such as physical contact, face-to-face opposition, strength, or aggressiveness, while feminine sports are typically characterized by expressivity, grace or esthetics (Chalabaev, Sarrazin, et al., 2013; Hardin & Greer, 2009). These stereotypes affect women's performance and contribute to keeping them out of the sport domain. In contrast, males continue to perceive themselves as characterized by higher strength and ability and give more importance to sports than females (Eccles & Harold, 1991; Fredricks & Eccles, 2004).

Stereotype Threat on Women in the Sport Domain

The effects of gender stereotype hold their roots in childhood, when parents' stereotypic beliefs about boys' and girls' abilities seem to play a significant role in children's self-perception of their abilities in various domains (Duriez & Soenens, 2009; O'Bryan, Fishbein,

& Ritchey, 2004; Parsons, Adler, & Kaczala, 1982; Tomasetto, Alparone, & Cadinu, 2011; Tomasetto, Mirisola, Galdi, & Cadinu, 2015). In this way, parents' stereotypic beliefs about male and female abilities interact with children's individual characteristics (e.g. temperament, personality traits, talents, gender, etc.) and shape their beliefs about their children's ability (Chalabaev, Sarrazin, et al., 2013; Eccles & Harold, 1991; Fredricks & Eccles, 2004).

Many authors (Chalabaev, Sarrazin, et al., 2008; Chalabaev, Stone, Sarrazin, & Croizet, 2008; Heidrich & Chiviacowsky, 2015; Hively & El-Alayli, 2014; Stone & McWhinnie, 2008) have focused their attention on the effects of stereotype threat on women's sport performances, using a variety of tasks belonging to different sports such as golf, soccer, basketball and tennis, among others.

Chalabaev, Sarrazin, et al. (2008) conducted an experiment in which a group of female soccer players was asked to complete a soccer dribbling task. Three framing conditions were created: (a) in the athletic ability condition, the task was presented as a measure of 'personal factors correlated with athletic ability', defined as ability related to speed, strength and accuracy; (b) in the technical soccer ability condition, the task was framed as a measure of 'personal factors correlated with technical soccer ability', that is, the ability to lead the ball with rapidity and precision and (c) in the control condition, the task was presented as a measure of 'psychological factors'. The authors hypothesized that the stereotypes related to poor athletic ability and poor technical soccer ability of women would have a harmful effect on their performance compared to the control condition. They found that participants in the athletic ability condition had the poorest performance, followed by women in the technical soccer ability condition and the control group.

Some studies hypothesized that stereotype threat could affect also the learning of motor gestures. Heidrich and Chiviacowsky (2015) ask a sample of 24 women that never

played soccer to complete a dribbling task. Participants were divided into two condition: in the stereotype threat condition, the task was introduced as showing that "women normally perform worse than men", while in the nullified stereotype threat condition the task was presented by telling that "women normally perform similarly than men" . The practice phase consisted in an immediate performance (10 min) and in a delayed (day 2) retention tests. The authors found that on the immediate retention test, participants in the nullified stereotype threat condition outperformed the stereotype threat group and similar results were found the next day in the delayed retention test.

The Present Study

Several meta-analyses investigating the effect of the threat produced by a stereotype(Flore & Wicherts, 2015; Nadler & Clark, 2011) in a variety of domains have been published, such as race and intellectual performances (Appel, Weber, & Kronberger, 2015), elderly and memory(Horton, Baker, Pearce, & Deakin, 2008; Lamont, Swift, & Abrams, 2015), women and math(Doyle & Voyer, 2016; Nguyen & Ryan, 2008). However, no meta-analysis about the influence of stereotype threat on women's sport performances has been conducted yet. Differently from other meta-analyses about stereotype threat that concern a deficit in a cognitive response, the present meta-analysis is related to a deficit mainly in physical activity.

Moreover, a number of moderators were codified to find which variables moderate the effect of gender stereotype in sport activities for women. Collected studies have been coded for the type of experimental manipulation: a blatant (i.e. the experimental task was explicitly framed as showing gender differences) versus a subtle threat to women's ability (i.e.

the presence of a male experimenter). Some studies (Seitchik, Jamieson, & Harkins, 2014; Stone & McWhinnie, 2008) specifically investigated the differential impact of blatant and subtle manipulation of stereotype threat. Their results showed that these mechanisms undermine people's performance at the same extent, although through different processes. According to Stone and McWhinnie (2008), blatant threat influences females' performance through explicit monitoring, thus inducing a prevention focus orientation that motivates women to avoid the failure but, at the same time, interrupts the automatic elements of execution. Subtle threat, instead, seems to operate as a distraction that provokes a cognitive overload.

In some of the studies, the gesture that participants had to complete was new to them and thus the experimental procedure required a training phase (Heidrich & Chiviacowsky, 2015). In other studies, participants were asked to try a well learned movement such as in (Chalabaev, Sarrazin, et al., 2008) in which female soccer players were asked to complete an exercise that was very common to them: dribbling obstacles. In these studies no training phase was performed by participants. According to the cognitive load hypothesis, new gestures should be more prone to such threat. Moreover, according to the motivational explanation of the stereotype threat, the so called mere effort account (Harkins, 2006; Huber, Seitchik, Brown, Sternad, & Harkins, 2015)people are generally motivated to do their best both when performing a well-learned tasks and when trying a new gesture. Nevertheless, when the prepotent response is correct (well-learned tasks) this motivation leads to performance increase, while when the prepotent response is incorrect (novel tasks) this motivation leads to performance decrease. On the contrary,if the explicit monitoring hypothesis is proven reliable, well-learned gestures should be those more affected.

The physical and sport activities of the studies selected for the meta-analysis may be

distinguished according to the amount of cognitive load required. Cognitive load can be individuated as the amount of decisional processes and strategic thinking required to perform the task. Some of the studies employed simple moves such as to contract the quadriceps as fast and as forcefully as possible (Chalabaev, Brisswalter, et al., 2013), while in other studies stereotype threat was directed to activities like chess playing that are mainly based on strategic thinking (Maass, D'Ettole, & Cadinu, 2008) . Since the load in working memory has been proposed as an explanation of poor performance under stereotype threat, the amount of strategic thinking and cognitive load required has been dummy coded (low vs. high) and its effect considered as a possible moderator of the stereotype threat effect on physical and sport activities.

In some studies, the threat was directed toward the specific activity that participants were going to perform (e.g. “Now you are probably aware that there are gender differences in sports performance. Previous studies using this test of natural athletic ability have reported differences in the performance of men and women. So even though there may be gender differences on this test, we ask that you give 100% effort on the task so we can accurately measure your natural skills.”) (Stone & McWhinnie, 2008, p. 448), while in others, the threat was somewhat indirect, thus the threat to physical activity was provoked by the elicitation of the weakness of women in mathematics and spatial abilities (“As you may know, there has been some controversy about whether there are gender differences in math and spatial ability. As demonstrated by Levy, Leifheit-Limson (2009), the influence of stereotype valence is stronger when the stereotype content matches the outcome domain compared to when this is less related.

Finally, not all the sports that appear in the sample of studies are equally ‘masculine’. There is probably a degree of masculinity varying along a bipolar continuum. Using an on

line questionnaire we derived an index of masculinity for all the activities mentioned in the collected studies and thus distinguished sports that are 'suited for males' from sports 'adaptable also for women'. This distinction distinction was introduced as a possible moderator in the current meta-analysis. (4)

Method

Literature Search and Inclusion Criteria(6-7)

Studies were collected using online databases, through a combination of the following terms: stereotype threat, female, sport. Our search yielded studies published between 2006 and 2017. Studies were included if they reported an objective measure of performance and expressed the difference between the stereotype threat condition and the no-stereotype threat condition (control condition).

All the authors were contacted via e-mail and were asked if they had some unpublished data set testing the effect of the gender stereotype on sport performance in women. In this way, two more studies have been found.

Summary of the Meta-analytic Data Set

Our meta-analysis was conducted on 24 effects (k = 24), 21 of which were obtained from published articles, 1 from Doctoral Dissertations and the remaining 2 from unpublished studies that authors gave us through personal communications. (10)

Studies published in the same article were considered independent, the effect sizes

were computed separately and each was included in the meta-analysis. (9)

Some of the studies included a training phase before the test trial. In these cases, the size of the effect was derived only from the test trial.

We followed Hedges and Olkin (1985) approach, an extension of Glass (1976)method. They proposed to estimate the effect size using Cohen's *d*, which expresses the difference between the experimental group mean and the control group mean in standard units. Thus, each collected effect was converted to Cohen's *d* to be easily suited for interpretation. In this case, the effect size was computed as the standardized mean difference of performance between threatened and nonthreatened groups. (13) Studies included in the present meta-analysis are reported in Table A1, where each effect is expressed in Cohen's *d*.

Publication Bias(15–22)

Publication bias (also known as 'file-drawer problem') is the main weakness of a meta-analysis and occurs because many journal editors are often reluctant to publish 'non-significant' results, thus leaving the so-called null results in the file drawer (Ioannidis, 2005; Rosenthal, 1979; Sterling, 1959) This practice can lead to serious inflation of estimated effect-sizes in a meta-analysis (Sutton, Duval, Tweedie, Abrams, & Jones, 2000).

In order to assess if the file-drawer problem existed in our sample of studies, we built a funnel plot, putting each effect, expressed in Cohen's *d*, on the X-axis and standard errors on the Y-axis (see Figure 1).

---

insert Figure 1 about here

Although the present meta-analysis has been conducted on a small number of effects, the funnel plot does not show the typical asymmetry that appears when a publication bias exists: most studies report effects concentrated on the average, while few studies report null or very large effects. This means that there is no systematic loss of effects in published literature. Moreover, as expected, variability across studies diminishes as the sample number increases. This commonly happens when no hidden specific bias plays a role in the probability of a study that is to be published.

Publication bias was also assessed by drawing a funnel plot with trim and fill (Duval & Tweedie, 2000; Taylor & Tweedie, 1998). This method both removes (trims) those studies that cause publication bias and fills the excluded studies with their missing counterparts (Mavridis & Salanti, 2014). In the present meta-analysis, the number of missing studies needed for obtaining a symmetric funnel plot is zero. Again, research literature in this domain does not show a systematic tendency to avoid null results.

Egger, Davey Smith, Schneider, and Minder (1997)proposed a significance test for assessing publication bias based on the quantification of the asymmetry of a funnel plot. This method uses the standard normal deviate, that is, the odds ratio divided by its standard error. It is regressed against the estimate's precision, defined as the inverse of standard error. The null hypothesis of Egger's test maintains the symmetry of the funnel plot, while the alternative indicates that asymmetry exists (Sedgwick, 2013).

The regression test of the present meta-analysis showed no asymmetry in its funnel plot (t = -1.71, p = 0.10). (16)

Mean Effect Size(21)

Mean effect size was calculated with the package Metafor of R (version 3.5.2), using a random effect model.

The results of this meta-analysis showed a significant estimate of d = 0.33 (95% CI .03;.63; p<.05), meaning that there is a remarkable difference between threat and non-threat conditions across the studies. Furthermore, the 95% confidence interval varies from 0.02 to 0.61, highlighting a significant departure from the chance of the average effect size. A graphical representation of the distribution of effects and their relative standard error of estimation is presented in Figure 2.

(20)

Insert Figure 2 about here

Considering the characteristic of sport that concerns the energy expenditure related to the movement, chess should not be included in the current meta-analysis. Thus the meta-analysis was performed again not including the studies based on chess activities. As showed in Table 2, the mean effect size on k=22 studies is now d=.30 (95% CI: -.03; .64). Chess is nevertheless considered a sport activity in 24 out of 28 member states of the European Union, and the World Chess Federation (FIDE; Federation Internationale des Echecs) reports that this game is recognized as a sport globally in 117 countries. Although Chess does not involve energy expenditure, the game implies the presence of an adversary, a competition, a winner

and a loser and a certain physical fit to maintain concentration throughout the game. According to these considerations, we decided to maintain all the 24 initially selected studies in the current meta-analysis.

Variability of the Studies and Fail-Safe N

The variability of effect sizes across the sample was analyzed by means of the Q statistics that test whether the sample of effects is homogeneous, or the observed variability goes beyond what is expected by chance. The Q value is high enough to exclude the chance as an explanation ($Q_{(23)} = 110.06$; $p < .0001$).

Finally, the stability of the meta-analysis was assessed calculating Fail-Safe N(Orwin, 1983; Rosenberg, 2005), which is the estimated number of non-significant results in the file drawers that would be required to turn a significant meta-analytic finding into one that is non-significant (Schwarzer, 1989). Conventionally, the finding is considered robust when the Fail-Safe N exceeds at least 6 times the number of effects included in the meta-analysis. In the case of the present work, the calculation of the Fail-Safe N yields the value of 411 studies, which is higher than the critical limit of 6×27=162 studies. It is then possible to conclude that the meta-analytic finding obtained is robust. As suggested by Rosenberg (2005), this means that the publication bias, if it exists, may be safely ignored.

Potential Moderators

In relation to the heterogeneity, the obtained value of Cochrane's Q brings to the rejection of the hypothesis of homogeneity among the studies. Even if the number of the

effects is quite low, six moderators were hypothesized: blatant versus subtle manipulation, new versus well-learned, cognitive versus physical tasks, domain-directed versus indirect threat, and the degree to which each study employs masculine versus feminine sport. Each moderator, except typicality, was treated as a dummy variable in terms of presence (1) or absence (0) of the characteristics. Criteria were jointly discussed and then two of the authors coded the studies separately.Agreement between coders wascomplete. The stereotypicality of the various sport activities employed in the studies included in the present meta-analysis was assessed separately via an online questionnaire. Specifically, people were contacted through social media. They were asked to indicate on a scale ranging from 1 to 5 to what extent each physical activity (e.g. dribbling) and the correspondent sport activity (e.g. soccer) that appears in the selected studies were commonly considered masculine or feminine by society (1 = absolutely masculine, 5 = absolutely feminine). Forty-one participants (9 males, 32 females, mean age = 26.44 years) answered this questionnaire. Table 1 reports the average stereotypicality of each activity.

Insert Table 1 about here

To test if people who answered the questionnaire were paying enough attention, the questionnaire included a question built following the instructional manipulation check that measures whether participants are reading the instruction (Oppenheimer, Meyvis, & Davidenko, 2009) and positioned at the end of the survey, before the sociodemographic questions. In particular, if participants were careful in reading the instructions, they had to

correctly select all the boxes. If people failed the instructional manipulation check, their answers were considered unreliable. The degree of stereotypicality was then considered as a possible moderator of the effect.

Among the possible moderators, only the distinction between more and less feminine sports proved to be a significant moderator of the heterogeneous distribution of effect sizes; this explains about the 32% of the variability (R^2 = .319).

Insert Table 2 about here

Based on the results of the present study, it seems that stereotype threat has a stronger effect for sports considered 'masculine' such as golf, soccer and strength exercises, while in more 'feminine' activities (e.g. tennis or balance), its effect is significantly weaker.

One of the most reliable explanations of stereotype threat is based on the cognitive load that comes with an awareness of being judged on performance in a domain that is considered problematic for the group with which one identifies(Schmader & Johns, 2003). In particular, the cognitive load would impair the ability to regulate attention during complex tasks, where, as mentioned earlier, it is necessary to inhibit thoughts, behaviours and feelings that are not related or are negatively related to the current task (Schmader, et al., 2008). This explanation was hypothesized in relation to stereotype threat regarding cognitive performance, but it was not tested for sport activities. Those activities requiring cognitive resources concern a new gesture since they involve a large amount of attentional resources for gesture

planning. The degree of learning of the sport gesture was codified as a moderator, but it explains a very small amount of variance. This means that the cognitive load explanation, which refers to cognitive threat, might not be a valid explanation for stereotype threat in physical activity. However, neither the explicit monitoring theory (Beilock & Carr, 2001)moderated this effect, Suggesting that both hypothesesremain plausibleand that sensorimotor performance may be reducedthrough different mechanism.

Discussion and Conclusions (24)

The effect of the awareness of belonging to a stereotyped group has received wide attention in literature since the phenomenon was first described in the mid-1990s (Steele & Aronson, 1995). Most of the collected data show how belonging to a discriminated group may lead to a poorer cognitive performance in some kinds of tasks. The effect has been reliably observed in relation to mathematical and logical abilities for women, IQ-related tasks for ethnic minorities and memory-related abilities in elderly. The published studies have been summarized in a number of meta-analyses.

Comparison with Other Meta-analyses about Stereotype Threat

At this point, we need to understand if the observed effect is comparable in size with that which is commonly observed in the meta-analysis conducted on cognitive performances. A comparison with all the systematic reviews showed that the finding of the present meta-analyses is in line with what has already been found in different but related domains.

To assess if the mean effect size is comparable with what we actually know about stereotype threat, a systematic search of meta-analyses, seeking the effect of stereotype activation on the performance of minority groups, was conducted. Twelve systematic

reviews were found. Mean effect size (expressed in Cohen's *d*), confidence interval and standard deviation of the effects were annotated for each meta-analysis and then reported in Table 3, together with the results of the present study. (16-23)

Insert Table 3 about here

In Doyle and Voyer (2016) two separate meta analytic effects were computed: one synthesizing the effect of gender and the other one the effect of ethnicity on performance on math and spatial performance. We considered both effects in the analysis.

The work of Walton and Spencer (2009)does not present confidence interval and standard deviation. Thus, it was not included in the subsequent analysis.

The weighted mean effects size of the selected meta-analysis has been calculated, together with the confidence interval around this estimate. The synthesis of the entire set of available meta-analysis on stereotype threat brought to an average effect of $d = 0.29$ (C.I. = 0.20, 0.38). The result of the current study is in line with the other findings on this topic.

Five of the 11 meta-analyses specifically tested the effects of the stereotype threat in women. These systematic reviews investigate the effect of stereotype activation on a variety of cognitive skills: math, spatial abilities, cognitive abilities in general and test performance. All the systematic reviews, except one (Doyle & Voyer, 2016), showed that women under stereotype threat perform worse than in control conditions. As before, the weighted mean effect size of the meta-analyses of studies conducted on women samples has been computed bringing to a synthetic $d = 0.25$ (C.I. = 0.15, 0.37). As previously observed, the effect size of

the present meta-analysis falls into the confidence interval.

Considering that the mean effect size computed in the present meta-analysis refers to an impairment which is not specifically cognitive but mainly in a physical activity, it can thus be stated that the threat for women directed toward physical skills and power may be more effective than the threat directed toward mental abilities. Nevertheless, its size is comparable with results found in previous quantitative syntheses of the effect of the activation of gender stereotype in women.

Conclusions

The present meta-analytic study is quite different from others because the effect of stereotype threat might not be necessarily related to the cognitive domain but rather to the characteristics of sport activity, such as gender typicality. Sport activities are historically considered part of a typically male domain. For this reason, we expected that priming the weaker-sex stereotype would result in a poorer performance in women. Literature has been thoroughly reviewed, and a number of studies directly testing this hypothesis sufficient for conducting a meta-analysis have been found.

The meta-analytic synthesis led to an average effect size of $d = 0.33$ that is in line with the effects of stereotype threat that can commonly be found in the literature. Furthermore, no publication bias was found. Among the hypothesized moderators, only the stereotypicality of the activity seems to have an influence on this phenomenon: gender stereotype threat is stronger for masculine and weaker for feminine sports.

Since the sample of the studies included in the present meta-analysis is limited, all comparisons between subsets of studies present a reduced statistical power.

Certainly it would have been interesting to go further in the analysis of moderators of the effect of gender stereotype threat. Again, the reduced number of available studies played against such a curiosity. Variables like domain identification or complexity of the physical task cannot be coded a posteriori and few published studies assessed these construct. For this reason, future research should directly manipulate the variables that may affect the impact of the gender stereotype on sport performance. Particularly interesting would be charging the activities with a varying amount of cognitive load to test whether the effect of stereotype threat depends on this variable. If the effect of stereotype threat varies depending on cognitive load, the most influential hypothesis about the origin of stereotype threat could be extended to physical activities.

References

Abbate, C. S., Boca, S., & Bocchiaro, P. (2004). Stereotypes in Persuasive Communication: Influence Exerted by a Disapproved Source. *Journal of Applied Social Psychology, 34*, 1191-1207.

Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: review and meta-analysis. *Frontiers in psychology, 6*, 900.

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of experimental psychology: General, 130*, 701.

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin, 32*, 1059-1071.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of experimental social psychology, 40*, 247-255.

Chalabaev, A., Brisswalter, J., Radel, R., Coombes, S. A., Easthope, C., & Clément-Guillotin, C. (2013). Can stereotype threat affect motor performance in the absence of explicit

monitoring processes?: Evidence using a strength task. *Journal of Sport and Exercise Psychology, 35*, 211-215.

Chalabaev, A., Sarrazin, P., Fontayne, P., Boiché, J., & Clément-Guillotin, C. (2013). The influence of sex stereotypes and gender roles on participation and performance in sport and exercise: Review and future directions. *Psychology of Sport and Exercise, 14*, 136-144.

Chalabaev, A., Sarrazin, P., Stone, J., & Cury, F. (2008). Do achievement goals mediate stereotype threat?: An investigation on females' soccer performance. *Journal of Sport and Exercise Psychology, 30*, 143-158.

Chalabaev, A., Stone, J., Sarrazin, P., & Croizet, J.-C. (2008). Investigating physiological and self-reported mediators of stereotype lift effects on a motor task. *Basic and Applied Social Psychology, 30*, 18-26.

Chasteen, A. L., Bhattacharyya, S., Horhota, M., Tam, R., & Hasher, L. (2005). How feelings of stereotype threat influence older adults' memory performance. *Experimental aging research, 31*, 235-260.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological review, 96*, 608.

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615-1628.

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47*, 103-116.

Duriez, B., & Soenens, B. (2009). The intergenerational transmission of racism: The role of right-wing authoritarianism and social dominance orientation. *Journal of research in Personality, 43*, 906-909.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel ‐ plot ‐ based method of testing and adjusting for publication bias in meta ‐ analysis. *Biometrics, 56*, 455-463.

Eccles, J. S., & Harold, R. D. (1991). Gender differences in sport involvement: Applying the Eccles' expectancy-value model. *Journal of Applied Sport Psychology, 3*, 7-35.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629-634.

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.

Fredricks, J. A., & Eccles, J. S. (2004). Parental influences on youth involvement in sports.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological bulletin, 132*, 692.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher, 5*, 3-8.

Gray, J. R. (2004). Integration of emotion and cognitive control. *Current directions in psychological science, 13*, 46-48.

Hardin, M., & Greer, J. D. (2009). The influence of gender-role socialization, media use and sports participation on perceptions of gender-appropriate sports. *Journal of Sport Behavior, 32*, 207.

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of personality and social psychology, 91*, 436.

Hedges, L. V., & Olkin, I. (1985). CHAPTER 5 - Estimation of a Single Effect Size: Parametric and Nonparametric Methods. In *Statistical Methods for Meta-Analysis* (pp. 75-106). San Diego: Academic Press.

Heider, F. (2013). *The psychology of interpersonal relations*: Psychology Press.

Heidrich, C., & Chiviacowsky, S. (2015). Stereotype threat affects the learning of sport motor skills. *Psychology of Sport and Exercise, 18*, 42-46.

Hermann, J. M., & Vollmeyer, R. (2016). "Girls should cook, rather than kick!"–Female soccer players under stereotype threat. *Psychology of Sport and Exercise, 26*, 94-101.

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58*, P3-P11.

Hively, K., & El-Alayli, A. (2014). "You throw like a girl:" The effect of stereotype threat on women's athletic performance and gender stereotypes. *Psychology of Sport and Exercise, 15*, 48-55.

Horton, S., Baker, J., Pearce, G., & Deakin, J. M. (2008). On the malleability of performance: Implications for seniors. *Journal of Applied Gerontology, 27*, 446-465.

Huber, M. E., Brown, A. J., & Sternad, D. (2016). Girls can play ball: Stereotype threat reduces variability in a motor skill. *Acta psychologica, 169*, 79-87.

Huber, M. E., Seitchik, A. E., Brown, A. J., Sternad, D., & Harkins, S. G. (2015). The effect of stereotype threat on performance of a rhythmic motor skill. *Journal of experimental psychology: human perception and performance, 41*, 525.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *Plos Medicine, 2*, 696-701.

Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and aging, 30*, 180.

Levy, B. R., & Leifheit-Limson, E. (2009). The stereotype-matching effect: greater influence on functioning when age stereotypes correspond to outcomes. *Psychology and Aging*, *24*(1), 230.

Maass, A., D'Ettole, C., & Cadinu, M. (2008). Checkmate? The role of gender stereotypes in the ultimate intellectual sport. *European Journal of Social Psychology, 38*, 231-245.

Martiny, S. E., Gleibs, I. H., Parks-Stamm, E. J., Martiny-Huenger, T., Froehlich, L., Harter, A.-L., & Roth, J. (2015). Dealing with negative stereotypes in sports: The role of cognitive anxiety when multiple identities are activated in sensorimotor tasks. *Journal of Sport and Exercise Psychology, 37*, 379-392.

Mavridis, D., & Salanti, G. (2014). Exploring and accounting for publication bias in mental health: a brief overview of methods. In: Royal College of Psychiatrists.

Nadler, J. T., & Clark, M. (2011). Stereotype threat: A meta‐analysis comparing African Americans to Hispanic Americans. *Journal of Applied Social Psychology, 41*, 872-890.

Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of applied psychology, 93*, 1314.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29*, 782-789.

O'Bryan, M., Fishbein, H. D., & Ritchey, P. N. (2004). Intergenerational transmission of prejudice, sex role stereotyping, and intolerance. *Adolescence, 39*, 407.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of experimental social psychology, 45*, 867-872.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of educational statistics, 8*, 157-159.

Parsons, J. E., Adler, T. F., & Kaczala, C. M. (1982). Socialization of achievement attitudes and beliefs: Parental influences. *Child Development*, 310-321.

Pinel, E. C. (1999). Stigma consciousness: the psychological legacy of social stereotypes.

Journal of personality and social psychology, *76*, 114.

Pyszczynski, T., & Greenberg, J. (1983). Determinants of reduction in intended effort as a strategy for coping with anticipated failure. *Journal of research in Personality, 17*, 412-422.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science, 14*, 287-290.

Rosenberg, M. S. (2005). The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution, 59*, 464-468.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin, 86*, 638.

Sanchez, A. N. P., Gates, S.R., King, C.J., Larson, A. D., Lutz, W.R., Sanker, J.E., Fondevilla, M.C.E., and Bainum, C.K. (2011). Shooting Hoops: why men's presence negatively affects women's athletic performance. . In.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of personality and social psychology, 85*, 440.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological review, 115*, 336.

Schwarzer, R. (1989). Meta-analysis programs. *Freie Universitat Berlin*.

Sedgwick, P. (2013). Simple linear regression. *Bmj-British Medical Journal, 346*.

Seitchik, A. E., Jamieson, J., & Harkins, S. G. (2014). Reading between the lines: Subtle stereotype threat cues can motivate performance. *Social Influence, 9*, 52-68.

Sherrow, V. (1996). *Encyclopedia of women and sports*: Abc-Clio Incorporated.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology, 35*, 4-28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist, 52*, 613.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology, 69*, 797.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In *Advances in experimental social psychology* (Vol. 34, pp. 379-440): Elsevier.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American statistical*

association, *54*, 30-34.

Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of experimental social psychology*, *44*, 445-452.

Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *BMJ, 320*, 1574-1577.

Taylor, S., & Tweedie, R. (1998). Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analyses. *Fort Collins, CO: Colorado State University*.

Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology, 47*, 943.

Tomasetto, C., Mirisola, A., Galdi, S., & Cadinu, M. (2015). Parents' math–gender stereotypes, children's self-perception of ability, and children's appraisal of parents' evaluations in 6-year-olds. *Contemporary Educational Psychology, 42*, 186-198.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of memory and language, 28*, 127-154.

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20*, 1132-1139.

Woolum, J. (1998). *Outstanding women athletes: who they are and how they influenced sports in America*: Greenwood Publishing Group.

Wout, D. A., Shih, M. J., Jackson, J. S., & Sellers, R. M. (2009). Targets as perceivers: How people determine when they will be negatively stereotyped. *Journal of personality and social psychology, 96*, 349.

Zimmerman, J. (2011). *Raising our athletic daughters: How sports can build self-esteem and save girls' lives*: Main Street Books.

Table A1 - Studies Included in the Present Meta-analysis

Study	*Conditions*	*Sport Domain*	*N*	*d*	*Publication Status*	*Manipulation (blatant vs subtle)*	*Gesture (well-learned vs new)*	*Threat (domain-directed vs domain uncorrelated)*	*Performance (sensorimotor vs cognitive)*
Chalabaev (2016)	Threat No threat	Soccer	20	-0.56	Unpublished	Blatant	Well-learned	Domain-directed	Sensorimotor
Chalabaev (2016)	Threat No threat	Soccer	43	0.55	Unpublished	Blatant	Well-learned	Domain-directed	Sensorimotor
~~Chalabaev (2006)~~	~~Adhesion to stereotype No adhesion~~	~~Soccer~~	~~68~~	~~0.47~~	~~Doctoral Dissertation~~			Domain-directed	Sensorimotor
~~Chalabaev (2006)~~	~~Physical Ability Sport Psychology~~	~~Soccer~~	~~34~~	~~1.19~~	~~Doctoral Dissertation~~			Domain-directed	Sensorimotor
~~Chalabaev (2006)~~	~~Technical Ability Sport Psychology~~	~~Soccer~~	~~34~~	~~0.58~~	~~Doctoral Dissertation~~				Sensorimotor
Chalabaev, Sarrazin, Stone & Cury (2008)	Athletic Ability Sport Psychology	Soccer	34	0.88	Published	Blatant	Well-learned	Domain-directed	Sensorimotor
Chalabaev, Sarrazin, Stone & Cury (2008)	Technical Ability Sport Psychology	Soccer	34	0.50	Published	Blatant	Well-learned	Domain-directed	Sensorimotor
Chalabaev, Stone, Sarrazin & Croizet (2008)	Threat No threat	Balance test	63	0.12	Published	Blatant	New gesture	Domain-directed	Sensorimotor
Chalabaev et al. (2013)	Threat No threat	Kinesiology	40	0.82	Published	Blatant	Well-learned	Domain-directed	Sensorimotor
Heidrich (2006)	Threat No threat	Soccer	36	0.84	Doctoral Dissertation	Blatant	Well-learned	Domain-directed	Sensorimotor
Heidrich & Chiviacowski (2015)	Threat No threat	Soccer	24	1.47	Published	Blatant	New gesture	Domain-directed	Sensorimotor
Heidrich & Chiviacowski (2015)	Threat No threat	Soccer	24	1.41	Published	Blatant	New gesture	Domain-directed	Sensorimotor
Heidrich & Chiviacowski (2015)	Threat No threat	Soccer	24	1.42	Published	Blatant	New gesture	Domain-directed	Sensorimotor
									Sensorimotor

Study	*Conditions*	*Sport Domain*	*N*	*d*	*Publication Status*	*Manipulation (blatant vs subtle)*	*Gesture (well-learned vs new)*	*Threat (domain-directed vs domain uncorrelated)*	*Performance (sensorimotor vs cognitive)*
Hermann & Vollmeyer (2016)	Threat No threat	Soccer	36	0.98	Published	Blatant	Well-learned	Domain-directed	Sensorimotor
Hively & El-Alayli (2014)	Threat No threat	Basketball	30	0.24	Published	Blatant	Well-learned	Domain-directed	Sensorimotor
Huber et al. (2016)	Threat No threat	Ball Bouncing Task	24	-1.76	Published	Blatant	New gesture	Domain-uncorrelated	Sensorimotor
Huber et al. (2015)	Threat No threat	Ball Bouncing Task	34	-0.70	Published	Blatant	New gesture	Domain-uncorrelated	Sensorimotor
Huber et al. (2015)	Threat No threat	Ball Bouncing Task	37	0.69	Published	Blatant	Well-learned	Domain-uncorrelated	Sensorimotor
Laurin (2013)	Gender threat No threat	Basketball	53	0.51	Published	Blatant	Well-learned	Domain-directed	Sensorimotor
Laurin (2017)	Threat No threat	Athletic	41	-1.37	Published	Blatant	New gesture	Domain-directed	Sensorimotor
Maas, D'Ettole & Cadinu (2008)	Threat No threat	Chess	42	0.68	Published	Subtle	Well-learned	Domain-directed	Cognitive
Maas, D'Ettole & Cadinu (2008)	Threat No threat	Chess	41	0.44	Published	Subtle	Well-learned	Domain-directed	Cognitive
Martiny et al. (2015)	Single identity Dual identity	Soccer	70	0.58	Published	Subtle	Well-learned	Domain-directed	Sensorimotor
Rothgerber & Wolsiefer (2014)	Threat No threat	Chess	199	0.78	Published	Subtle	Well learned	Domain-directed	Cognitive
Sanchez et al. (2011)	Threat No threat	Basketball	33	0.89	Published	Subtle	Well-learned	Domain-directed	Sensorimotor
Stone & McWhinnie (2008)	Gender threat Control	Golf	110	0.34	Published	Blatant	New gesture	Domain-directed	Sensorimotor
Stone & McWhinnie (2008)	Gender threat Control	Golf	110	0.39	Published	Subtle	New gesture	Domain-directed	Sensorimotor

Table 1 – Scores of the masculinity linked to the following sports/physical activities

Sports/Physical Activity	*Score*
Basket	1.88
Tennis	3.06
Golf	1.41
Video-terminal exercise	2.18
Balance exercise	4.06
Soccer	1.12
Strength exercise	1.29
Jump	3.47
Chess	2.06

Table 2.

Potential Moderators

Moderators	*Z-score*	*p-value*
Experimental Manipulation (Blatant/Subtle)	-1.03	0.29
Well-learned/New Gesture	-0.05	0.96
Cognitive/Sensorimotor Task	1.13	0.26
Domain-directed/Domain-uncorrelated Threat	-1.63	0.09
Masculinity/Femininity	-2.54	0.01**

Table 3

Meta-analyses about Stereotype Threat

Authors	*d*	SE	*K*	*95% CI*
[60] Appel, Weber & Kronberger (2015)	.63	.12	19	[0.40; 0.86]
[61] Doyle & Voyer (2016)	.32	.04	109	[0.24; 0.40]
[61] Doyle & Voyer (2016)	.12	.09	22	[-0.06; 0.30]
[62]Flore & Wicherts (2015)	.22	.03	47	[0.08; 0.32]
[63] Horton, Baker, Pearce & Deakin (2008)	.38	.09	17	[0.20; 0.56]
[64] Lamont, Swift & Abrams (2015)	.28	.09	82	[0.10: 0.44]
[65] Nadler & Clark (2011)	.52	.05	23	[0.42; 0.62]
[66] Nguyen & Ryan (2008)	.21	.06	72	[0.08; 0.36]
[66] Nguyen & Ryan (2008)	.32	.07	44	[0.19; 0.49]
[67] Picho, Rodriguez & Finnie (2013)	.24	.05	103	[0.14; 0.35]
[68] Walton & Spencer (2009)	.44	NA	39	NA
Current meta-analysis	.32	.15	24	[0.02; 0.61]

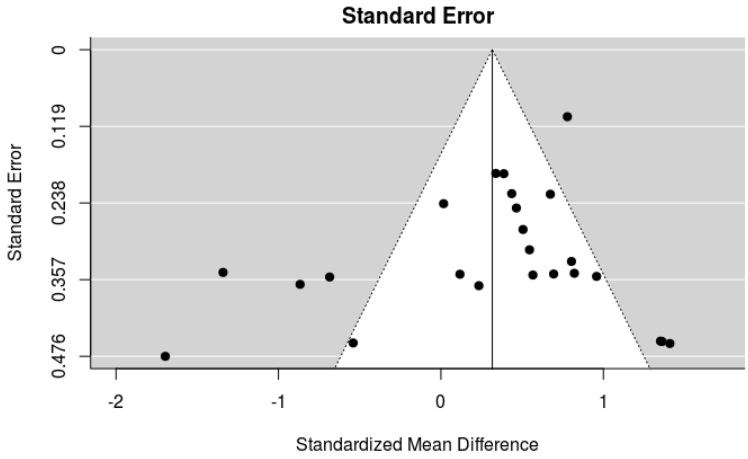


Figure 1. Funnel plot assessing the existence of publication bias.

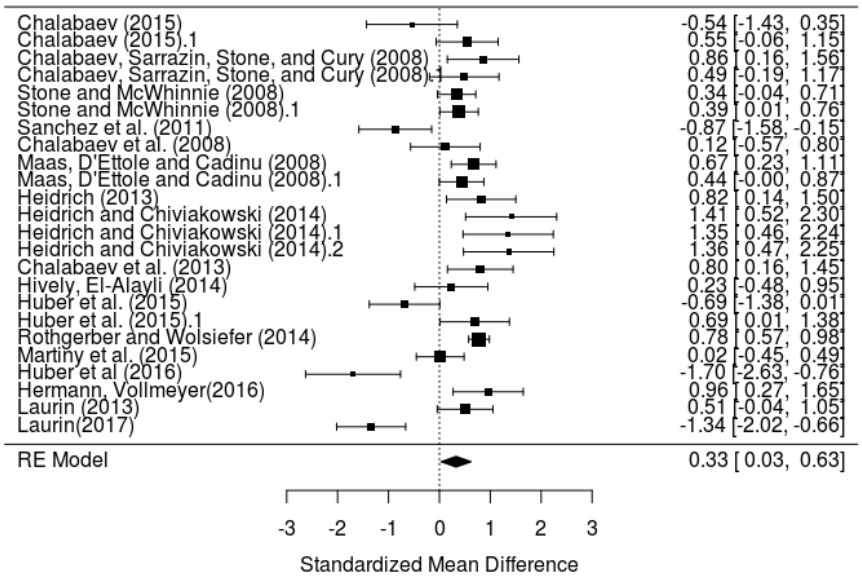


Figure 2. Forest plot of the studies included in the current meta-analysis

Highlights

- Stereotype threat affects sport performances of women.
- Women experience stereotype threat especially when the sport is typically masculine.
- Both cognitive overload hypothesis and explicit monitoring hypothesis are plausible in sensorimotor task.