



Rebounding from threat: Examining the relationship between memory recall and self-affirmation in female athletes under social identity threat

**N.A.J. van Doremalen**

SNR: 2014143

**Bachelor's thesis**

Psychology

Major Social Psychology

Tilburg School of Social and Behavioral Sciences

Tilburg University, Tilburg

Supervisor: M. Bender, Ph.D.

Final version

July 2020

## Abstract

Despite the fact that autobiographical memories have been shown to be pertinent to numerous domains of psychological functioning, few researchers have yet investigated whether these memories can also play a role in self-affirmation. I conducted an online experiment to assess whether self-affirmation through the recall of mastery autobiographical memories can help female athletes participating in a predominantly male-dominated sport cope with gender stereotype threat. Participants, who were 193 female ice hockey players, were subjected to a gender stereotype of lower female athleticism and subsequently instructed to recall either mastery or routine autobiographical memories. The results indicate no statistically significant difference in personal self-esteem between participants who either recall mastery or routine autobiographical memories. The results also indicate that global self-esteem and group identification strength, which have been proposed as moderators of frequent occurrence in previous research, do not seem to moderate the relationship between the type of memory recall (i.e., mastery or routine autobiographical memories) and personal self-esteem. My findings make an important contribution to the current body of literature on threat management and social identity threat and highlight possible boundaries of the fruitfulness of the use of autobiographical memories as a particular machinery in self-affirmation.

*Keywords:* autobiographical memory, self-affirmation, female athletes, stereotype threat, social identity threat

It appears to be the case that gender differences perceived in athletics are not as well-grounded as they give the impression of being. The fact that on average women run less fast than men seems to go without saying, but it has actually been shown that gender only predicts 5% of the variance in physical abilities (Eagly, 1995). Be that as it may, Stereotypes of male superiority as athletes persist in the physical activity domain and the threat brought about by such stereotypes has far-reaching consequences for female athletes. In fact, previous research on stereotype threat has shown that individuals can underperform to a great extent at pursuits when pondering about the negative performance anticipations for their social group (Steele, 1997, 1998). In sports, emphasizing a woman's lower athletic ability compared to men may seriously derogate her athletic performance, not to mention her self-esteem may be severely compromised in a negative fashion. Previous research has shown that, generally, people have a basic need to uphold a positive self-evaluation (Steele, 1988; Tesser, 1988) and diverse strategies to achieve the aforementioned have been put forward. One such strategy is self-affirmation, which implies that people uphold a positive self-evaluation by displacing their attention to a valued domain of the self that is not threatened and thereby render threats to the self less appalling. While people can affirm themselves in many different ways, contemporary research has accentuated recalling a certain type of autobiographical memory as a specific machinery of self-affirmation (Tavitian-Elmadjian, Bender, Van de Vijver, Chasiotis, & Harb, 2020). Although more research on this specific machinery is necessary, recalling the latter type of memory might be an extremely effective way for female athletes to fight down negative threats to identity and to deal with the adversity brought about by stereotypes of lower female athleticism.

### **What is Stereotype Threat**

Over the past decades, studies on stereotype threat have proliferated. Steele, Spencer, and Aronson (2002) offered the following definition of stereotype threat: "When a negative

stereotype about a group that one is part of becomes personally relevant, usually as an interpretation of one's behavior or an experience one is having, stereotype threat is the resulting sense that one can then be judged or treated in terms of the stereotype or that one might do something that would inadvertently confirm it" (Steele et al., 2002, p. 389). The jeopardy of being looked at in terms of those negative stereotypes may prompt a troublesome state that could potentially negatively affect performance in realms as disparate as elderly people and memory exercises (Mazerolle et al., 2012; Chasteen, Bhattacharyya, Horhota, Tam, & Hasher, 2005; Hess, Auman, Colcombe, & Rahhal, 2003), the race gap in academics (Steele & Aronson, 1995; Steele et al., 2002), and decision making concerning finances (Carr & Steele, 2010). Specifically, the gender gap in the math domain has been studied and debated on a large scale (Doyle & Voyer, 2016; Nguyen & Ryan, 2008; Picho et al., 2013; Stoet & Geary, 2012; Walton & Cohen, 2003; Walton & Spencer, 2009).

From the above cited definition it follows that stereotype threat is a situational threat (Steele et al., 2002). It originates from situational cues indicating that an adverse stereotype is now pertinent as an interpretation of one's behavior in a particular environment. Thence Steele et al. (2002) advocate that, even though just about everyone can experience stereotype threat at some point (i.e., every person has an identity for which adverse stereotypes exist), the threat would solely be experienced by those people and in those circumstances to which the stereotype is appropriate. This automatically implies that when people are doing things in circumstances to which the stereotype does not apply, they would not experience the threat.

### **The Multi-Threat Framework**

Despite the fact that the negative consequences of stereotype threat are far-reaching (e.g., underperformance and diminished interest in domains and careers relevant to the stereotype (Steele & Aronson, 1995; Davies, Spencer, Quinn, & Gerhardstein, 2002), adverse

health consequences (Blascovich, Spencer, Quinn, & Steele, 2001), and decreased self-efficacy (Aronson & Inzlicht, 2004)), it has formerly been conceptualized as a singular construct. In spite of that, contemporary research gives prominence to the variability in conceptualizations of stereotype threat, disclosing stereotype threat is frequently utilized to describe and explicate processes and phenomena that appear to be significantly distinct (Shapiro & Neuberg, 2007). This contemporary prevailing research emphasizes that different negatively stereotyped groups appear to be vulnerable for different stereotype threats (e.g., different stereotypes related with sports will in all probability have different consequences on the sport performance of women; different stereotypes related with academics will in all probability have different consequences on the academic performance of individuals). A single stereotype threat is most likely out of the question and this effect will manifest itself differentially under various situations.

Taking this unevenness into account, Shapiro and Neuberg (2007) introduced the Multi-Threat Framework comprising six qualitatively distinct stereotype threats. The latter excavate from the crossing of two dimensions—the source of the threat (i.e., who will criticize these stereotype-suitable actions: outgroup others, ingroup others, or the individual him- or herself) and the target of the threat (i.e., who will one's stereotype-suitable actions pore on: the group or the individual him- or herself). This framework is nowadays reinforced by several studies that acknowledge the distinct forms of stereotype threats and their influences on different negatively stereotyped groups (Shapiro & Neuberg, 2007; Shapiro, 2011; Shapiro, Williams, & Hambarchyan, 2013; Wout, Danso, Jackson, & Spencer, 2008).

The stereotype threat literature struggles with a large share of heterogeneity amidst effect sizes. Idem, some studies successfully corroborate predictions drawn from stereotype threat theory regarding moderating variables and some fail to corroborate. According to Flore (2018) and in line with the Multi-Threat Framework, this substantial heterogeneity in data

most likely emanates from the covering of diverse contexts, social groups, and trials by different studies. That is, previous research may yield fused results in unique studies, not because the stereotype threat effect does not exist in general but because there is low risk for that specific stereotype threat in that clear-cut context (Shapiro & Neuberg, 2007). If one is aware of the variability in conceptualizations of stereotype threat, then it follows logically that findings from some studies will not easily generalize to different contexts and social groups. The variety of contexts and individual differences moderate the effects of the multiple types of stereotype threats (Shapiro, 2011). A more consistent subset of variables will in all probability result in a less heterogeneous collection of data and, therefore, contribute to the vigour of the results and the generalizability of the stereotype threat phenomenon (Flore, 2018).

### **Divergent Findings Surrounding Stereotype Threat**

Plentiful studies have in point of fact shown that stereotype threat can lead to decrements in performance in an unimaginably large amount of domains (i.e., focusing on tasks based on cognition) (Flore & Wicherts, 2015; Nguyen & Ryan, 2008; Walton & Cohen, 2003). However, even though the existence of the stereotype threat effect seems to be accepted over a large area, these affirmative effects diverge from recent failings to find such an effect (Ganley et al., 2013; Cherney & Campbell, 2011). Recent replication efforts that were unsuccessful in finding a robust effect provide reasons to distrust the, repeatedly asserted to be, sturdy stereotype threat phenomenon (Gibson, Losee, & Vitiello, 2014; Moon & Roeder, 2014; Finnigan & Corker, 2016; Flore, Mulder, & Wicherts, 2018). According to Flore (2018), the genuine probability of publication bias could be one of the possible reasons for questioning the robustness of the stereotype threat effect.

Publication bias can lead to grave exaggerations of effect-sizes by predominantly publishing significant results (Ioannidis, 2005; Rosenthal, 1979; Sterling, 1959). According to Ioannidis (2005), publication bias is extraordinarily likely to misrepresent a research field if the latter (1) is well-liked (i.e., features many studies) and attracts the attention of politicians and finance; (2) implicates studies with dozens of choices to discover interesting results and analyze the data; (3) implicates studies with small sample sizes; (4) pays attention to heaps of relations, and (5) examines small effect sizes. All five features are, in some studies more than in others, present in virtually all stereotype threat research concentrating on the cognitive domain.

With the former in mind, the field appears to be exceptionally susceptible to publication bias which might have misconstrued our notion of the phenomenon (i.e., overly positive) (Zigerell, 2017; Ganley et al., 2013; Stoet & Geary, 2012). Additionally, a number of methodological matters like Questionable Research Practices, Hypothesizing After Results are Known, and underpowered studies are expected to worsen obstacles related to publication bias (Doyle & Voyer, 2016; Bakker et al., 2012). These concerns lead to oftentimes complicated to replicate outcomes (Bakker et al., 2012), refer to the complexity of the stereotype threat phenomenon, and leave room for a lot of ambiguities. By all means, no researcher seems to dispute the fact that stereotypes might heavily affect an individual's life; yet, based on the results of previous studies on stereotype threat (e.g., Gibson et al., 2014; Flore, 2018)), there are a number of reasons to question whether the stereotype threat effect is as sound as it is not infrequently asserted to be.

### **Specific Stereotype Threat: Gender and Sport**

An example of a negatively stereotyped group and domain are female athletes participating in a sport predominantly male-dominated. Hitherto, most of the work on

stereotype threat has been limited to tasks based on cognition (Mazerolle et al., 2012; Steele & Aronson, 1995) and the stereotype threat literature fixating on sports has received scant scrutiny (Hively & El-Alayli, 2014). In the meantime, more and more women are competing in conventionally male sports (Zimmerman, 2011) such as football, wrestling, rugby, and ice hockey characterized by aspects such as strength, physical contact, and aggressiveness (Hardin & Greer, 2009; Chalabaev, Sarrazin, Fontayne, Boiché, & Clément-Guillotin, 2013). According to the National Federation of State High School Associations (NFHS), women's participation in conventionally male sports has grown exponentially in the past ten years. From 2008-09 to 2018-19, the number of women playing football grew from 759 to 2404, ice hockey from 8261 to 9650, rugby from 100 to 678, and the number of women participating in wrestling from 6025 to 21124 (in the U.S. alone). Notwithstanding, negative stereotypes about women (e.g., "You throw like a girl"; "Girls should cook, rather than kick!") persist in the physical activity domain. A handful of authors have investigated the effects of these negative stereotypes on women's sport performance in a variety of sports (e.g., soccer (Chalabaev, Sarrazin, Stone, & Cury, 2008; Heidrich & Chiviacowsky, 2015), golf (Stone & McWhinnie, 2008), tennis and basketball (Hively & El-Alayli, 2014)) and all have found a negative effect of the threat produced by a stereotype. That is, all the authors have found that the threat brought about by a stereotype negatively impacts women's performance in sport.

The world of sports has for centuries been deemed an exclusively masculine domain in which women continue to face stereotype vulnerability. Gender stereotypes of lower female athleticism are omnipresent in media coverages where women are primarily depicted as sex objects (to the exclusion of their athleticism) (Knight & Giuliano, 2001), referees' decisions during matches where they tend to penalize women more than men (gender stereotypes could effectively affect decision making) (Souchon, Coulomb-Cabagno, Traclet, & Rasclé, 2004), and funding of various sports programs where, for example, men's football



and basketball took up an average of almost 75% of all (Division I-A) budgets in 2000 (Hardin, Simpson, Whiteside, & Garris, 2007; Priest, 2003). The above mentioned sources, among others, investigate the ubiquitous held conviction that, compared to men, women have inferior athletic ability (Knight & Giuliano, 2001; Biernat & Vescio, 2002) and support the notion that people hold negative stereotypes about female athletes. For instance, an image of women as possessing poor athletic ability readily comes to mind merely by thinking about universal expressions like, for example, “Stop playing like a girl!”. By and large, people seem to associate athletic ability with manliness rather than womanliness (Fredricks & Eccles, 2005; Koivula, 1999) and a staunch belief that men outperform women at a myriad of sports (especially at sports considered to be predominantly male-dominated) continues to exist.

### **The Causal Mechanism of Stereotype Threat in Sport**

Whilst the presence of negative gender stereotypes regarding athleticism carries on to impair the sports performance of women and may even thwart them from at any point performing to their genuine capacity in athletic settings (Chalabaev et al., 2013), extremely few research to date has investigated how stereotype threat might bring to bear its unpleasant effect in sport (Beilock & McConnell, 2004). Although more research is necessary, our understanding of the latter may be considerably enlightened by work on explicit monitoring (often referred to as “choking under pressure”). The explicit monitoring hypothesis (Beilock & Carr, 2001; Schmader, Johns, & Forbes, 2008) argues that motor skills deteriorate under pressure because such settings trigger people to expand the amount of scrutiny they dedicate to monitoring step-by-step performance in order to assure a favourable result (Lewis & Linder, 1997; Beilock & Carr, 2001; Baumeister, 1984). Regrettably, this expanded scrutiny can boomerang and throw into disarray what should have been a smooth execution of motor skills. Support for the above mentioned route by which stereotype threat might derogate athletic performance comes from current work in, among others, baseball batting (Gray,

2004) and golf putting (Perkins-Ceccato, Passmore, & Lee, 2003; Beilock, Bertenthal, McCoy, & Carr, 2004), indicating that motor skills are disadvantaged by excessively much attention to execution. That is, stereotype threat may be a kind of “choking under pressure”, pressure as a consequence of the anxiety of affirming a negative stereotype concerning how one ought to perform.

Strikingly, women do not need to agree with gender stereotypes in order for their performance to be negatively affected (Chalabaev et al., 2008). The fear of a poor performance might be extant nonetheless, even if they disagree or consider the gender stereotype not to personally pertain to them (Steele, 1997). Hence, adverse comparisons to men remain serving as a prominent burden (Stone & McWhinnie, 2006) and the reminder and sheer existence of gender stereotypes of lower female athleticism might go on to derogate women’s athletic performance.

### **Empirical Evidence (and Meta-Analyses) on Stereotype Threat and Gender**

According to the previously mentioned Multi-Threat Framework, discoveries from preceding research on stereotype threat (i.e., focusing on cognition-based tasks) might be relevant chiefly to that particular form of stereotype threat, but might not generalize beyond that form or to distinctive contexts that might trigger a different form of stereotype threat (Shapiro & Neuberg, 2007). The latter sparked Gentile, Boca, and Giammusso (2018) to provide an overview of the stereotype threat phenomenon and its potentially negative effect on women’s performance in sport. They performed a meta-analysis on 24 effects in order to grant an assessment of the effect of stereotype threat on women’s sport performance. This meta-analysis reveals that stereotype threat unfavourably affects the physical activity of women and that this effect is significantly more prominent for female athletes participating in a relatively male-dominated sport compared to female athletes participating in a relatively

female-dominated sport characterized by aesthetics, grace, and expressivity (Hardin & Greer, 2009; Chalabaev et al., 2013). With effect sizes ranging from  $d = 0.24$  (Picho et al., 2013) to  $d = 0.48$  (Walton & Spencer, 2009), preceding meta-analyses targeting the gender stereotype threat phenomenon (i.e., limited to cognitively-based tasks) in women (Doyle & Voyer, 2016; Nguyen & Ryan, 2008; Picho et al., 2013; Stoet & Geary, 2012; Walton & Cohen, 2003; Walton & Spencer, 2009) found results comparable to those found by Gentile et al. (2018) ( $d = 0.33$ ).

However, in contrast to work on stereotype threat limited to cognitively-based tasks, research literature that focuses on the physical activity domain does not seem to display a systematic bias to exclusively embrace positive results (i.e., publication bias). In order to determine whether publication bias existed in their sample of selected studies, Gentile et al. (2018) carried out a number of tests on the funnel plot to estimate funnel plot asymmetry (i.e., Begg and Mazumdar's rank correlation test (Begg & Mazumdar, 1994), trim and fill (Duval & Tweedie, 2000), and Egger's test (Sterne & Egger, 2005)). When publication bias is absent, studies with high exactitude will be plotted not far off the average and studies with low exactitude will not be plotted near the average, but will diffuse evenly on both sides of the average. As a result, an approximately funnel-shaped distribution will be created. Any deviation from this funnel-shaped distribution can point to publication bias. The tests on the funnel plot carried out by Gentile et al. (2018) did not display the prototypical asymmetry normally present when publication bias exists. That is, research literature that focuses on the physical activity domain does not seem to display a methodical inclination to exclusively report significant results.

The results of the latter meta-analysis corroborate results from an array of like-minded studies that re-evaluated the prevailing stereotype threat literature and scrutinized how this effect could account for performance decrements in sports (Beilock & McConnell, 2004),

tested whether performance of females decreased when gender stereotypes of athleticism were triggered (Hermann & Rumrich, 2018), and investigated various mediators hypothesized in the stereotype threat literature concentrating on athletics (Hermann & Vollmeyer, 2016). Even though virtually every previously mentioned study concerning stereotype threat and sports examined slightly different manipulations and control conditions, populations from contrasting countries, possible causal mechanisms, and moderators and mediators, stereotype threat is, precisely as it is in academics, viable in sport (Beilock & McConnell, 2004).

According to Shapiro and Neuberg (2007), the latter inference is a reasonably logical one. Foundational to their Multi-Threat Framework is the awareness that the phenomenon will manifest itself dissimilarly under various domains and groups of which sports and athletes is one of many. Altogether, stereotype threat appears to be a robust phenomenon that could manifest itself across a wide range of domains and social groups (Beilock & McConnell, 2004). By virtue of stereotype threat having far-reaching consequences for sports (e.g., undermining the athletic performance of women), one of those are, without a doubt, female athletes participating in a sport predominantly male-dominated.

### **Divergent Strategies to Cope With Stereotype Threat**

As previously noted, the jeopardy of being looked at in terms of those negative stereotypes may prompt a troublesome psychological state and self-esteem may be affected. Generally, people have a basic need to uphold a positive self-evaluation (Steele, 1988; Tesser, 1988). That is to say that, when individuals face such a negative stereotype, they will in all probability attempt to abate the troublesome psychological state. Diverse strategies to deal with such uneasiness have been put forward, such as identification and disidentification.

According to the rejection-identification model (Branscombe, Schmitt, & Harvey, 1999), the recognition of being a member of a stereotyped social group might lead to a more powerful identification with this particular group. When faced with a negative stereotype, a social identity can accommodate people with a “cushion” (i.e., psychological resources) to handle drawbacks (Haslam & Reicher, 2006; Ashmore, Deaux, & McLaughlin-Volpe, 2004). Moreover, identifying with the stereotyped social group can cause individuals to become aware of resemblances with others who share an identical psychological pain (Bastian, Jetten, & Ferris, 2014). The feeling of fellowship makes people feel bolstered and appreciated, and thus is a way of abating the troublesome psychological state caused by a negative stereotype and protecting the individuals’ well-being (Branscombe et al., 1999).

Alternatively, the rejection-disidentification model (Jasinskaja-Lahti, Liebkind & Solheim, 2009) posits that, to cope with negative stereotype threat, members of a stereotyped group might want to dissociate themselves from that particular group. If individuals anticipate more fulfilment of acceptance needs in a different group, they may lessen their dedication to the stereotyped group, choosing for discontinuation or avoidant reactions (Smart Richman & Leary, 2009). Notwithstanding, while the strategy of identification ordinarily displays a protective role in the well-being of individuals, the strategy of disidentification usually causes individuals to experience unfavourable attitudes and emotions towards the stereotyped group (Jasinskaja-Lahti et al., 2009).

An additional and less defensive means of abating the troublesome psychological state and asserting a positive self-evaluation is self-affirmation (Steele, 1988; Sherman & Cohen, 2006). Self-affirmation implies that people uplift the self’s resources by displacing their attention to a valued domain of the self that is not threatened and thereby render threats to the self psychologically less appalling (Steele, 1988; Sherman & Cohen, 2006). The latter accommodates individuals with a means of inserting the threat within a larger context that

accentuates a feeling of being deserving and capable (Sherman et al., 2014; Sherman & Cohen, 2006). Individuals will in all likelihood go through less sorrow and, importantly, react less defensively when challenged by a negative stereotype. Previous research has shown that, just as it is effective in domains as diverse as improving performance in academic settings (Protzko & Aronson, 2016) and strengthening a feeling of fitting in (Layous et al., 2017), self-affirmation is effective in fighting down negative threats to identity (e.g., national identity, gender) (Cook, Purdie-Vaughns, Garcia, & Cohen, 2012; Sherman et al., 2014; Layous et al., 2017). Moreover, the process of self-affirmation seems to be less defensive to a great extent compared to the rather defensive strategies such as identification and disidentification. Of note, however, is that previous research has shown that self-affirmation must take place in a domain unrelated to the one that is threatened (Stone & Cooper, 2003; Blanton et al., 1997). When self-affirmation takes place in the domain that is threatened, defensive reactions tend to escalate. By and large, self-affirmation may be employed by people as some sort of “buffer” to cope with stereotype threat (Liu & Steele, 1986; Steele, Spencer, & Lynch, 1993).

### **Moderators: Global Self-Esteem and Group Identification Strength**

Global self-esteem and group identification strength (importance of the group in the concept of the self) have been proposed as moderators of frequent occurrence in previously conducted research that focused on social identity threat (Schmader, Block, & Lickel, 2015; MCQueen & Klein, 2006). These factors presumably play a role in the effectiveness of particular self-affirmation mechanisms (Sherman & Cohen, 2002).

A variety of researchers have demonstrated that, among individuals with low global self-esteem juxtaposed to those with high global self-esteem, self-affirmations in a domain unrelated to the one that is threatened bring about greater perceived stress (Creswell et al., 2005) and attitude change (Stone & Cooper, 2003; Steele et al., 1993). Also, individuals with

low global self-esteem may worry less about threats to their social identity because this feedback suits with their already negative self-view. Still and all, other studies were unsuccessful in corroborating the latter results and suggest that self-affirmation in a domain unrelated to the one that is threatened affords identical effects for individuals high and low on global self-esteem (Spencer, Fein, & Lomore, 2001). Despite the specific causal mechanisms through which global self-esteem affects self-affirmation remaining unknown, low and high levels of global self-esteem should in all probability be distinguished to improve our understanding of the usefulness of self-affirmation through the recall of mastery autobiographical memories.

Moreover, previous research has shown that different classes of responses to categorization threat are likely to be exhibited by people who vary in the degree to which they feel committed to a particular social group (e.g., national identity, gender) (Schmader et al., 2015; Armenta, 2010; Spears, Doosje, & Ellemers, 1997; McCoy & Major, 2003). Recall that categorization threat is a form of social identity threat where, in inter-group scenarios, a stereotypical group characteristic is ascribed to a group of people (Branscombe, Ellemers, Spears, & Doosje, 1999). According to Branscombe et al. (1999), internal and external categorizations will in all probability be contradictory or incompatible when such a stereotypical group characteristic is explicitly imposed on an individual who identifies weakly with that particular group. As a consequence, individuals who identify weakly with their group tend to show resistance to such threats by, for example, stressing ingroup heterogeneity, stressing unique personal qualities, and further disidentification (Branscombe et al., 1999). The latter automatically implies that individuals who identify strongly with their group tend to be more receptive to such threats.

Gender differences between male and female athletes have a lengthy history of discourse and carry a considerable burden for female athletes (especially for female athletes

participating in a relatively male-dominated sport (Gentile et al., 2018)). Moreover, the suggestion that women possess poor athletic ability is broadly embraced as a negative stereotype about female athletes (Knight & Giuliano, 2001; Biernat & Vescio, 2002). Accordingly, a gender stereotype of lower female athleticism would presumably be identity-relevant. Interestingly, the latter stereotype is likely to be relevant to two social categories, namely women and female athletes (who of course are women). As mentioned above, different classes of responses to the gender stereotype are likely to be exhibited by female athletes who vary in how committed they are as women (i.e., gender identity strength) and as (female) athletes (athletic identity strength).

### **Specific Self-Affirmation: Autobiographical Memories**

Previous research on self-affirmation has mainly concentrated on value affirmation, which admits one to prolong or reinstate one's overall positive self-image and a sense of global self-worth by giving consideration to important personal values like relationships, religion, norms, and skills (i.e., valued aspects of the self about which one feels competent and that are not interconnected with the threatened domain) (Cook et al., 2012; Cohen & Sherman, 2014; Steele, 1988; Sherman & Cohen, 2006). On the other hand, Tavitian-Elmadjian et al. (2020) demonstrated that, just as it can be essential for a wide range of other domains (e.g., resolving difficulties (Cohen, 1998), conveying empathy (Bluck, Alea, Habermas, & Rubin, 2005), and conserving social relationships (Alea & Bluck, 2003; Nelson, 1993; Neisser, 1988)), recalling autobiographical memories can idem be essential for self-affirmation.

Autobiographical memories can best be thought of as the branch of human memory associated with the self (Brewer, 1986) and a wide variety of types of autobiographical memories exist (Rubin, 1988). An example of such a type is what Brewer (1986) refers to as



a personal memory, which is a remembrance of a specific experience from the yore of an individual (i.e., a single event). Naturally, the content of this experience may vary widely. One might, for example, call to mind the first time he or she fell in love, a terrible black-out during an important exam, or the one time when his or her team won the most important game of the season.

The type of personal memory Tavitian-Elmadjian et al. (2020) focused on in their study was a mastery memory, which is a recollection of an instance of success and getting grip on a challenge. Their research revealed that, much like thinking about important personal values (Sherman & Cohen, 2006), the recall of mastery autobiographical memories could be of service to establish a sense of worth and capability when faced with a negative stereotype threat. Notwithstanding, Tavitian-Elmadjian et al.'s (2020) experiments displayed that the recall of mastery autobiographical memories was only advantageous in tackling a negative stereotype threat when the latter was identity-relevant. That is, the threat ought to aim at a valued element of the self (e.g., national identity). When the threat aimed at an unvalued element of the self (i.e., identity-irrelevant), the recall of mastery autobiographical memories was not of assistance in tackling the latter.

Tavitian-Elmadjian et al.'s (2020) study was the first to display that self-affirmation by summoning back a past success can buffer against negative stereotype threat and small steps have already been taken towards attempting to conceptually replicate their results within a different context (Tavitian, Bender, Van de Vijver, Chasiotis, & Vosgerichian, 2019). However, to establish that the detected phenomenon is vigorous, more conceptual replication efforts or extensions with other participants, instruments, and experimental manipulations are necessary. I, therefore, set out to assess whether self-affirmation through the recall of mastery autobiographical memories can also help female athletes participating in

a predominantly male-dominated sport (ice hockey) cope with gender stereotype threat (RQ1).

### **The Present Study**

For the present study, I looked into mastery recall of autobiographical memories as a specific machinery of self-affirmation in a frame of reference of categorization threat based on gender. Categorization threat based on gender is a form of social identity threat where, in inter-group scenarios, a stereotypical group characteristic based on gender is ascribed to a group of people (Branscombe et al., 1999). For my study, a gender stereotype of lower female athleticism was ascribed to female athletes participating in a predominantly male-dominated sport (ice hockey). I anticipate participants subjected to a gender stereotype of lower female athleticism, who as well recall mastery autobiographical memories, to possess higher personal self-esteem compared to participants who recall routine autobiographical memories (H1).

The gender stereotype of lower female athleticism was ascribed to the female athletes by means of a vignette formulated by me in which I refer to all female athletes as having inferior athletic ability (i.e., categorization threat) compared to male athletes and back up the assertion with false statistics (see Appendix A for full script). I employed a threat to the athletic ability as “poor athletic ability” is broadly embraced as a negative stereotype about female athletes (Knight & Giuliano, 2001; Biernat & Vescio, 2002). If the threat manipulation was regarded by participants in the purposed manner, I anticipate ratings of perceived vignette content and author attitudes to be predominantly negative versus neutral or positive (H2).

In previously conducted research that focused on social identity threat, global self-esteem and group identification strength have been proposed as moderators of frequent

occurrence (Schmader et al., 2015; MCQueen & Klein, 2006). In line with other corresponding studies (Tavitian-Elmadjian et al., 2020; Tavitian et al., 2019), it seemed to me necessary to equally take into account these frequently occurring moderators in my study. I, therefore, set out to assess whether global self-esteem, gender identity strength, and athletic identity strength moderate the relationship between the type of memory recall (i.e., mastery or routine autobiographical memories) and personal self-esteem (RQ2).

Firstly, I anticipate participants subjected to a gender stereotype of lower female athleticism, who as well recall mastery autobiographical memories and are high on global self-esteem, to score higher on personal self-esteem compared to participants who recall routine autobiographical memories (H3). Also, I do not anticipate personal self-esteem to differ across recall types for participants who are low on global self-esteem (H4).

Secondly, I anticipate participants subjected to a gender stereotype of lower female athleticism, who as well recall mastery autobiographical memories and are high on athletic identity strength, to score higher on personal self-esteem compared to participants who recall routine autobiographical memories (H5). Also, I do not anticipate personal self-esteem to differ across recall types for participants who are low on athletic identity strength (H6).

Finally, I anticipate participants subjected to a gender stereotype of lower female athleticism, who as well recall mastery autobiographical memories and are high on gender identity strength, to score higher on personal self-esteem compared to participants who recall routine autobiographical memories (H7). Also, I do not anticipate personal self-esteem to differ across recall types for participants who are low on gender identity strength (H8).

## Method

### Participants

A between-participants factorial analysis of covariance (ANCOVA) was used to investigate whether differences exist between the two groups (mastery versus non-mastery) on the dependent variable (personal self-esteem), after controlling for covariates (athletic and gender identity strength). In an a priori power analysis, I computed the minimum required sample size  $N$  ( $N = 210$ ) using G\*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). I calculated the latter as a function of the effect size  $f$  ( $f = .25$ ), the pre-specified significance level  $\alpha$  ( $\alpha = .05$ ), and the requisite power level ( $1 - \beta = .95$ ). Participants (i.e., female ice hockey players of various levels, from beginners to professionals) were recruited via snowball sampling through social media platforms (Facebook, Instagram, and WhatsApp) and email (see Appendix B for advertisement title and text). To account for non-eligible participants and memories, and to oversample for mastery memories, I targeted 384 participants. I retained a total of 193 participants after excluding those who did not practice the sport of ice hockey ( $N = 14$ ), were not female ( $N = 8$ ), completed the Dutch variant of the survey<sup>1</sup> ( $N = 32$ ), and whose autobiographical memory entries were either absent ( $N = 20$ ), unclear ( $N = 14$ ), or non-eligible ( $N = 71$ ). Sociodemographic attributes of the sample are presented in Tables 1 and 2.

---

<sup>1</sup> I assembled the present study on Qualtrics in two different languages, namely English and Dutch, so that not merely Dutch female ice hockey players, but also players from additional countries could participate in the study. However, I was unable to find validated Dutch translations of the scales used, which signified that I had to translate the latter into Dutch myself. As a result, it was not unlikely that the scales were no longer valid, which made it difficult to easily merge Dutch and non-Dutch participants. That is why I later decided to exclude participants who completed the Dutch variant of the survey.

Table 1

*Cultural Background by Country Living and Country Born (N = 226)*

Characteristic		Living				Total
		Netherlands	U.S.A	Canada	Other	
Born	Netherlands	30	0	0	0	30
	U.S.A.	0	132	0	1	133
	Canada	0	4	24	0	28
	Other	2	6	1	26	35
Total		32	142	25	27	226

Table 2

*Sociodemographic Attribute of the Sample*

Attribute	N	M (SD)	Minimum	Maximum
Age	226	34.05 (12.49)	12	66

*Note.* Participants' age was assessed in years.

## Procedure

Followed by obtaining approval from the Ethics Review Board (reference EC-2019.EX148t), I assembled the present study on Qualtrics. To not directly disclose the experimental design to participants, I inquired them that the study was “understanding different attitudes among female athletes of many different ages and countries”. After reading the consent form, participants had to indicate whether they agreed to participate in the study. Subsequently, I asked participants two questions to exclude potential non-eligible participants. First, I asked whether they indeed practiced the sport of ice hockey. Second, I asked whether they were indeed female. I asked these questions to make sure that the gender stereotype threat of lower female athleticism is indeed pertinent to a valued feature of the self and that the sample would be as homogeneous as possible, since homogeneity is likely to contribute to better interpretability of results. Thereafter, I administered measures of global self-esteem, gender identity strength, and athletic identity strength in randomized order.

Of note is that global self-esteem can be considered here as the baseline and personal self-esteem (the dependent variable) as the outcome and, therefore, I had a pre-post design. If I used the same scale to measure both global self-esteem and personal self-esteem, potential carry-over effects could occur. Since I was not interested in the nuances of self-esteem, I used a validated one-item version of the Rosenberg Self-Esteem (RSE) scale (Rosenberg, 1965) to evaluate global self-esteem in 50% of the participants. For the other half of the participants, global self-esteem was not evaluated. However, the dependent variable (i.e., personal self-esteem) was evaluated using the complete RSE. In this way, effects could be checked for any carry-over.

Following the administration of the above mentioned measures, I subjected participants to a gender stereotype of lower female athleticism (threat manipulation). I presented a vignette featuring a blatant threat (I did not particularly target high identifiers and, in such occasions, it may be required to heighten the salience of the threat for it to be operative (see Keller, 2002; Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003)) in which I refer to all female athletes as having inferior athletic ability (i.e., categorization threat) compared to male athletes and back up the assertion with false statistics. Since the dependent variable was personal self-esteem and not a measure of performance, I specifically formulated the vignette for the present study as opposed to modifying one from another study. For my study, I employed what we generally know about categorization threat and adjusted the content so that it matches threats faced by female athletes. On the other hand, alternative studies that likewise focused on the effect of stereotype threat on women's performance in sport tended to consider performance on an athletic task as the dependent variable where threat was typically brought about by manipulating task instructions (Hively & El-Alayli, 2014; Gentile et al., 2018; Martiny et al., 2015; Beilock & McConnell, 2004).

After reading the vignette, participants were asked to briefly summarize the latter, indicate whether the content of it is positive or negative, and rate the attitudes of the author of it to the group portrayed, which served as a manipulation check (to determine whether the vignette was regarded by participants in the purposed manner) (adapted from Phinney, Chavira, & Tate, 1993). Both the perceived vignette content and author attitudes were rated on a 5-point Likert-type scale ranging from 1 “extremely positive” to 5 “extremely negative”. Next, I randomly assigned participants to one of the two experimental conditions with recall type as the independent variable. Subsequently, as the dependent variable, I assessed participants’ overall subjective evaluation of their own worth (i.e., personal self-esteem). Afterwards, I asked participants to indicate whether they could recall the gender of the author of the presented vignette. I asked this question to find out if differences in the perceived source of the threat (i.e., whether the threat was perceived as coming from a man or a woman) might induce different responses to it. Then, participants were asked to supply some demographic information, namely how old they are in years, the country in which they were born, the country in which their mother was born, the country in which their father was born, and the country in which they currently live. Next to that, I requested participants to indicate what they thought the study was about. I asked this question to find out whether the experimental design was perhaps too obvious, which might have prevented participants from objectively partaking in the study. Finally, participants were debriefed about the aims of the study in which they just took part.

## **Measures**

To evaluate global self-esteem in 50% of the participants, I used the Single-Item Self-Esteem (SISE) scale (Robins, Hendin, & Trzesniewski, 2001) encompassing 1 item (i.e., “I have high self-esteem”) rated on a 7-point Likert-type scale ranging from 1 “very true of me” to 7 “not very true of me”.

I measured gender identity strength by asking participants to answer 4 items (e.g., “Being a woman is an important part of myself”) adapted from the identification measure used by Doosje, Branscombe, Spears and Manstead (1998) concerning their identification as women on a 7-point Likert-type scale ranging from 1 “strongly agree” to 7 “strongly disagree” ( $\alpha = .70$ ).

Athletic identity strength was measured by administering the Athletic Identity Measurement Scale (AIMS) (Brewer, Van Raalte, & Linder, 1993) encompassing 10 items (e.g., “Sport is the most important part of my life”) rated on a 7-point Likert-type scale ranging from 1 “strongly agree” to 7 “strongly disagree” ( $\alpha = .93$ ) to participants.

For their study, Tavitian-Elmadjian et al. (2020) settled on assessing collective self-esteem as a measure of collective self-worth in order to enlarge the fit with the threatened domain of identity (i.e., the group). In the end, they were found not to be entirely satisfied with this choice as they propose future studies should assess the effect of mastery recall on general (personal) self-esteem. According to them, assessing the effect of mastery recall on general (personal) self-esteem would possibly match the personal content of the self-affirmation task better and, consequently, yield a more sizeable effect. Based on the above, I settled on assessing the effect of mastery recall on personal self-esteem. To estimate the latter, I administered the RSE encompassing 10 items (e.g., “I feel that I am a person of worth”) rated on a 4-point Likert-type scale ranging from 1 “strongly agree” to 4 “strongly disagree” ( $\alpha = .90$ ) to the full number of participants.

Note that because of the way all the latter scales are arranged, lower (mean) scores equate to higher (mean) global self-esteem, gender identity strength, athletic identity strength, and personal self-esteem (i.e., the scales were not coded the reverse to indicate high scores



equals high construct levels). This is of great importance to keep in mind when interpreting the results.

### **Pre-Testing the Experimental Vignette**

To determine whether the vignette would be perceived by participants in the purposed manner, I collected some qualitative pre-test data prior to collecting the actual data. For this I personally approached a number of female ice hockey players ( $N = 5$ ) and had the vignette read by them. After reading the vignette, they were asked to evaluate the content of the latter and the attitudes of the author of it to the group portrayed. The content of the vignette was evaluated in a similar way by all female ice hockey players, namely extremely negative. It made them feel inferior to men, threatened, and demotivated to participate in sports. In addition, they perceived the content to be very stereotypical. The attitudes of the author of the vignette to the group portrayed were evaluated as negatively as the content of it. The female ice hockey players indicated that the author has an enormously negative view of the female and a superior view of the male athlete. They also suspected the author to be a misogynist. Based on the above, I concluded that the vignette would most likely be perceived by participants in the purposed manner. To substantiate this conclusion, some statements of the female ice hockey players can be seen in Appendix C.

### **Autobiographical Memories: Recall Type Manipulation and Coding**

To appraise the role of recalling mastery autobiographical memories as a specific machinery of self-affirmation in a frame of reference of categorization threat based on gender, I asked participants to take a few minutes to call to mind and write as detailed as possible about one of two types of autobiographical memories. I either primed participants to remember and write about an occasion in their life when they came up against a challenge and how they got grip on the latter (mastery recall, affirmation condition) or their daily

morning routine (routine recall, non-affirmation condition) (see Appendix D for full scripts). According to Cohen, Aronson, and Steele (2000), participants may employ, in order to self-affirm, any self-contemplative writing opportunity. That is why, in the non-affirmation condition, I primed participants to call to mind memories of their daily morning routine, which is oftentimes neutral and therefore generally does not contribute to potential self-affirmation. By priming participants to remember and write about one of the above mentioned two types of autobiographical memories, I evoked mastery memories (e.g., *“In Middle School I was a bit overweight and for that reason I was often bullied. Therefore, I often felt alone and sad. To overcome these negative feelings my grandmother told me every day that I was beautiful just the way I was. Also, that I should try to ignore the bullying. Once I started believing what my grandmother said I felt a lot better”*) and routine morning memories (e.g., *“I wake up. Then I turn off my alarm clock, get out of bed, and put on my clothes. After that, I kiss my girlfriend goodbye. Next thing I do is eat a sandwich and feed my cat. Then I brush my teeth. I depart from home once I’m ready and drive to work. Then I get to work and drink a cup of coffee. After that, I start working”*). As hypothesized by self-affirmation theory (Aronson et al., 1999), neither type of autobiographical memory should be related to the domain that is threatened. Furthermore, routine autobiographical memories, in contrast to mastery autobiographical memories, would not qualify as an occasion in one’s life that could be of service to establish a sense of worth and capability when faced with a negative stereotype threat. That is, routine autobiographical memories would not meet the requirements to serve as a means of identity affirmation.

To be able to assess whether self-affirmation through the recall of mastery autobiographical memories could serve as a means of identity affirmation, it is naturally greatly key that participants who were primed to recall a mastery autobiographical memory actually did so. Consistent with the study by Tavitian-Elmadjian et al. (2020), I employed the

coding manual for self-defining events by Thorne and McLean (2001) to exclude non-eligible autobiographical memory entries. All entries that gave prominence to the chronicles of one's individual or family/group's vigorous attempts at accomplishment and were not interrelated with the content of the threat were coded as mastery memories (Thorne & McLean, 2001). On the other hand, all entries that underscored topics of life-threatening events, recreational experiences, guilt/shame, and relationships were coded as non-mastery memories. Likewise, routine recalls were regarded as non-eligible if they did not give prominence to a routine and were not neutral.

### **Mastery Memories: Coping vs. Preventing Agents Against Categorization Threat Based on Gender**

A factor that may influence the process of self-affirmation is whether the self-affirmation manipulation is presented before (preventive mechanism) or after (coping mechanism) the threat manipulation. Several studies have shown that the recall of mastery autobiographical memories could serve as the mechanism through which self-worth is maintained after the threat manipulation (Aronson, Cohen, & Nail, 1999; Steele, 1988; Sherman & Cohen, 2006). However, both mechanisms are plausible founded on the self-affirmation literature (Steele, 1988). That is, the recall of mastery autobiographical memories could also serve as the mechanism through which self-worth is maintained before the threat manipulation (Sherman, Nelson, & Steele, 2000; Tavitian-Elmadjian et al., 2020; MCQueen & Klein, 2006). Based on intuition, I tested self-affirmation as a coping mechanism.

## **Results**

### **Manipulation Check**

A one-sample t-test was conducted to assess whether the threat manipulation was regarded by participants in the purposed manner (i.e., predominantly negative versus neutral

or positive (H2)). Both the perceived vignette content ( $M = 4.50$ ,  $SD = .72$ ) and author attitudes ( $M = 4.41$ ,  $SD = .78$ ) were rated as significantly more negative compared to neutral or positive<sup>2</sup>;  $t(192) = 28.89$ ,  $MD = 1.50$ ,  $p = .00$ ,  $d = 2.08$  and  $t(192) = 25.19$ ,  $MD = 1.42$ ,  $p = .00$ ,  $d = 1.81$ . That is to say that, ratings of perceived vignette content and author attitudes were in the anticipated direction and the threat manipulation was in all likelihood regarded by participants in the purposed manner.

To explore the distributions of the ratings of perceived vignette content and author attitudes a bit further, I conducted an exploratory analysis examining the latter variables using boxplots. The boxplots for both variables yielded three outliers that were identical to each other. These outliers belonged to participants who experienced the perceived vignette content and author attitudes as overly positive. Because a few outliers can sometimes make a disproportionate difference, I again conducted a one-sample t-test to explore how and if the results would change when they were excluded from the analysis.

Results indicated that both the perceived vignette content ( $M = 4.54$ ,  $SD = .66$ ) and author attitudes ( $M = 4.45$ ,  $SD = .72$ ) were rated as significantly more negative compared to neutral or positive;  $t(189) = 32.44$ ,  $MD = 1.54$ ,  $p = .00$ ,  $d = 2.35$  and  $t(189) = 27.66$ ,  $MD = 1.45$ ,  $p = .00$ ,  $d = 2.01$ . These results are not extraordinarily surprising given that the perceived vignette content and author attitudes were already seen as significantly more negative compared to neutral or positive when the overly positive outliers were included in the analysis. The already negative compared to neutral or positive mean scores became even more negative in the absence of these overly positive data entries and larger effect sizes were found ( $d = 2.35$  and  $d = 2.01$  instead of  $d = 2.08$  and  $d = 1.81$ ). That is to say that, ratings of

---

<sup>2</sup> Both the perceived vignette content and author attitudes were rated on a 5-point Likert-type scale ranging from 1 "extremely positive" to 5 "extremely negative". That is, the higher the mean score, the more negative the perceived vignette content and author attitudes.

perceived vignette content and author attitudes were in the anticipated direction and the threat manipulation was in all likelihood regarded by participants in the purposed manner when outliers were included and even more so when excluded. A description and overview of the sample, as well as the zero-order correlations per condition can be seen in Tables 3 and 4.

Table 3

*GSE, AI, GI, and PSE Total Score Descriptives per Experimental Condition*

Type of memory		<i>N</i>	<i>M</i>	<i>SD</i>	Minimum	Maximum
Mastery ( <i>N</i> = 71)	GSE	34	2.53	1.11	1	5
	AI	71	32.87	10.27	14.00	60.00
	GI	71	7.04	2.48	4.00	14.00
	PSE	71	18.87	4.91	10.00	32.00
Daily morning routine ( <i>N</i> = 122)	GSE	58	2.53	1.19	1	7
	AI	122	32.66	9.22	12.00	57.00
	GI	122	7.19	3.49	4.00	21.00
	PSE	122	19.14	4.31	10.00	29.00

*Note.* GSE = global self-esteem; AI = athletic identity; GI = gender identity; PSE = personal self-esteem.

Table 4

*Zero-Order Correlations per Experimental Condition*

Type of memory			GSE	GI	AI	PSE
Mastery	GSE	<i>r</i>	1	.01	.02	.64**
		<i>p</i>		.94	.92	.00
	GI	<i>r</i>	.01	1	.24*	.07
		<i>p</i>	.94		.045	.55
	AI	<i>r</i>	.02	.24*	1	-.11
		<i>p</i>	.92	.045		.36
	PSE	<i>r</i>	.64**	.07	-.11	1
		<i>p</i>	.00	.55	.36	
Daily morning routine	GSE	<i>r</i>	1	.09	.08	.71**
		<i>p</i>		.50	.53	.00
	GI	<i>r</i>	.09	1	.05	.05
		<i>p</i>	.50		.62	.57
	AI	<i>r</i>	.84	.05	1	-.11
		<i>p</i>	.53	.62		.22
	PSE	<i>r</i>	.71**	.05	-.11	1
		<i>p</i>	.00	.57	.22	

*Note.* GSE = global self-esteem; AI = athletic identity; GI = gender identity; PSE = personal self-esteem.

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

**Hypothesis Test: Mastery Memories as a Specific Machinery of Self-Affirmation**

To assess whether participants subjected to a gender stereotype of lower female athleticism, who as well recall a mastery autobiographical memory as opposed to a routine morning autobiographical memory, would have higher personal self-esteem (H1), I conducted a between-participants factorial ANCOVA with personal self-esteem as the dependent variable and athletic and gender identity strength as covariates. Global self-esteem was not included in this analysis as a covariate because it was evaluated in only 50% of the participants. Including the latter as a covariate would reduce the total sample size and, as a result, reduce the statistical power. Global self-esteem was idem not included in any further between-participants factorial analyses of covariance. It was, however, later assessed as a moderator of the effect of social identity threat on self-affirmation.

Athletic identity strength ( $b = -1.01$ ,  $t(189) = -1.92$ ,  $SE_b = .53$ ,  $p = .06$ ,  $\eta_p^2 = .02$ ) and gender identity strength ( $b = .54$ ,  $t(189) = 1.07$ ,  $SE_b = .51$ ,  $p = .29$ ,  $\eta_p^2 = .01$ ) emerged as statistically insignificant covariates.

Controlled for athletic and gender identity strength, the effect for recall type (i.e., mastery memory versus routine morning memory) was not statistically significant ( $F(1, 189) = .15$ ,  $MSE = 20.35$ ,  $p = .70$ ,  $\eta_p^2 = .001$ ). On average, participants subjected to a gender stereotype of lower female athleticism, who as well recalled a mastery autobiographical memory as opposed to a routine morning autobiographical memory, did not have significantly higher personal self-esteem (see Table 5). That is, the data do not support hypothesis 1.

Table 5

*Mean PSE per Recall Type*

Type of memory	<i>M</i>	<i>SD</i>	<i>N</i>
Mastery	18.87	4.91	71
Daily morning routine	19.14	4.31	122
Total	19.04	4.53	193

*Note.* PSE = personal self-esteem.

***Explorations of the effect for recall type.***

As mentioned earlier, outliers in the dataset may vigorously affect the results. They can increase the variability in the data and, subsequently, reduce the statistical power. Additionally, measures of central tendency that try to describe the center point or middle of a distribution may be strongly influenced by them. When outliers are present in the dataset, the mean can easily be dragged up or down. To investigate whether outliers in the dataset are responsible for the fact that the data do not support hypothesis 1, I explored the distributions of personal self-esteem, athletic identity, and gender identity a bit further. I conducted an

exploratory analysis examining the latter variables using boxplots. While the boxplots for both personal self-esteem and athletic identity yielded one outlier, the boxplot for gender identity yielded five. To explore how and if the results would change when these seven outliers were excluded from the analysis, I again conducted a between-participants factorial ANCOVA with personal self-esteem as the dependent variable and athletic and gender identity strength as covariates.

While gender identity strength emerged as a statistically insignificant covariate ( $b = .62$ ,  $t(182) = 1.22$ ,  $SE_b = .51$ ,  $p = .23$ ,  $\eta_p^2 = .01$ ), athletic identity strength emerged as a statistically significant one;  $b = -1.19$ ,  $t(182) = -2.20$ ,  $SE_b = .54$ ,  $p = .03$ ,  $\eta_p^2 = .03$ .

Controlled for athletic and gender identity strength, the effect for recall type was not statistically significant ( $F(1, 182) = .25$ ,  $MSE = 20.22$ ,  $p = .62$ ,  $\eta_p^2 = .001$ ). On average, participants subjected to a gender stereotype of lower female athleticism, who as well recalled a mastery autobiographical memory as opposed to a routine morning autobiographical memory, did not have significantly higher personal self-esteem (see Table 6). That is, the data do not support hypothesis 1.

Table 6

*Mean PSE per Recall Type for the Analysis Without Outliers*

Type of memory	<i>M</i>	<i>SD</i>	<i>N</i>
Mastery	18.69	4.91	68
Daily morning routine	19.02	4.32	118
Total	18.90	4.53	186

*Note.* PSE = personal self-esteem.

Results indicated that participants subjected to a gender stereotype of lower female athleticism, who as well recalled a mastery autobiographical memory as opposed to a routine morning autobiographical memory, did not have significantly higher personal self-esteem,



both when outliers were included and excluded. Apart from the fact that athletic identity strength emerged as a statistically significant covariate, the results barely changed when the seven outliers were excluded from the analysis (the effect sizes were even of exactly the same order of magnitude, namely  $\eta_p^2 = .001$ ). That is to say that, the seven outliers that were yielded by the boxplots for personal self-esteem and athletic and gender identity are most likely not responsible for not finding a statistically significant effect for recall type. Based on the latter I decided not to exclude them from further analyses.

Another factor that could be responsible for the fact that the data do not support hypothesis 1 is lower power. Power is lower in view of the sample size ( $N = 193$ ). However, if all participants who completed the Dutch variant of the survey ( $N = 32$ ) were included in the analysis, I would have enough power for the experimental design as a whole, as the calculations of G\*Power 3.1 suggest (Faul et al., 2007). To explore how and if the results would change when the latter participants were included in the analysis, I again conducted a between-participants factorial ANCOVA with personal self-esteem as the dependent variable and athletic and gender identity strength as covariates.

Athletic identity strength ( $b = -.06$ ,  $t(221) = -1.96$ ,  $SE_b = .03$ ,  $p = .052$ ,  $\eta_p^2 = .02$ ) and gender identity strength ( $b = .15$ ,  $t(221) = 1.64$ ,  $SE_b = .09$ ,  $p = .10$ ,  $\eta_p^2 = .01$ ) emerged as statistically insignificant covariates.

Controlled for athletic and gender identity strength, the effect for recall type was not statistically significant ( $F(1, 221) = .36$ ,  $MSE = 20.36$ ,  $p = .55$ ,  $\eta_p^2 = .002$ ). On average, participants subjected to a gender stereotype of lower female athleticism, who as well recalled a mastery autobiographical memory as opposed to a routine morning autobiographical memory, did not have significantly higher personal self-esteem (see Table 7). That is, the data do not support hypothesis 1.

Table 7

*Mean PSE per Recall Type for the Analysis With Dutch Surveys*

Type of memory	<i>M</i>	<i>SD</i>	<i>N</i>
Mastery	18.85	4.70	81
Daily morning routine	19.31	4.47	144
Total	19.14	4.55	225

*Note.* PSE = personal self-esteem.

Results indicated that participants subjected to a gender stereotype of lower female athleticism, who as well recalled a mastery autobiographical memory as opposed to a routine morning autobiographical memory, did not have significantly higher personal self-esteem, both when participants who completed the Dutch variant of the survey were included and excluded. The results barely changed when the latter participants were included in the analysis and the absence of a statistically significant effect for recall type is most likely not due to lower power.

### **Hypothesis Test: Global Self-Esteem Moderates the Effect of Social Identity Threat on Self-Affirmation**

I anticipated participants subjected to a gender stereotype of lower female athleticism, who as well recall mastery autobiographical memories and are high (i.e., those who score one standard deviation or more above the sample mean) on global self-esteem, to score higher on personal self-esteem compared to participants who recall routine autobiographical memories (H3). Also, I did not anticipate personal self-esteem to differ across recall types for participants who are low (i.e., those who score one standard deviation or more below the sample mean) on global self-esteem (H4). I tested hypotheses 3 and 4 using a between-participants factorial analysis of variance (ANOVA) with personal self-esteem as the dependent variable and recall type and global self-esteem as fixed factors. I examined the main effects for recall type and global self-esteem, as well as their interaction effect.

While the main effect for recall type was not statistically significant ( $F(1, 86) = .14$ ,  $MSE = 13.12$ ,  $p = .71$ ,  $\eta_p^2 = .002$ ), the main effect for global self-esteem was;  $F(2, 86) = 24.99$ ,  $MSE = 13.12$ ,  $p = .00$ ,  $\eta_p^2 = .37$ . Participants subjected to a gender stereotype of lower female athleticism, who as well are high on global self-esteem, scored significantly higher on personal self-esteem ( $M = 13.28$ ,  $SE = .98$ ,  $N = 16$ ) compared to those who are average (i.e., those who score anywhere between one standard deviation above or below the sample mean) ( $M = 18.39$ ,  $SE = .49$ ,  $N = 60$ ) or low ( $M = 22.73$ ,  $SE = .91$ ,  $N = 16$ ) on global self-esteem. In turn, participants subjected to a gender stereotype of lower female athleticism, who as well are average on global self-esteem, scored significantly higher on personal self-esteem ( $M = 18.39$ ,  $SE = .49$ ) compared to those who are low ( $M = 22.73$ ,  $SE = .91$ ) on global self-esteem. These results are not extraordinarily surprising given that global and personal self-esteem are highly correlated, regardless of experimental condition (see Table 4). The interaction between recall type and global self-esteem was not statistically significant ( $F(2, 86) = .02$ ,  $MSE = 13.12$ ,  $p = .98$ ,  $\eta_p^2 = .001$ ).

The data do not support hypotheses 3 and 4. Participants subjected to a gender stereotype of lower female athleticism, who as well recalled mastery autobiographical memories and are high on global self-esteem, did not score significantly higher on personal self-esteem compared to participants who recalled routine autobiographical memories (H3). Also, personal self-esteem did not differ significantly across recall types for participants who are low on global self-esteem (H4). Global self-esteem does not seem to moderate the effect of social identity threat on self-affirmation in this specific sample of female athletes.

## **Hypothesis Test: Athletic Identity Strength Moderates the Effect of Social Identity**

### **Threat on Self-Affirmation**

I anticipated participants subjected to a gender stereotype of lower female athleticism, who as well recall mastery autobiographical memories and are high (i.e., those who score one standard deviation or more above the sample mean) on athletic identity strength, to score higher on personal self-esteem compared to participants who recall routine autobiographical memories (H5). Also, I did not anticipate personal self-esteem to differ across recall types for participants who are low (i.e., those who score one standard deviation or more below the sample mean) on athletic identity strength (H6). I tested hypotheses 5 and 6 using a between-participants factorial ANOVA with personal self-esteem as the dependent variable and recall type and athletic identity strength as fixed factors. I examined the main effects for recall type and athletic identity strength, as well as their interaction effect.

The main effects for both recall type ( $F(1, 187) = .20, MSE = 20.68, p = .65, \eta_p^2 = .001$ ) and athletic identity strength ( $F(2, 187) = 1.59, MSE = 20.68, p = .21, \eta_p^2 = .02$ ) were not statistically significant. The interaction between recall type and athletic identity strength was also not statistically significant ( $F(2, 187) = .02, MSE = 20.68, p = .98, \eta_p^2 = .00$ ).

The data do not support hypotheses 5 and 6. Participants subjected to a gender stereotype of lower female athleticism, who as well recalled mastery autobiographical memories and are high on athletic identity strength, did not score significantly higher on personal self-esteem compared to participants who recalled routine autobiographical memories (H5). Also, personal self-esteem did not differ significantly across recall types for participants who are low on athletic identity strength (H6). Athletic identity strength does not seem to moderate the effect of social identity threat on self-affirmation in this specific sample of female athletes.

**Hypothesis Test: Gender Identity Strength Moderates the Effect of Social Identity****Threat on Self-Affirmation**

I anticipated participants subjected to a gender stereotype of lower female athleticism, who as well recall mastery autobiographical memories and are high (i.e., those who score one standard deviation or more above the sample mean) on gender identity strength, to score higher on personal self-esteem compared to participants who recall routine autobiographical memories (H7). Also, I did not anticipate personal self-esteem to differ across recall types for participants who are low (i.e., those who score one standard deviation or more below the sample mean) on gender identity strength (H8). I tested hypotheses 7 and 8 using a between-participants factorial ANOVA with personal self-esteem as the dependent variable and recall type and gender identity strength as fixed factors. I examined the main effects for recall type and gender identity strength, as well as their interaction effect.

While the main effect for recall type was not statistically significant ( $F(1, 187) = 1.08$ ,  $MSE = 19.74$ ,  $p = .30$ ,  $\eta_p^2 = .01$ ), the main effect for gender identity strength was;  $F(2, 187) = 6.18$ ,  $MSE = 19.74$ ,  $p = .003$ ,  $\eta_p^2 = .06$ . Participants subjected to a gender stereotype of lower female athleticism, who as well are high or low on gender identity strength, scored significantly higher on personal self-esteem ( $M = 17.39$ ,  $SE = .68$ ,  $N = 47$  and  $M = 17.88$ ,  $SE = .81$ ,  $N = 34$ ) compared to those who are average (i.e., those who score anywhere between one standard deviation above or below the sample mean) ( $M = 19.95$ ,  $SE = .43$ ,  $N = 112$ ) on gender identity strength. Then again, participants subjected to a gender stereotype of lower female athleticism, who as well are high on gender identity strength, did not score significantly different on personal self-esteem ( $M = 17.39$ ,  $SE = .68$ ) compared to those who are low ( $M = 17.88$ ,  $SE = .81$ ) on gender identity strength. The interaction between recall type and gender identity strength was not statistically significant ( $F(2, 187) = .66$ ,  $MSE = 19.74$ ,  $p = .52$ ,  $\eta_p^2 = .01$ ).

The data do not support hypotheses 7 and 8. Participants subjected to a gender stereotype of lower female athleticism, who as well recalled mastery autobiographical memories and are high on gender identity strength, did not score significantly higher on personal self-esteem compared to participants who recalled routine autobiographical memories (H7). Also, personal self-esteem did not differ significantly across recall types for participants who are low on gender identity strength (H8). Gender identity strength does not seem to moderate the effect of social identity threat on self-affirmation in this specific sample of female athletes.

*An exploration of the main effect for gender identity strength.*

The main effect for gender identity strength described above is quite remarkable, which sparked a more thorough investigation. To explore the distribution of gender identity a bit further, I conducted an exploratory analysis examining the latter variable using the Normal Q-Q plot and boxplot. Both tests of normality (i.e., Kolmogorov-Smirnov and Shapiro-Wilk) yielded a statistically significant result ( $p < .001$ ). That is, gender identity is in all likelihood not normally distributed. A graphical representation of the distribution of gender identity can be seen in Figure 1. Notice that the points do not fall on the straight line. Normal Q-Q plots that display this conduct ordinarily mean that the data have more extreme values than would be anticipated if it genuinely came from a distribution that is normal. Namely, if the data genuinely came from a distribution that is normal, the points should form a line that is approximately straight. To investigate whether extreme values (i.e. outliers) in the dataset are responsible for finding such a remarkable main effect, I explored the boxplot for gender identity in more detail. The latter boxplot yielded six extreme values. To explore how and if the results would change when these six extreme values were excluded from the analysis, I again conducted a between-participants factorial ANOVA with personal self-

esteem as the dependent variable and recall type and gender identity strength as fixed factors.

I examined the main effect for gender identity strength.

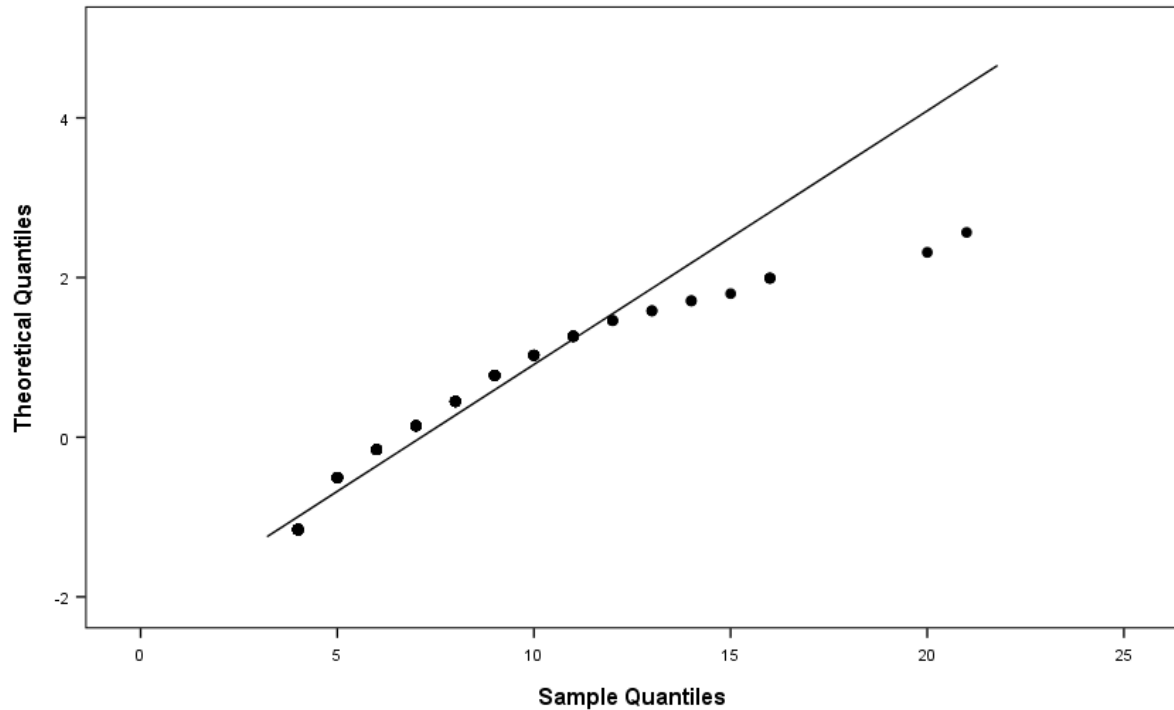


Figure 1. Normal Q-Q plot of the distribution of gender identity

Results indicated that the main effect for gender identity strength was statistically significant ( $F(2, 181) = 5.27, MSE = 20.05, p = .006, \eta_p^2 = .06$ ). Participants subjected to a gender stereotype of lower female athleticism, who as well are high on gender identity strength, scored significantly higher on personal self-esteem ( $M = 17.39, SE = .69, N = 47$ ) compared to those who are average ( $M = 20.01, SE = .47, N = 96$ ) on gender identity strength. Then again, participants subjected to a gender stereotype of lower female athleticism, who as well are low on gender identity strength, did not score significantly different on personal self-esteem ( $M = 18.52, SE = .69, N = 44$ ) compared to those who are high ( $M = 17.39, SE = .69$ ) or average ( $M = 20.01, SE = .47$ ) on gender identity strength.

Although the distribution of gender identity seemed to approach normality a little more when the extreme values were excluded from the analysis (see Figure 2), both tests of normality still yielded a statistically significant result ( $p < .001$ ). Apart from the fact that the results changed to a small extent when the extreme values were excluded from the analysis (participants who are low on gender identity strength no longer scored significantly higher on personal self-esteem compared to those who are average on gender identity strength), the main effect for gender identity strength remained quite remarkable. That is, extreme values in the dataset are most likely not responsible for finding such a remarkable main effect for gender identity strength.

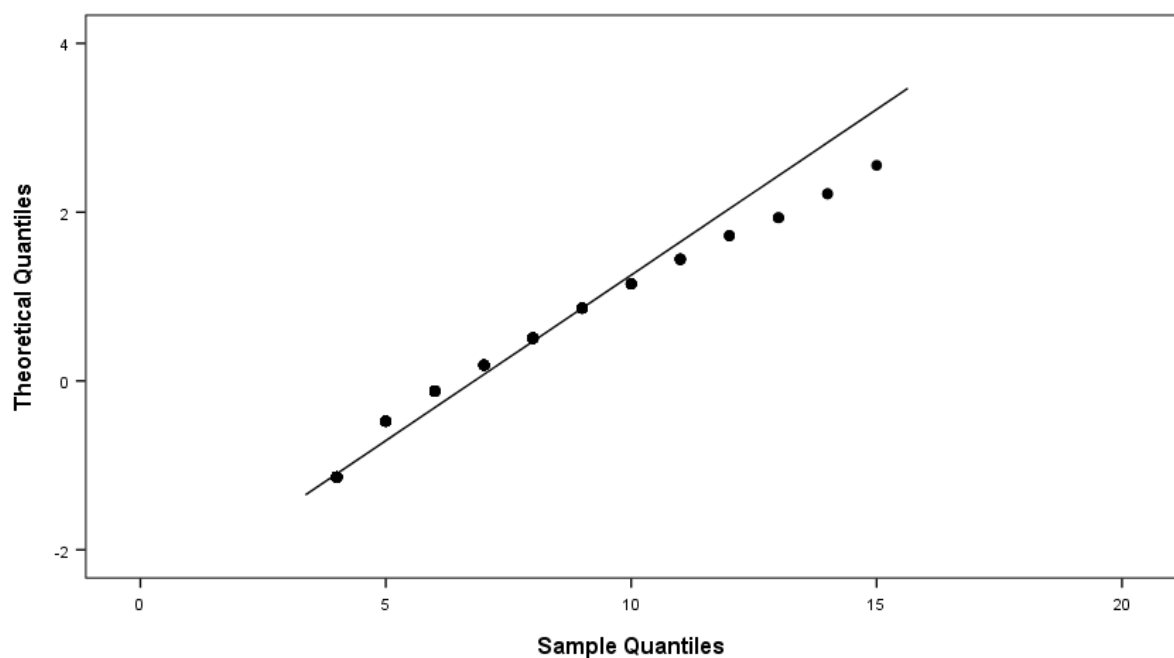


Figure 2. Normal Q-Q plot of the distribution of gender identity for the analysis without outliers

### Exploratory Analyses

The fact that hypothesis 1 is not supported by the data prompted new analyses to be performed. I employed the coding manual for self-defining events by Thorne and McLean



(2001) to exclude non-eligible autobiographical memory entries. All entries that gave prominence to the chronicles of one's individual or family/group's vigorous attempts at accomplishment and were not interrelated with the content of the threat were coded as mastery memories. On the other hand, all entries that underscored topics of life-threatening events, recreational experiences, guilt/shame, and relationships were coded as non-mastery memories. To explore how and if the results would change when a less restrictive coding approach (i.e., when all entries that were interrelated with the content of the threat were included in the analysis) and an approach without coding (i.e., when all entries that were interrelated with the content of the threat and underscored topics of life-threatening events, recreational experiences, guilt/shame, and relationships were included in the analysis) were used, I again conducted between-participants factorial analyses of covariance with personal self-esteem as the dependent variable and athletic and gender identity strength as covariates. Of note is that, for both exploratory approaches (i.e., less restrictive and without coding), autobiographical memory entries that were absent were not included in the analysis.

***A less restrictive coding approach.***

While gender identity strength emerged as a statistically insignificant covariate ( $b = .64$ ,  $t(202) = 1.25$ ,  $SE_b = .51$ ,  $p = .21$ ,  $\eta_p^2 = .01$ ), athletic identity strength emerged as a statistically significant one;  $b = -1.03$ ,  $t(202) = -1.99$ ,  $SE_b = .52$ ,  $p = .048$ ,  $\eta_p^2 = .02$ .

Controlled for athletic and gender identity strength, the effect for recall type was not statistically significant ( $F(1, 202) = .41$ ,  $MSE = 21.11$ ,  $p = .53$ ,  $\eta_p^2 = .002$ ). On average, participants subjected to a gender stereotype of lower female athleticism, who as well recalled a mastery autobiographical memory as opposed to a routine morning autobiographical memory, did not have significantly higher personal self-esteem (see Table 8). That is, apart from the fact that athletic identity strength emerged as a statistically

significant covariate, the results hardly changed when a less restrictive coding approach was used.

Table 8

*Mean PSE per Recall Type for the Less Restrictive Coding Approach*

Type of memory	<i>M</i>	<i>SD</i>	<i>N</i>
Mastery	18.80	5.05	84
Daily morning routine	19.14	4.31	122
Total	19.00	4.62	206

*Note.* PSE = personal self-esteem.

***An approach without coding.***

Athletic identity strength ( $b = -.55$ ,  $t(252) = -1.11$ ,  $SE_b = .50$ ,  $p = .27$ ,  $\eta_p^2 = .005$ ) and gender identity strength ( $b = .53$ ,  $t(252) = 1.12$ ,  $SE_b = .47$ ,  $p = .27$ ,  $\eta_p^2 = .005$ ) emerged as statistically insignificant covariates.

Controlled for athletic and gender identity strength, the effect for recall type was not statistically significant ( $F(1, 252) = .03$ ,  $MSE = 22.22$ ,  $p = .86$ ,  $\eta_p^2 = .00$ ). On average, participants subjected to a gender stereotype of lower female athleticism, who as well recalled a mastery autobiographical memory as opposed to a routine morning autobiographical memory, did not have significantly higher personal self-esteem (see Table 9). That is, the results hardly changed when an approach without coding was used.

Table 9

*Mean PSE per Recall Type for the Approach Without Coding*

Type of memory	<i>M</i>	<i>SD</i>	<i>N</i>
Mastery	19.07	5.10	130
Daily morning routine	19.13	4.29	126
Total	19.10	4.71	256

*Note.* PSE = personal self-esteem.

### **Exploratory Analyses: Different Sources of Threat and Resulting Responses**

Recall that categorization threat is a form of social identity threat where, in inter-group scenarios, a stereotypical group characteristic is ascribed to a group of people (Branscombe et al., 1999). According to Branscombe et al. (1999), this involuntary categorization may be extraordinarily threatening in a setting where that group membership insinuates poor performance or competence. When individuals find themselves in such a setting, they frequently look for cues to find out whether or not they will be criticized founded on their social identities (Steele et al., 2002). For example, given that negative stereotypes exist about women's athletic abilities, women may be particularly affected by and perceptive to cues in sports settings that notify whether or not they fit and have the means to triumph in this domain. Previous research suggests that feedback, which is defined as "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007, p. 81), is such a key situational cue that can either improve or deteriorate the outcomes of those at risk of encountering categorization threat (Cohen, Steele, & Ross, 1999; Yeager et al., 2014; Park, Kondrak, Ward, & Streamer, 2018). According to Park et al. (2018), the source of feedback (i.e., whether the feedback was coming from a man or a woman), among others, may particularly influence the outcomes of groups that are negatively stereotyped. Their research revealed that women who received *positive* feedback in a field in which they face stereotype vulnerability (e.g., particular sports or math settings) from a male authority figure compared to a female authority figure felt more confident, competent, and reported greater feelings of belonging. That is, the source of feedback may induce different responses in women who are negatively stereotyped in a particular field (Park et al., 2018). Although Park et al. (2018) mainly focused on *positive* feedback in their study, I have emphasized the opposite in my study, namely *negative* feedback. In my study, a gender stereotype of lower

female athleticism was ascribed to female athletes participating in a predominantly male-dominated sport. In other words, female athletes participating in a predominantly male-dominated sport received *negative* feedback regarding their athletic qualities (of note is that this feedback was made-up and therefore not specifically coming from a man or woman). Despite the fact that future research is necessary, it intuitively makes sense to expect that the outcomes of women who receive *negative* (versus *positive*) feedback in a field in which they face stereotype vulnerability from a (recalled) male authority figure will be more negative (versus positive) compared to the outcomes of women who receive *negative* (versus *positive*) feedback in a field in which they face stereotype vulnerability from a (recalled) female authority figure.

To explore the latter, I conducted between-participants factorial analyses of variance per recall condition (i.e., mastery memory versus routine morning memory) with personal self-esteem as the dependent variable and the recalled gender of the author of the presented vignette (i.e., male, female, or not recalled) as the independent variable. I conducted between-participants factorial analyses of variance separately per recall condition to disentangle the effect of recalled gender from the recall of either a mastery or a routine morning memory.

The majority of participants indicated that they remembered that the author of the presented vignette is male (see Table 10).

Table 10

*Frequency of Recalled Genders per Recall Condition*

Type of memory	Male <i>N</i>	Female <i>N</i>	Not recalled <i>N</i>	Total <i>N</i>
Mastery	46	13	12	71
Daily morning routine	93	11	18	122
Total <i>N</i>	139	24	30	193

While the results indicated no statistically significant difference in personal self-esteem between participants who did not recall the gender of the author ( $M = 17.17$ ,  $SD = 4.51$ ), recalled the gender of the author to be male ( $M = 19.24$ ,  $SD = 4.89$ ), or recalled the gender of the author to be female ( $M = 19.15$ ,  $SD = 5.35$ ) for the mastery memory recall condition ( $F(2, 68) = .87$ ,  $MSE = 24.17$ ,  $p = .42$ ,  $\eta_p^2 = .03$ ), the opposite was true for the routine morning memory recall condition;  $F(2, 119) = 8.57$ ,  $MSE = 16.53$ ,  $p = .00$ ,  $\eta_p^2 = .13$ . For the latter condition, results indicated that participants subjected to a gender stereotype of lower female athleticism, who as well did not recall the gender of the author ( $M = 16.11$ ,  $SD = 3.97$ ), scored significantly higher on personal self-esteem compared to participants who recalled the gender of the author to be either male ( $M = 19.34$ ,  $SD = 4.08$ ) or female ( $M = 22.36$ ,  $SD = 4.11$ ). Additionally, participants subjected to a gender stereotype of lower female athleticism, who as well recalled the gender of the author to be male ( $M = 19.34$ ,  $SD = 4.08$ ), scored significantly higher on personal self-esteem compared to participants who recalled the gender of the author to be female ( $M = 22.36$ ,  $SD = 4.11$ ). The mean differences in personal self-esteem per recalled gender for the routine morning memory recall condition can be seen in Table 11.

Table 11

*Mean Difference in PSE per Recalled Gender*

Recalled gender		<i>MD</i>	<i>SE</i>	<i>p</i>
Male	Female	-3.02	1.30	.02
	Not recalled	3.23	1.05	.003
Female	Male	3.02	1.30	.02
	Not recalled	6.25	1.56	.00
Not recalled	Male	-3.23	1.05	.003
	Female	-6.25	1.56	.00

*Note.* PSE = personal self-esteem.

I expected the outcomes of women who receive negative feedback in a field in which they face stereotype vulnerability from a (recalled) male authority figure to be more negative compared to the outcomes of women who receive negative feedback in a field in which they face stereotype vulnerability from a (recalled) female authority figure. The data do not support this expectation, regardless of recall condition. For the routine morning memory recall condition, the outcomes of women who received negative feedback from a (recalled) female authority figure were significantly more negative compared to the outcomes of women who received negative feedback from a (recalled) male authority figure. That is, the (recalled) source of feedback did induce significantly different responses in women who are negatively stereotyped in a particular field (which is in line with Park et al.'s (2018) finding), but not in the direction I anticipated. For the mastery memory recall condition, the (recalled) source of feedback did not induce significantly different responses in women who are negatively stereotyped in a particular field, which is not in line with Park et al.'s (2018) finding and my expectation.

### **Discussion**

Previous research has shown that self-affirmation can mitigate the adverse consequences of social identity threat through diverse mechanisms (Tavitian-Elmadjian et al., 2020; Layous et al., 2017; Sherman et al., 2014; Cook et al., 2012). For the present study, I looked into mastery recall of autobiographical memories as such a specific mechanism of self-affirmation in a frame of reference of categorization threat based on gender. Specifically, I set out to assess whether self-affirmation through the recall of mastery autobiographical memories could help female athletes participating in a predominantly male-dominated sport (ice hockey) cope with gender stereotype threat (RQ1). Previously done research has shown that recalling mastery autobiographical memories can be advantageous for self-esteem when confronted with a threat to social identity (Tavitian-Elmadjian et al., 2020). However, in my

study, the notion that self-affirmation via the recall of mastery memories safeguards self-esteem when threatened was not supported. For my study, a gender stereotype of lower female athleticism was ascribed to female athletes participating in a predominantly male-dominated sport. I anticipated participants subjected to a gender stereotype of lower female athleticism, who as well recall mastery autobiographical memories to possess higher personal self-esteem compared to participants who recall routine morning autobiographical memories (H1). However, I found that participants subjected to a gender stereotype of lower female athleticism, who as well recalled mastery autobiographical memories, did not have significantly higher personal self-esteem compared to participants who recalled routine morning autobiographical memories.

In addition to assessing whether self-affirmation through the recall of mastery autobiographical memories could help female athletes participating in a predominantly male-dominated sport cope with gender stereotype threat, I set out to assess whether global self-esteem and the strength of athletic and gender identity moderate the relationship between the type of memory recall (i.e., mastery or routine morning autobiographical memories) and personal self-esteem (RQ2). Previous research has shown that different classes of responses to social identity threat are likely to be exhibited by people who vary in the degree to which they identify with a particular social group (Schmader et al., 2015; Armenta, 2010; Spears et al., 1997; McCoy & Major, 2003; Tavitian-Elmadjian et al., 2020). Besides, previous work that focused on the moderating role of self-esteem in social identity threat management has produced mixed results. Some researchers suggest that self-affirmation affords identical effects for individuals high and low on self-esteem (Spencer et al., 2001). Others, however, suggest that, among individuals who are low on self-esteem, adverse outcomes follow self-affirmation (Creswell et al., 2005; Stone & Cooper, 2003; Steele et al., 1993). My results displayed that participants subjected to a gender stereotype of lower female athleticism, who

as well recalled mastery autobiographical memories and are high on global self-esteem, did not score significantly higher on personal self-esteem compared to participants who recalled routine autobiographical memories (H3). Also, personal self-esteem did not differ significantly across recall types for participants who are low on global self-esteem (H4). These findings are not in line with hypotheses 3 and 4. Neither in line with the hypotheses (i.e., hypotheses 5, 6, 7, and 8), my results displayed that participants subjected to a gender stereotype of lower female athleticism, who as well recalled mastery autobiographical memories and are high on athletic (H5) or gender (H7) identity strength, did not score significantly higher on personal self-esteem compared to participants who recalled routine autobiographical memories. Also, personal self-esteem did not differ significantly across recall types for participants who are low on athletic (H6) or gender (H8) identity strength. That is, I did not have a statistically significant interaction term for both global self-esteem and group identification strength.

In brief, the results of the current study do not support the notion that self-affirmation through the recall of mastery autobiographical memories helps female ice hockey players cope with gender stereotype threat, as well as the notion that global self-esteem and group identification strength moderate the relationship between the type of memory recall and personal self-esteem.

### **Mastery Memories as Tools for Self-Affirmation**

In my study, I found no evidence for the idea that mastery autobiographical memories are a way of self-affirmation for female athletes when threatened. Even though the results indicated that the threat manipulation was regarded by participants in the purposed manner (i.e., predominantly negative versus neutral or positive (H2)), I found no statistically significant difference in personal self-esteem between participants who were asked to either



recall a mastery or a routine morning autobiographical memory. Be that as it may, the threat manipulation most likely prompted a troublesome psychological state that affected self-esteem. Since people have a congenital need to uphold a positive self-evaluation (Steele, 1988; Tesser, 1988), it is expected that mechanisms directed at upholding or reinstating the latter were activated (Sherman et al., 2000; Aronson et al., 1999). Diverse mechanisms have been put forward to uphold or reinstate a positive self-evaluation, one of which is self-affirmation by recalling a mastery autobiographical memory (Tavitian-Elmadjian et al., 2020). However, in my study, the notion that self-affirmation via the recall of mastery memories safeguards self-esteem when threatened was not supported. The first possible explanation for this could be that participants engaged in a different cognitive response to deal with the troublesome psychological state prompted by the threat manipulation.

Participants were asked to read a vignette featuring a blatant threat in which I explicitly refer to female athletes in negative terms. In accordance with previous research on stereotype reactance (Kray, Thompson, & Galinsky, 2001), the nature of this blatant threat-triggering sign may have culminated in threat-relevant defensiveness (i.e., reactance). That is, the application of a blatant threat could have contributed to a reactance effect and, as a consequence, may have resulted in defensive coping (e.g., a more powerful identification or disidentification with the threatened social group) regardless of whether participants were given the chance to self-affirm by recalling a mastery memory. Such a reactance effect could be a possible explanation for not finding a statistically significant difference in personal self-esteem between participants who were asked to either recall a mastery or a routine morning autobiographical memory. Participants who were given the chance to self-affirm may not have done so and, as a result, the latter did not help them to avert the negative effects of threat and boost self-esteem, which may have led to a near equalization of the average score

on personal self-esteem of participants who were either given the opportunity to self-affirm ( $M = 18.87$ ) or not ( $M = 19.14$ ).

I did not expect to find a statistically insignificant difference in personal self-esteem between participants who were asked to either recall a mastery or a routine morning autobiographical memory, let alone a near equalization of the average scores on the latter variable. In addition to the aforementioned first possible explanation for the latter, I suggest that COVID-19 could be a second possible explanation. Worldwide, the outbreak of the Coronavirus turned a lot of human lives upside down. People were no longer allowed to go out on the streets without a valid reason and were expected to work from home. In addition, gyms, restaurants, and other entertainment venues were not accessible. All of the latter prevented people from living their lives as they usually do. That is, for the majority of people, their routines were broken. According to Cohen et al. (2000), individuals may employ, in order to self-affirm, any self-contemplative writing opportunity. Therefore, participants who were not given the opportunity to self-affirm were asked to call to mind memories of their daily morning routine, which is oftentimes neutral and therefore generally does not contribute to potential self-affirmation. The data collection took place during this unusual period, and since the majority of people did not have a daily morning routine at the time, participants may have gotten very confused when asked to remember and write about their daily morning routine. Unfortunately, since limited cognitive control is inherent to online experiments (and the data collection took place online), it is not possible for me to know exactly what the participants recalled, but to get an idea of the amount of participants who actually named the Coronavirus, I examined the qualitative morning routine autobiographical memory entries in greater detail. As it turned out, only 16 of the 122 (13%) participants who were asked to call to mind memories of their daily morning routine actually named it (i.e., the Coronavirus). Despite the fact that only 13% of the participants actually mentioned COVID-19, it cannot be

ruled out that the latter did not also affect other participants. Participants may have been unsure about which routine to write down (i.e., the routine before or after the outbreak of the virus) and this uncertainty could have led to other feelings and memories emerging. Since individuals may employ, in order to self-affirm, any self-contemplative writing opportunity (Cohen et al., 2000), it could be that the emergence of these other feelings and memories has contributed to possible self-affirmation. As a result, an artificially low ceiling might have been created, making it problematic to compare two groups with a hypothesis such as “The mean of group A will be higher than that of group B” (i.e., a ceiling effect might have happened). Note that although participants did not actually write down any other memories than daily morning routines, they may still have recalled one that contributed to self-affirmation. However, as mentioned above, limited cognitive control makes it impossible for me to know this.

### **Global Self-Esteem and Group Identification Strength as Moderators**

When investigating the effects of identity threat and affirmation, it is of value to give some thought to potential moderators. Global self-esteem and group identification strength have been proposed as moderators of frequent occurrence in previously conducted research that focused on social identity threat (Schmader et al., 2015; MCQueen & Klein, 2006). These factors presumably play a role in the effectiveness of particular self-affirmation mechanisms (Sherman & Cohen, 2002). In line with previous work (e.g., Tavitian-Elmadjian et al., 2020; Tavitian et al., 2019; Schmader et al., 2015; McCoy & Major, 2003; Spencer et al., 2001; Steele et al., 1993), it seemed to me necessary to explore these frequently occurring moderators in the present study as well. My results demonstrated that global self-esteem and group identification strength did not moderate the relationship between self-affirmation and personal self-esteem. In other words, I did not have a statistically significant interaction term for either variable when exploring the interaction. A possible explanation for not finding

statistically significant interaction terms could be that the (average) scores on all three variables (i.e., global self-esteem and athletic and gender identity strength) are rather low (as previously indicated, lower (average) scores equate to higher global self-esteem, athletic identity strength, and gender identity strength), indicating a potential ceiling effect. 92% of participants scored lower than the scale midpoint anchored at 4 “neither agree nor disagree” on global self-esteem ( $M = 2.53$ ). Besides, 97% of participants scored lower than the average of 16 on gender identity strength ( $M = 7.13$ ) and 76% of participants scored lower than the average of 40 on athletic identity strength ( $M = 32.74$ ). Based on the latter, there does indeed appear to have been a potential ceiling effect. A large part of the participants scores low on the variables, which makes mutual discrimination very difficult or even impossible (separating all low scores into low and high relative to each other simply does not work) and impedes meaningful analysis of the results. Note that a ceiling effect generally occurs when most values obtained for a variable converge on the upper limit of the scale used in its assessment. However, in my study, a potential ceiling effect occurs when most values obtained for a variable converge on the lower limit of the scale used in its assessment.

### **Limitations and Directions for Future Research**

The present study comes with some limitations that may serve as valuable input for future research. First of all is the question of whether participants are not working their way through the survey without paying sufficient attention to the assignments at hand, which is inherent to experiments conducted online. To assess whether participants paid sufficient attention to completing the survey, I added a timing question to the latter. In particular, I added a timing question to the page on which participants were asked to carefully read the experimental vignette and accompanying instruction text. This timing question kept track of how many total seconds had passed before the respondent moved on to the next page of the survey (i.e., the total amount of time the respondent spent on the page reading the

experimental vignette and accompanying instruction text). For the routine morning memory recall condition, participants spent an average of 11.17 seconds on the page ( $Mdn = 8.03$ ). For the mastery memory recall condition, the average amount of time spent on the page was even less, namely 10.07 seconds ( $Mdn = 8.66$ ). That is, a remarkable number of participants spent little time (some even less than 2 seconds) on the page, particularly when looking at the medians. Reading both the experimental vignette and accompanying instruction text or even only the instruction text in such little time is humanly impossible. Given all of the latter, it is likely that participants did not pay sufficient attention to completing the survey, which is a critical issue in the current study and another (third) possible explanation for not finding a statistically significant effect for recall type (the majority of participants did not carefully read the experimental vignette, which likely prevented them from feeling threatened by it, which in turn likely meant that there was no defensive response for the self-affirmation to ameliorate). A possible explanation for the fact that a remarkable number of participants spent little time reading the experimental vignette could be that they recognized the threat quickly and moved on to the next page of the survey. The ubiquitous held conviction that, compared to men, women have inferior athletic ability (Knight & Giuliano, 2001; Biernat & Vescio, 2002) has been around for centuries and most female athletes immediately recognize this stereotype. As a result, participants may not have bothered to reread the same old stereotype they have heard so many times by now. Future studies may benefit from enabling participants to move on to the next page only after a certain number of seconds, likely preventing them from progressing through the survey without paying sufficient attention to the assignments at hand. In other words, doing the latter likely increases the chance that participants will pay sufficient attention to completing the survey.

Another limitation of the present study is the (non-probability) sampling method used. Participants were recruited via snowball sampling through social media platforms and email.

While snowball sampling has some significant advantages (e.g., it is generally much quicker to find samples), it also has a number of crucial disadvantages. One of these critical drawbacks is that, since individuals allude to those whom they are acquainted with and have comparable traits, this sampling method may hold a possible margin of error and sampling bias. This means that I may have only reached a small group of individuals and therefore may not have finished the study with conclusive results. Although the representativeness of the sample is generally not guaranteed when using the snowball sampling method (it is extremely challenging to get an idea of the true distribution of both the sample and the population), using the latter can be an effective way (and sometimes even the only way) to collect meaningful information and data if the sample for the study is very exceptional or restricted to a small subset of the population (e.g., female ice hockey players) (Baltar & Brunet, 2012). Baltar and Brunet's (2012) research has shown that the use of virtual networks (e.g., Facebook) in non-probabilistic samples can broaden the geographical scope of the latter and eases the identification of individuals with barriers to access. Therefore, the principal benefits of this technique are that it can enlarge the sample size and its representativeness. Since the sample for my study is restricted to a small subset of the population, the purposive sampling method used is an effective way to collect meaningful information and data. In other words, it is not likely that the snowball sampling method used is a drastic issue in the current research.

It is also sensible to consider the way the memories were coded as a possible limitation. Consistent with the study by Tavitian-Elmadjian et al. (2020), I employed the coding manual for self-defining events by Thorne and McLean (2001) to exclude non-eligible autobiographical memory entries. However, while two independent coders coded the entries in the study by Tavitian-Elmadjian et al. (2020), the entries in my study were coded by only one coder (i.e., by me). Previous research has shown that, since coding (of autobiographical memory entries) is based on a certain degree of subjective interpretation, it is in all likelihood

superior to do it with two (or more) independent coders (Berends & Johnston, 2005; Ryan, 1999). For example, if two (or more) independent coders code the same autobiographical memory entry as a non-mastery memory, its interpretation is likely not a product of guesswork and therefore more trustworthy. According to Ryan (1999), agreement among two (or more) independent coders suggests that they are administering the codes in the same way and are thus acting as reliable measurement instruments and yields evidence that a code has external validity and is not merely an invention of the researcher's imagination. In other words, agreement among multiple coders is often used to check the reliability and validity of the coding process. The results of both exploratory analyses I conducted (i.e., a less restrictive coding approach and an approach without coding) indicated that the effect for recall type was not statistically significant, regardless of coding approach. Considering these results, it is improbable that the fact that the entries were coded by only one coder is a drastic obstacle in the present study.

Future studies may benefit from assessing the effect of mastery recall on female athletes' personal self-esteem during a less unusual period. Formulated differently, future studies may benefit from replicating the present study when all the hustle and bustle surrounding the Coronavirus has (largely) passed and people have been able to resume their routines. For my study, the data collection took place during the outbreak of the Coronavirus, which turned a lot of human lives upside down and prevented people from living their lives as they usually do. Since the majority of people did not have a daily morning routine at the time, participants may have gotten very confused when asked to remember and write about their daily morning routine. They may have been unsure about which routine to write down (i.e., the routine before or after the outbreak of the virus), which could have led to other feelings and memories emerging, which in turn could have contributed to possible self-affirmation. In the future, it is important to assess the effect of mastery recall on personal

self-esteem when daily morning routines are back in place and the possible emergence of other feelings and memories may no longer contribute to self-affirmation. Then, a statistically significant difference in personal self-esteem between participants who were asked to either recall a mastery or a routine morning autobiographical memory might be found.

Future studies may also benefit from assessing the effect of mastery recall on personal self-esteem in specific cultural groups. Previous research has shown that culture moderates the effectiveness of specific self-affirmation mechanisms (Hoshino-Browne et al., 2005). Hoshino-Browne et al.'s (2005) research highlights how culture moderates the effectiveness of self-affirmation. They suggest that, in different cultures where the self is structured dissimilarly, what constitutes a self-affirmation and a threat are dissimilar too. For example, to figure out how self-affirmation functions cross-culturally, the effect of mastery recall (as a specific mechanism of self-affirmation) on personal self-esteem could be assessed in a more collectivist culture (e.g., Japan) rather than in a more individualist culture (e.g., the U.S.A.). By doing this, strikingly different results might be found. The results of the pre-test of the experimental vignette indicated that the latter would most likely be perceived by participants in the purposed manner (i.e., threatening). However, these results came from Dutch female ice hockey players. Perhaps participants in my study, who are mainly from the U.S.A., did not perceive the experimental vignette as threatening. As a consequence, there may have been no defensive response for the self-affirmation to ameliorate, which in turn could have led to a statistically insignificant effect for recall type. Given that both the Netherlands and the U.S.A are White, Educated, Industrialized, Rich, and Democratic (WEIRD) countries (Henrich, Heine, & Norenzayan, 2010), it is improbable, but not entirely ruled out, that the experimental vignette was not perceived as threatening by participants in my study. That is to say that, the latter may be yet another (fourth) possible explanation for not finding a statistically significant effect for recall type.



Last of all, future studies may also address the role of the source of the threat in triggering different responses in women who are negatively stereotyped in a particular field. Previous research has shown that women who receive positive feedback in a field in which they face stereotype vulnerability from a male authority figure compared to a female authority figure feel more confident, competent, and report greater feelings of belonging (Park et al., 2018). Based on intuition, I expected the outcomes of women who receive negative feedback in a field in which they face stereotype vulnerability from a (recalled) male authority figure to be more negative compared to the outcomes of women who receive negative feedback in a field in which they face stereotype vulnerability from a (recalled) female authority figure. However, the data do not support this expectation, regardless of recall condition. For the mastery memory recall condition, the (recalled) source of feedback did not induce significantly different responses in women who are negatively stereotyped in a particular field. However, for the routine morning memory recall condition, the outcomes of women who received negative feedback from a (recalled) female authority figure were significantly more negative compared to the outcomes of women who received negative feedback from a (recalled) male authority figure. Due to a lack of previous research on this topic and the fact that my expectation was based only on intuition, it is difficult for me to provide a significant explanation for the latter results. However, it could be that the outcomes of women who recalled the negative feedback to come from another woman were the most negative because these women felt betrayed (at least, for the routine morning memory recall condition). They may have wondered why another woman, who is likewise the target of the negative feedback and therefore should share an identical psychological pain, would write something along those lines, which could have negatively affected their self-esteem (even more so than when they recalled the negative feedback to come from a man with whom they do not share this resemblance). Future studies may benefit from assessing the specific causal

mechanisms by which the (perceived) source of the threat induces different responses of female athletes to it, as well as specific self-affirmation interventions targeting female athletes. It might also be interesting for future studies to assess how the Multi-Threat Framework (Shapiro & Neuberg, 2007), which posits that six qualitatively distinct stereotype threats excavate from the crossing of the source and the target of the threat, relates to the role of the source of the threat in triggering different responses in negatively stereotyped women.

## **Conclusion**

To my knowledge, the current study, in addition to Tavitian-Elmadjian et al.'s (2020) and Tavitian et al.'s (2019) studies, is the third to look into mastery recall of autobiographical memories as a specific machinery of self-affirmation in a frame of reference of categorization threat. In accordance with Tavitian-Elmadjian et al.'s (2020) work that advocates expanding research to other instruments, participants, and experimental manipulations, my study places the interpretation of threat management and identity threat within the confines of a bigger context. Particularly, in this study, I set out to assess whether self-affirmation through the recall of mastery autobiographical memories could help female ice hockey players cope with gender stereotype threat. Unlike Tavitian-Elmadjian et al. (2020), who assessed collective self-esteem as a measure of collective self-worth in order to enlarge the fit with the threatened domain of identity (i.e., the group), I assessed the effect of mastery recall on personal self-esteem. According to them, assessing the effect of mastery recall on personal self-esteem would possibly match the personal content of the self-affirmation task better and, consequently, yield a more sizeable effect. A significant amount of researchers have shown that autobiographical memories can be essential for a wide range of domains, such as maintaining collective self-esteem (Tavitian-Elmadjian et al., 2020), conserving social relationships (Alea & Bluck, 2003; Nelson, 1993; Neisser, 1988), resolving difficulties (Cohen, 1998), and conveying empathy (Bluck, Alea, Habermas, & Rubin, 2005). However,

in my study, the functional use of autobiographical memories was not supported.

Specifically, in my study, the notion that drawing on one's personal memories is conducive to self-esteem when confronted with a threat to social identity was not supported. The latter demonstrates that considering the specificities of the threat context is of great importance for comprehending identity processes, since anticipations deduced from one context may not always hold true in another. This also applies to global self-esteem and group identification strength as moderators of the relationship between the type of memory recall and personal self-esteem, for which I did not have statistically significant interaction terms in my study. To find out whether or not the use of mastery memories as a specific machinery in self-affirmation is a robust phenomenon, I point to the need for future research into identity threat and affirmation, including other manipulations, (cultural) groups, and measures. In view of the present study specifically, I yielded clear ideas as to why I possibly did not find support for the notion that drawing on one's personal memories is conducive to self-esteem when confronted with a threat to social identity, which could be beneficial for future research. It is possible that, under alternative circumstances, mastery memories are helpful tools of self-affirmation for female athletes, which could be of great added value for, among others, sports and performance psychologists, coaches, and the athletes themselves. For example, if mastery memories turn out to be helpful tools of self-affirmation, female athletes may use them to maintain self-esteem in the face of negative gender stereotypes regarding athleticism and perhaps improve their performance accordingly. The current study already provides some insights for future research to investigate the effect under alternative circumstances and, with its findings, makes an important contribution to the present body of literature on threat management and social identity threat, highlighting possible boundaries of the fruitfulness of the use of autobiographical memories as a particular machinery in self-affirmation.

## References

- Alea, N., & Bluck, S. (2003). Why are you telling me that? A conceptual model of the social function of autobiographical memory. *Memory*, 11(2), 165–178.  
<https://doi.org/10.1080/741938207>
- Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology*, 16(1), 94–98. <https://doi.org/10.1037/a0017564>
- Aronson, J., Cohen, G., & Nail, P. R. (1999). Self-affirmation theory: An update and appraisal. In E. Harmon-Jones & J. Mills (Eds.), *Science conference series. Cognitive dissonance: Progress on a pivotal theory in social psychology* (p. 127–147). American Psychological Association. <https://doi.org/10.1037/10318-006>
- Aronson, J., & Inzlicht, M. (2004). The Ups and Downs of Attributional Ambiguity. *Psychological Science*, 15(12), 829–836. <https://doi.org/10.1111/j.0956-7976.2004.00763.x>
- Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An Organizing Framework for Collective Identity: Articulation and Significance of Multidimensionality. *Psychological Bulletin*, 130(1), 80–114.  
<https://doi.org/10.1037/0033-2909.130.1.80>
- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science*, 7(6), 543–554.  
<https://doi.org/10.1177/1745691612459060>

Baltar, F., & Brunet, I. (2012). Social research 2.0: virtual snowball sampling method using Facebook. *Internet Research*, 22(1), 57–74.

<https://doi.org/10.1108/10662241211199960>

Bastian, B., Jetten, J., & Ferris, L. J. (2014). Pain as Social Glue: Shared Pain Increases Cooperation. *Psychological Science*, 25(11), 2079–2085.

<https://doi.org/10.1177/0956797614545886>

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, 46(3), 610–620. <https://doi.org/10.1037/0022-3514.46.3.610>

Begg, C. B., & Mazumdar, M. (1994). Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*, 50(4), 1088–1101. <https://doi.org/10.2307/2533446>

Beilock, S. L., Bertenthal, B. I., Mccoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review*, 11(2), 373–379.

<https://doi.org/10.3758/bf03196585>

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, 130(4), 701–725. <https://doi.org/10.1037/0096-3445.130.4.701>

Beilock, S. L., & McConnell, A. R. (2004). Stereotype Threat and Sport: Can Athletic Performance Be Threatened? *Journal of Sport and Exercise Psychology*, 26(4), 597–609. <https://doi.org/10.1123/jsep.26.4.597>

- Berends, L., & Johnston, J. (2005). Using multiple coders to enhance qualitative analysis: The case of interviews with consumers of drug treatment. *Addiction Research & Theory*, 13(4), 373–381. <https://doi.org/10.1080/16066350500102237>
- Biernat, M., & Vescio, T. K. (2002). She Swings, She Hits, She's Great, She's Benched: Implications of Gender-Based Shifting Standards for Judgment and Behavior. *Personality and Social Psychology Bulletin*, 28(1), 66–77. <https://doi.org/10.1177/0146167202281006>
- Blanton, H., Cooper, J., Slkurnik, I., & Aronson, J. (1997). When Bad Things Happen to Good Feedback: Exacerbating the Need for Self-Justification with Self-Affirmations. *Personality and Social Psychology Bulletin*, 23(7), 684–692. <https://doi.org/10.1177/0146167297237002>
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and High Blood Pressure: The Role of Stereotype Threat. *Psychological Science*, 12(3), 225–229. <https://doi.org/10.1111/1467-9280.00340>
- Bluck, S., Alea, N., Habermas, T., & Rubin, D. C. (2005). A TALE of Three Functions: The Self-Reported Uses of Autobiographical Memory. *Social Cognition*, 23(1), 91–117. <https://doi.org/10.1521/soco.23.1.91.59198>
- Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N., Ellemers, R., Spears, & B., Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 35– 58). Oxford, England: Blackwell.
- Branscombe, N. R., Schmitt, M. T., & Harvey, R. D. (1999). Perceiving pervasive discrimination among African Americans: Implications for group identification and

well-being. *Journal of Personality and Social Psychology*, 77(1), 135–149.

<https://doi.org/10.1037/0022-3514.77.1.135>

Brewer, B. W., Van Raalte, J. L., & Linder, D. E. (1993). Athletic identity: Hercules' muscles or Achilles heel? *International Journal of Sport Psychology*, 24(2), 237–254.

Brewer, W. F. (1986). What is autobiographical memory? In D. C. Rubin (Ed.), *Autobiographical memory* (p. 25–49). Cambridge University Press.

<https://doi.org/10.1017/CBO9780511558313.006>

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: the effect of expectancy on performance. *European Journal of Social Psychology*, 33(2), 267–285. <https://doi.org/10.1002/ejsp.145>

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>

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(2), 136–144. <https://doi.org/10.1016/j.psychsport.2012.10.005>

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(2), 143–158. <https://doi.org/10.1123/jsep.30.2.143>

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(3), 235–260.

<https://doi.org/10.1080/03610730590948177>

Cherney, I. D., & Campbell, K. L. (2011). A League of Their Own: Do Single-Sex Schools Increase Girls' Participation in the Physical Sciences? *Sex Roles*, 65(9–10), 712–724.

<https://doi.org/10.1007/s11199-011-0013-6>

Cook, J. E., Purdie-Vaughns, V., Garcia, J., & Cohen, G. L. (2012). Chronic threat and contingent belonging: Protective benefits of values affirmation on identity development. *Journal of Personality and Social Psychology*, 102(3), 479–496.

<https://doi.org/10.1037/a0026312>

Cohen, G. (1998). Aging and autobiographical memory. In C. P. Thompson, D. J. Herrmann, D. Bruce, J. D. Read, D. G. Payne, & M. P. Toglia (Eds.), *Autobiographical memory: Theoretical and applied perspectives* (pp. 105–123). Hillsdale, NJ: Erlbaum.

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When Beliefs Yield to Evidence: Reducing Biased Evaluation by Affirming the Self. *Personality and Social Psychology Bulletin*, 26(9), 1151–1164. <https://doi.org/10.1177/01461672002611011>

Cohen, G. L., & Sherman, D. K. (2014). The Psychology of Change: Self-Affirmation and Social Psychological Intervention. *Annual Review of Psychology*, 65(1), 333–371.

<https://doi.org/10.1146/annurev-psych-010213-115137>

Creswell, J. D., Welch, W. T., Taylor, S. E., Sherman, D. K., Gruenewald, T. L., & Mann, T. (2005). Affirmation of Personal Values Buffers Neuroendocrine and Psychological Stress Responses. *Psychological Science*, 16(11), 846–851.

<https://doi.org/10.1111/j.1467-9280.2005.01624.x>



- 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(12), 1615–1628. <https://doi.org/10.1177/014616702237644>
- Doosje, B., Branscombe, N. R., Spears, R., & Manstead, A. S. R. (1998). Guilty by association: When one's group has a negative history. *Journal of Personality and Social Psychology*, 75(4), 872–886. <https://doi.org/10.1037/0022-3514.75.4.872>
- 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. <https://doi.org/10.1016/j.lindif.2015.12.018>
- 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(2), 455–463. <https://doi.org/10.1111/j.0006-341x.2000.00455.x>
- Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist*, 50(3), 145–158. <https://doi.org/10.1037/0003-066x.50.3.145>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/bf03193146>
- Finnigan, K. M., & Corker, K. S. (2016). Do performance avoidance goals moderate the effect of different types of stereotype threat on women's math performance? *Journal of Research in Personality*, 63, 36–43. <https://doi.org/10.1016/j.jrp.2016.05.009>

Flore, P. C. (2018). *Stereotype threat and differential item functioning: A critical assessment*.

Enschede: Gildeprint Drukkerijen.

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(1), 25–

44. <https://doi.org/10.1016/j.jsp.2014.10.002>

Fredricks, J. A., & Eccles, J. S. (2005). Family Socialization, Gender, and Sport Motivation and Involvement. *Journal of Sport and Exercise Psychology*, 27(1), 3–31.

<https://doi.org/10.1123/jsep.27.1.3>

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>

Gentile, A., Boca, S., & Giammusso, I. (2018). '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 and Exercise*, 39, 95–103.

<https://doi.org/10.1016/j.psychsport.2018.07.013>

Gibson, C. E., Losee, J., & Vitiello, C. (2014). A Replication Attempt of Stereotype

Susceptibility (). *Social Psychology*, 45(3), 194–198. [https://doi.org/10.1027/1864-](https://doi.org/10.1027/1864-9335/a000184)

[9335/a000184](https://doi.org/10.1027/1864-9335/a000184)

- Gray, J. R. (2004). Integration of Emotion and Cognitive Control. *Current Directions in Psychological Science*, 13(2), 46–48. <https://doi.org/10.1111/j.0963-7214.2004.00272.x>
- 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(2), 207-226.
- Hardin, M., Simpson, S., Whiteside, E., & Garriss, K. (2007). The Gender War in U.S. Sport: Winners and Losers in News Coverage of Title IX. *Mass Communication and Society*, 10(2), 211–233. <https://doi.org/10.1080/15205430701265737>
- Haslam, S. A., & Reicher, S. (2006). Stressing the group: Social identity and the unfolding dynamics of responses to stress. *Journal of Applied Psychology*, 91(5), 1037–1052. <https://doi.org/10.1037/0021-9010.91.5.1037>
- Heidrich, C., & Chiviawowsky, S. (2015). Stereotype threat affects the learning of sport motor skills. *Psychology of Sport and Exercise*, 18, 42–46. <https://doi.org/10.1016/j.psychsport.2014.12.002>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/s0140525x0999152x>
- Hermann, J. M., & Rumrich, K. (2018). Empirische Arbeit: Stereotype Threat im Sportunterricht. *Psychologie in Erziehung und Unterricht*, 65(4), 288. <https://doi.org/10.2378/peu2018.art17d>

- 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. <https://doi.org/10.1016/j.psychsport.2016.06.010>
- 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(1), P3–P11. <https://doi.org/10.1093/geronb/58.1.p3>
- 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(1), 48–55. <https://doi.org/10.1016/j.psychsport.2013.09.001>
- Hoshino-Browne, E., Zanna, A. S., Spencer, S. J., Zanna, M. P., Kitayama, S., & Lackenbauer, S. (2005). On the cultural guises of cognitive dissonance: The case of Easterners and Westerners. *Journal of Personality and Social Psychology*, 89(3), 294–310. <https://doi.org/10.1037/0022-3514.89.3.294>
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Jasinskaja-Lahti, I., Liebkind, K., & Solheim, E. (2008). To Identify or Not To Identify? National Disidentification as an Alternative Reaction to Perceived Ethnic Discrimination. *Applied Psychology*, 58(1), 105–128. <https://doi.org/10.1111/j.1464-0597.2008.00384.x>
- Keller, J. (2002). Blatant stereotype threat and women’s math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, 47(3/4), 193–198. <https://doi.org/10.1023/a:1021003307511>

- Knight, J. L., & Giuliano, T. A. (2001). He's a Laker; She's a "Looker": The Consequences of Gender-Stereotypical Portrayals of Male and Female Athletes by the Print Media. *Sex Roles*, 45(3/4), 217–229. <https://doi.org/10.1023/a:1013553811620>
- Koivula, N. (1999). Sport participation: Differences in motivation and actual participation due to gender typing. *Journal of Sport Behavior*, 22, 360–380. Retrieved from <http://www.southalabama.edu/psychology/journal.html>.
- 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>
- Layous, K., Davis, E. M., Garcia, J., Purdie-Vaughns, V., Cook, J. E., & Cohen, G. L. (2017). Feeling left out, but affirmed: Protecting against the negative effects of low belonging in college. *Journal of Experimental Social Psychology*, 69, 227–231. <https://doi.org/10.1016/j.jesp.2016.09.008>
- Lewis, B. P., & Linder, D. E. (1997). Thinking about Choking? Attentional Processes and Paradoxical Performance. *Personality and Social Psychology Bulletin*, 23(9), 937–944. <https://doi.org/10.1177/0146167297239003>
- Liu, T. J., & Steele, C. M. (1986). Attributional analysis as self-affirmation. *Journal of Personality and Social Psychology*, 51(3), 531–540. <https://doi.org/10.1037/0022-3514.51.3.531>
- 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(4), 379–392.  
<https://doi.org/10.1123/jsep.2014-0284>
- Mazerolle, M., Régner, I., Morisset, P., Rigalleau, F., & Huguet, P. (2012). Stereotype Threat Strengthens Automatic Recall and Undermines Controlled Processes in Older Adults. *Psychological Science*, 23(7), 723–727. <https://doi.org/10.1177/0956797612437607>
- McCoy, S. K., & Major, B. (2003). Group Identification Moderates Emotional Responses to Perceived Prejudice. *Personality and Social Psychology Bulletin*, 29(8), 1005–1017.  
<https://doi.org/10.1177/0146167203253466>
- McQueen, A., & Klein, W. M. P. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity*, 5(4), 289–354.  
<https://doi.org/10.1080/15298860600805325>
- Moon, A., & Roeder, S. S. (2014). A Secondary Replication Attempt of Stereotype Susceptibility (). *Social Psychology*, 45(3), 199–201. <https://doi.org/10.1027/1864-9335/a000193>
- Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology*, 1(1), 35–59.  
<https://doi.org/10.1080/09515088808572924>
- Nelson, K. (1993). The Psychological and Social Origins of Autobiographical Memory. *Psychological Science*, 4(1), 7–14. <https://doi.org/10.1111/j.1467-9280.1993.tb00548.x>
- 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(6), 1314–1334. <https://doi.org/10.1037/a0012702>

- Perkins-Ceccato, N., Passmore, S. R., & Lee, T. D. (2003). Effects of focus of attention depend on golfers' skill. *Journal of Sports Sciences*, 21(8), 593–600.  
<https://doi.org/10.1080/0264041031000101980>
- Phinney, J. S., Chavira, V., & Tate, J. D. (1993). The Effect of Ethnic Threat on Ethnic Self-Concept and Own-Group Ratings. *The Journal of Social Psychology*, 133(4), 469–478. <https://doi.org/10.1080/00224545.1993.9712171>
- Picho, K., Rodriguez, A., & Finnie, L. (2013). 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), 299–333.  
<https://doi.org/10.1080/00224545.2012.737380>
- Priest, L. (2003). The Whole IX Yards: the impact of Title IX: the Good, the Bad and the Ugly. *Women in Sport & Physical Activity Journal*, 12(2), 27+. Retrieved from <https://link.gale.com/apps/doc/A112542448/AONE?u=tilburgb&sid=AONE&xid=ebe45399>
- Protzko, J., & Aronson, J. (2016). Context Moderates Affirmation Effects on the Ethnic Achievement Gap. *Social Psychological and Personality Science*, 7(6), 500–507.  
<https://doi.org/10.1177/1948550616646426>
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 27(2), 151–161.  
<https://doi.org/10.1177/0146167201272002>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Rubin, D. C. (1988). *Autobiographical Memory*. Cambridge, Verenigd Koninkrijk: Cambridge University Press.
- Ryan, G. (1999). Measuring the Typicality of Text: Using Multiple Coders for More Than Just Reliability and Validity Checks. *Human Organization*, 58(3), 313–322. <https://doi.org/10.17730/humo.58.3.g224147522545rln>
- Schmader, T., Block, K., & Lickel, B. (2015). Social Identity Threat in Response to Stereotypic Film Portrayals: Effects on Self-Conscious Emotion and Implicit Ingroup Attitudes. *Journal of Social Issues*, 71(1), 54–72. <https://doi.org/10.1111/josi.12096>
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115(2), 336–356. <https://doi.org/10.1037/0033-295x.115.2.336>
- Shapiro, J. R. (2011). Different Groups, Different Threats: A Multi-Threat Approach to the Experience of Stereotype Threats. *Personality and Social Psychology Bulletin*, 37(4), 464–480. <https://doi.org/10.1177/0146167211398140>
- Shapiro, J. R., & Neuberg, S. L. (2007). From Stereotype Threat to Stereotype Threats: Implications of a Multi-Threat Framework for Causes, Moderators, Mediators, Consequences, and Interventions. *Personality and Social Psychology Review*, 11(2), 107–130. <https://doi.org/10.1177/1088868306294790>
- Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of*



*Personality and Social Psychology*, 104(2), 277–288.

<https://doi.org/10.1037/a0030461>

Sherman, D. K., & Cohen, G. L. (2002). Accepting Threatening Information: Self-Affirmation and the Reduction of Defensive Biases. *Current Directions in Psychological Science*, 11(4), 119–123. <https://doi.org/10.1111/1467-8721.00182>

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology*, 38, 183–242.

[https://doi.org/10.1016/S0065-2601\(06\)38004-5](https://doi.org/10.1016/S0065-2601(06)38004-5)

Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., ... Cohen, G. L. (2014). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology*, 104(4), 591–618.

<https://doi.org/10.1037/a0031495>

Sherman, D. K., Nelson, L. D., & Steele, C. M. (2000). Do Messages about Health Risks Threaten the Self? Increasing the Acceptance of Threatening Health Messages Via Self-Affirmation. *Personality and Social Psychology Bulletin*, 26(9), 1046–1058.

<https://doi.org/10.1177/01461672002611003>

Smart Richman, L., & Leary, M. R. (2009). Reactions to discrimination, stigmatization, ostracism, and other forms of interpersonal rejection: A multimotive model. *Psychological Review*, 116(2), 365–383. <https://doi.org/10.1037/a0015250>

Souchon, N., Coulomb-Cabagno, G., Traclet, A., & Rasclé, O. (2004). Referees' Decision Making in Handball and Transgressive Behaviors: Influence of Stereotypes About

Gender of Players? *Sex Roles*, 51(7/8), 445–453.

<https://doi.org/10.1023/b:sers.0000049233.28353.f0>

Spears, R., Doosje, B., & Ellemers, N. (1997). Self-Stereotyping in the Face of Threats to Group Status and Distinctiveness: The Role of Group Identification. *Personality and Social Psychology Bulletin*, 23(5), 538–553.

<https://doi.org/10.1177/0146167297235009>

Spencer, S. J., Fein, S., & Lomore, C. D. (2001). Maintaining One's Self-Image Vis-à-Vis Others: The Role of Self-Affirmation in the Social Evaluation of the Self. *Motivation and Emotion*, 25(1), 41–65. <https://doi.org/10.1023/a:1010659805978>

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 261–302). New York: Academic Press.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613–629. <https://doi.org/10.1037/0003-066x.52.6.613>

Steele, C. M. (1998). Stereotyping and its threat are real. *American Psychologist*, 53(6), 680–681. <https://doi.org/10.1037/0003-066x.53.6.680>

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: Vol 34*, (pp. 379–440). Elsevier.
- Steele, C. M., Spencer, S. J., & Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology*, 64(6), 885–896. <https://doi.org/10.1037/0022-3514.64.6.885>
- 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(285), 30–34. <https://doi.org/10.1080/01621459.1959.10501497>
- Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta- analysis: Prevention, assessment and adjustments*. (pp. 99–110). New York, NY: Wiley.
- 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>
- Stone, J., & Cooper, J. (2003). The effect of self-attribute relevance on how self-esteem moderates attitude change in dissonance processes. *Journal of Experimental Social Psychology*, 39(5), 508–515. [https://doi.org/10.1016/s0022-1031\(03\)00018-0](https://doi.org/10.1016/s0022-1031(03)00018-0)
- Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology*, 44(2), 445–452. <https://doi.org/10.1016/j.jesp.2007.02.006>
- Tavitian, L. R., Bender, M., Van de Vijver, F. J. R., Chasiotis, A., & Vosgerichian, H. A. (2019). Collective References and Identity Threat among Lebanese and Lebanese-

Armenians in the Context of Continuous Intergroup Conflict. *Psychological Research Record*, 9(3), 25-43.

Tavitian-Elmadjian, L., Bender, M., Van de Vijver, F. J. R., Chasiotis, A., & Harb, C. (2020).

Autobiographical recall of mastery experiences is a mechanism of self-affirming under social identity threat. *The Journal of Social Psychology*, 160(1), 39–60.

<https://doi.org/10.1080/00224545.2019.1606775>

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L.

Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 181–227). New York: Academic Press.

Thorne, A., & McLean, K. C. (2001). *Manual for coding events in self-defining memories (Unpublished Manuscript)*. University of California, Santa Cruz, US.

Thorne, A., McLean, K. C., & Lawrence, A. M. (2004). When Remembering Is Not Enough:

Reflecting on Self-Defining Memories in Late Adolescence. *Journal of*

*Personality*, 72(3), 513–542. <https://doi.org/10.1111/j.0022-3506.2004.00271.x>

Walton, G. M., & Cohen, G. L. (2003). Stereotype Lift. *Journal of Experimental Social*

*Psychology*, 39(5), 456–467. [https://doi.org/10.1016/s0022-1031\(03\)00019-2](https://doi.org/10.1016/s0022-1031(03)00019-2)

Walton, G. M., & Spencer, S. J. (2009). Latent Ability. *Psychological Science*, 20(9), 1132–

1139. <https://doi.org/10.1111/j.1467-9280.2009.02417.x>

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat:

Group- and self-threat. *Journal of Experimental Social Psychology*, 44(3), 792–799.

<https://doi.org/10.1016/j.jesp.2007.07.005>

Zigerell, L. J. (2017). Potential publication bias in the stereotype threat literature: Comment on Nguyen and Ryan (2008). *Journal of Applied Psychology*, 102(8), 1159–1168.

<https://doi.org/10.1037/apl0000188>

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

### **Appendix A. Manipulation Vignette**

Compared to men, women are known to possess poor athletic ability. This imbalance is particularly noticeable in physical competitions such as ice hockey where aspects such as speed, strength, and robustness are essential. Athletically speaking, women are slow (when corrected for leg length, the average speed of women is  $\pm 78\%$  of that of men), fragile, and incapable. Moreover, women are physically weak with an average strength of only  $1/3$  that of men. These aspects are not in line with the requirements of full-contact sports such as ice hockey. Because of this poor athletic ability that women possess, they will most likely never be able to perform as well as men who do have these essential aspects. The latter is clearly reflected in, for example, the number of men's and women's teams participating in professional ice hockey competitions worldwide. 288 men's teams participate as opposed to just 5 women's teams.

**Appendix B. Advertisement Title and Text****Title**

Wanted, female ice hockey players of many different ages and countries.

**Text**

Hey! My name is Nicky van Doremalen and I study Psychology at Tilburg University. At the moment I am (hopefully) graduating and for this I need a little bit of your help. For my thesis I am looking for a lot of female ice hockey players of many different ages and countries who would like to complete my survey. Completing the survey takes about 10-15 minutes. Do you practice the sport of ice hockey and are you female? Would you please take part in my research then? You would really help me with that (and not only me, but probably many female athletes)! By clicking on the link below, you will be automatically redirected to the English version of the survey. Thank you very much in advance for your time and effort! Oh, and do not hesitate to share the link to my research with your friends and teammates!

### **Appendix C. Pre-Test Statements**

#### **Interviewee 1**

“The author of the description must be a guy who hates women”

“The content of the description is very negative and I feel threatened by it”

#### **Interviewee 2**

“I think the female athlete is portrayed very negatively in this description”

“Men have always been seen as more important, stronger, and bigger in sports and everyone seems to accept that image indiscriminately. I think the content of this description emphasizes that very much”

#### **Interviewee 3**

“The content of the description makes me feel inferior to men and not motivated to do my best in sports. I will probably never be as good as a man anyway”

“The author of this description has a very negative image of female and a superior image of male athletes”



## **Appendix D. Recall Type Manipulation**

### **Mastery recall**

Please take a few minutes to remember a time in your life when you experienced a difficult situation, and recall how you overcame it.

After that, please start writing down your memory in as much detail as possible. There are no right or wrong descriptions. You might want to include the setting or location, who else was there, and what happened, so that the event is described in a way that others will understand. Writing down the memory should not take longer than 5–10 min.

### **Routine morning recall**

Please take a few minutes to remember your typical, everyday morning routine. Recall what you usually do during a regular morning.

After that, please start writing down your memory in as much detail as possible. There are no right or wrong descriptions. You might want to include the setting or location, who else was there, and what happened, so that the event is described in a way that others will understand. Writing down the memory should not take longer than 5–10 min.