Accident Analysis and Prevention 85 (2015) 199–206

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Women drive better if not stereotyped

Angelica Moè[a,*], Mara Cadinu[a], Anne Maass[b]

[a] Department of General Psychology, University of Padova, Italy
[b] Department of Developmental and Social Psychology, University of Padova, Italy

ARTICLE INFO

Article history:
Received 21 April 2015
Received in revised form 2 August 2015
Accepted 24 September 2015

Keywords:
Driving
Women
Stereotype threat

ABSTRACT

A commonly held stereotype is that women are poor drivers. This stereotype is recognized and endorsed by women and girls very early on, long before taking their driving licence, nevertheless they are less involved in accidents and drive safer and less fast than men. In line with the stereotype threat theory, the present study tests the hypothesis that making the driving stereotype salient will lead women to underperform in a driving simulation task. In Experiment 1women in the stereotype threat condition were told that the aim of the study was to detect gender differences in driving whereas in a control condition no study aim was provided. In Experiment 2, two conditions were compared: stereotype threat (same instructions as in Experiment 1), and stereotype boost (the alleged goal was to compare driving ability of young vs. old people). As predicted, the results of both experiments showed that women under stereotype threat, as compared to either control or stereotype boost participants, doubled the number of mistakes. Nevertheless, they overall expected/self-reported to drive/have driven poorly. Importantly, their level of expectation was a significant predictor of their actual driving performance only in the stereotype threat condition. Implications of these effects of stereotype threat on women's driving performance and self-assessment are discussed.

1. Introduction

A commonly held stereotype leads to think that women are poor drivers (Chateignier et al., 2011). They are considered hesitant, overcautious and unsure behind the wheel (Berger, 1986; Harris and Miller, 2000), because of a presumed difficulty in making fast decisions (e.g., swerving or braking to avoid a collision). This supposed inability is considered the source of omission errors (e.g., stay too long at a stop sign, take too much time to start or to park), which are considered typical of women's driving style (Lawrence and Richardson, 2005). These beliefs are endorsed to different degrees, but they are widely shared, because they depend on gender role socialization and identification (Sibley and Harré, 2009); they define what is typically masculine or feminine, and driving is conceived as a masculine task. Consistent with these stereotypes, women tend to self-assess their driving abilities as poor (Özkan and Lajunen, 2006; Tronsmoen, 2008) and to feel less safe than men when driving (Bergdahl, 2007). Nevertheless, compared to men, they are more compliant with traffic rules (Yagil, 1998), engage less in risk behaviors (Bina et al., 2006), and therefore report fewer traffic violations (Lonczak et al., 2007) and are less likely to die in a car accident (Hanna et al., 2006). This suggests a discrepancy between self-assessment and driving performance by women.

Since the age of 10 people describe women as unskilled, but careful and compliant drivers (Granié and Papafava, 2011). In addition, this image of women as poor drivers becomes more negative with increasing age (from 10 to 16), importantly before people obtain their driving licence. This further supports the assumption that these beliefs do not depend on actual performance, i.e. on experiencing difficulties in driving, which would then cause women's underperformance. Therefore, one possibility is that stereotypical beliefs affect women's self-evaluations of driving abilities as well as their actual performance. Specifically, in the present research we test the prediction that activating the stereotype of women as poor drivers will impair women's driving performance and their self-assessments.

1.1. Stereotype threat and female driving

Following the stereotype threat (ST) theory (Steele and Aronson, 1995), stereotyped individuals tend to underperform when a negative belief about the performance of their group is made salient in a testing situation (e.g. 'women are poor drivers' before taking a driving test). Such negative stereotypes may be activated by

* Corresponding author at: Via Venezia, 8, 35131 Padova, Italy.
E-mail addresses: angelica.moe@unipd.it (A. Moè), mara.cadinu@unipd.it (M. Cadinu), anne.maass@unipd.it (A. Maass).

http://dx.doi.org/10.1016/j.aap.2015.09.021
0001-4575/

reminding people's social category (e.g. asking for gender, ethnicity or age before performing the task) or by stressing a task-relevant stereotype (e.g. providing instructions about supposed gender differences). Fear of failure, reduced expectations, performance anxiety, worries about (not) confirming the negative stereotype about the group one belongs to are among the factors responsible for performance decrements (Steele, 1997). Being negatively stereotyped induces self-uncertainty and triggers performance-related thoughts, often accompanied by a physiological stress response and by an effort to control the negative thoughts and expectations (Schmader and Croft, 2011). These processes reduce the cognitive resources and, in particular working memory capacity needed to perform the task, thus leading to impaired performance. ST effects occur mainly in evaluative contexts, in which negatively stereotyped individuals face difficult tasks (Aronson et al., 1998).

Effects of ST have been observed in a wide range of tasks and populations, such as African-Americans (Steele and Aronson, 1995) and low socioeconomic status individuals with intelligence tests (Croizet and Claire, 1998), and women with math (Cadinu et al., 2005; Good et al., 2008) and with spatial abilities (Moè and Pazzaglia, 2006; Moè, 2009). To the best of our knowledge, only two studies so far have addressed the potential role of ST in women's driving abilities. In Chateignier et al.'s (2011) study, women in the ST condition were told that the test measured their attitudes toward driving and they were induced to think about women's poor driving performance. Women under ST failed more questions about driving behaviors in hypothetical situations than women in a control condition who were told that the aim of the study was to test materials for future research. However, this study only assessed driving knowledge rather than actual driving behavior. This limitation was overcome in a study by Yeung and von Hippel (2008), who examined the effects of ST on female's driving performance in a simulation paradigm. They found that women under ST (told that the goal of the study was to understand why men drive better than women) were less able to face an unexpected event (a group of pedestrians crossing the road) than women in the control condition (told that the goal of the study was to investigate the mental processes involved in driving). Although very telling about ST in driving, the study of Yeung and von Hippel (2008) addresses a type of error that is atypical of women's real-life driving behavior given that women generally cause fewer severe accidents than men do (ISTAT, 2013; NHTSA, 2012). A more typical type of error, namely incompetent parking, was addressed in two additional studies (Derks et al., 2011; Wolf et al., 2010). However, neither of these included a control condition without stereotype activation, making it impossible to judge whether ST does, indeed, have detrimental effects on driving performance.

1.2. Stereotype threat and underlying mechanisms

ST may cause underperformance through a range of mechanisms. Among the most studied are anxiety (Osborne, 2007), intrusive thoughts (Cadinu et al., 2005), poor expectations of success (Cadinu et al., 2003), and reduction of working memory capacities (Schmader and Johns, 2003; Beilock et al., 2007). The fear of underperforming or the concern to demonstrate that the stereotype is not true require emotional regulation processes that absorb working memory resources that would otherwise be devoted to the task (Johns et al., 2008). This has been demonstrated in a wide range of tasks, including the area of driving, confirming that ST reduces working memory resources (Yeung and von Hippel, 2008) and that it affects cardiovascular responses indicative of threat or challenge in car parking (Derks et al., 2011). Using a lexical decision task, Chateignier et al. (2011) also found that women who were told that their driving skills would be measured recognized the word *anger* faster than controls, and the level of anger experienced caused underperformance in driving knowledge, supposedly because women were involved in inner speech and wished to disconfirm the stereotype. In line with this interpretation, Yeung and von Hippel (2008) reported that women under ST were more motivated to show they were good drivers than women in the control condition, and Chateignier et al. (2011) (Study 1) found that women under ST did not endorse the stereotype of women as poor drivers, and, in fact, considered themselves better drivers than men (for similar effects of driving stereotypes on elderly participants see Joanisse et al., 2013; Levy et al., 2013).

Together, these studies show that ST produces effects on performance through the involvement of emotional regulation and working memory reduction, but possibly also through self-evaluations and expectations, which can play a crucial role, given that self-assessment of overall driving ability – in the absence of stereotype induction – has been shown to be related to parking and overall driving performance (Tronsmoen, 2008; Wolf et al., 2010). In addition, these self-perceptions could be related with intrinsic motivation toward driving such as perceived difficulty of the task and safety, satisfaction, and enjoyment. Yet these variables have not been examined in relation to driving tasks and when a ST is induced.

1.3. The present study

Previous studies have found effects of ST induction on women, who showed performance decrements on knowledge tests about driving behavior in hypothetical situations (Chateignier et al., 2011) and a higher likelihood to hit pedestrians in a driving simulation task (Yeung and von Hippel, 2008). However, the effects of ST on daily driving under normal conditions have not been tested yet. This is surprising, because the content of the commonly held stereotype is that women are poor drivers (i.e. make more mistakes and are uncertain when making fast decisions on the road), not that they have more accidents. On the contrary, men are stereotypically perceived as reckless drivers, prone to drive fast, and to cause accidents (Mast et al., 2008).

Therefore, the goal of the present study is to test the effects of ST on women's common driving performance, measured through the number of mistakes, which overtly show poor driving skills in daily driving, and could lead to accidents under certain conditions. The aim was to simulate what happens in daily driving situations to women who are reminded of the negative stereotype about female drivers.

In addition to the number of driving mistakes, we measured women's expectations, the self-assessment of their driving performance, and their intrinsic motivation (perceived difficulty and safety, satisfaction, and enjoyment), which may also be affected by the ST induction. To the best of our knowledge, the effects of a ST induction on these variables have not been tested before. We hypothesized that a stereotypical message may exert negative effects on expectations, self-assessment of driving performance, and intrinsic motivation.

In sum, it was predicted that a ST message would affect women's performance, causing more driving mistakes, and possibly also lower expectations, self-assessment of driving performance, and intrinsic motivation.

2. Experiment 1

2.1. Method

2.1.1. Participants

Eighty-one women (M_{age} = 23.24, SD = 2.53) attending the University of Padua, Italy, participated on a voluntary basis. They had a driving licence for an average of 4.43 years (SD = 2.20), and self-rated their frequency of driving as M = 3.00 (SD = 1.73), on a scale 0 = never to 5 = everyday.

2.1.2. Materials

The following materials were used:

2.1.2.1. Stereotype threat and control instructions. Participants were asked to read on paper the following instructions, delivered twice in Italian (the authors can provide the original versions upon request), at the arrival in the lab and just before the driving simulation task: 'You are going to engage in a driving simulation task. The aim of the study is to detect differences in driving ability between women and men. Follow the driving rules as close as possible' (ST condition) or 'You are going to engage in a driving simulation task. Follow the driving rules as close as possible' (control condition).

2.1.2.2. Driving simulation task. The driving simulation was delivered by a pc connected with two pedals for accelerator and brake, and a steering-wheel. On the screen participants saw a road, a map (on the lower right corner) and a driving mirror (on the upper right corner), as shown in Fig. 1. The route was along a two-lane road. There were other vehicles on the road, buildings and green areas around, some stops and crossings, a bridge, and a gallery.

2.1.2.3. Self-assessment of driving performance. Self-assessed driving performance was measured twice, as expectation and as appraisal of performance, just before and just after the driving simulation task through the following two questions: 'how well do you think you will perform/have performed?', 'how many mistakes do you think you will make/have made?', on a scale ranging from 1 = very poorly/no mistake, 7 = very well, many mistakes. The second item was reverse scored so that higher scores indicate better driving performance. A single score was computed given that the two ratings were inter-correlated, r = .64, p < .001, and r = .77, p < .001, respectively before and after the driving simulation task.

2.1.2.4. Intrinsic motivation toward driving. After the driving simulation, participants were asked the following four questions about perceived difficulty, and safety, satisfaction, and enjoyment: 'How difficult was the driving task?' (reverse scored), 'Did you feel unsure during the driving task?' (reverse scored), 'Are you satisfied with your driving?', 'Did you enjoy the driving task?' (anchor points: 1 = not at all, 7 = very much). Cronbach alpha was .67, so a single intrinsic motivation score was computed.

Fig. 1. Example of screen images in the driving simulation task.

2.1.3. Procedure

Participants were tested individually by two female experimenters, blind to the condition that participants had been randomly assigned to, namely ST (n = 40) or control (n = 41). At the arrival in the lab participants were greeted and asked to sign a consent form in which they were informed that they would be asked to perform a driving simulation task and to answer some questions.

Participants were then delivered the instructions written on a sheet and asked to read them carefully. Right after, they were asked to answer the two questions about expected driving performance and expected mistakes, and to sit down in front of the pc. Then, the apparatus (a screen showing the course, the steering-wheel, the accelerator and the brake) was explained in detail to them. They were asked to put their feet on the pedals and their hands on the steering wheel and familiarize by driving freely for a couple of minutes.

2.1.3.1. Driving simulation. Once familiarized, participants were given a sheet in which they were reminded of the instructions together with the aim of the study (which varied according to experimental condition) and asked to perform the 5-min driving simulation task following the driving rules as closely as possible. At the end of the 5 min the driving simulation stopped, even if the participants had not finished the course.

During the driving task the two experimenters sat one on the right, the other on the left of the participant, so as to eliminate any potential bias due to visual perspective (counterbalanced across experimenters). They looked always at the driving simulation, and took note of the same kind of mistakes independently, by crossing on a list containing the following ten: (1) passing over a 100% red traffic light [0–4], (2) delay in starting (>5 s) [0–3], (3) late braking (>3 s) [0–2], (4) unmotivated stop [0–3], (5) road edge excursion [0–1], (6) crash [0–1], (7) taking the wrong side of the road [0–2], (8) passing the centreline [0–4], (9) following the wrong direction [0–4], (10) collision [0–4]. There were four traffic lights so that the maximum number of mistakes was four for point (1). For all other errors, the maximum number could not be predicted a priori. We therefore report the range of mistakes actually committed by our participants in square parentheses.

At the end of the driving simulation, participants were asked to respond to the self-assessment questions on the driving performance and on intrinsic motivation toward driving and to reply to the manipulation check question ('What were you told in the instructions?'). Finally participants were debriefed: they were explained that they were randomly assigned to one of two conditions, both of which were described, and they were informed about the aim of the study. Finally they were thanked for the participation. The whole procedure lasted 20–30 min.

2.1.3.2. Scoring of driving errors. Following the same procedure adopted in previous studies (Joanisse et al., 2013; Chateignier et al., 2011), and given that for most of the errors, the range was very restricted (0–2), thus giving unreliable measures, a global performance score was calculated for each participant by each experimenter by summing the number of participants' mistakes. Because the inter-rater agreement was r = .99, an average score was computed. Looking at the single disagreements between the two experimenters, there was complete agreement regarding the following five mistakes: passing over a red traffic light, delay in starting, unmotivated stop, passing the centreline, and following the wrong direction, while they disagreed on no more than 1 error for the other kinds of mistakes, with an overall disagreement of 2 mistakes (totaling 48 mistakes for experimenter 1 and 50 for experimenter 2). This suggests that there was no experimenter bias, that is sometimes experimenter 1 crossed an error which experimenter 2 did not notice, while sometimes it was the opposite.

Table 1
Expected, actual and post-task estimated driving performance in the two experiments. Participants in the stereotype threat (ST) condition differ from control or stereotype boost only in the number of actual mistakes. Standard deviations are in parenthesis.

	Experiment 1		Experiment 2	
	ST	Control	ST	Stereotype boost
Expected # of driving mistakes	–	–	11.20 (4.29)	10.60 (3.78)
Expected global driving performance (min–max 1–7)	4.11 (.88)	4.05 (1.05)	–	–
Actual # of mistakes in the driving simulation	6.50 (3.94)	3.27 (2.97)	8.29 (3.17)	2.77 (2.16)
Estimated # of driving mistakes	–	–	9.49 (5.48)	7.69 (5.16)
Estimated global driving performance (min–max 1–7)	3.14 (1.36)	3.15 (1.45)	3.46 (1.12)	3.86 (1.11)
Intrinsic motivation toward driving (min–max 1–7)	3.77 (1.02)	4.09 (.95)	3.82 (1.21)	3.79 (.98)

Note. Estimated global driving performance in Experiment 1 was the average of two items and in Experiment 2 was measured by just one item.

Having completed the itinerary within the 5-min limit is an index of good driving. In order to take this into account, and given that not all the participants finished the itinerary whereas some drove even longer (we set the time, 5 min, not the end point), the global performance score was corrected considering the point at which the participants had arrived. To do so, we took the raw score and added 1 error point if they arrived as far as passing the bridge, but did not finished the course, 2 error points if they did not even pass the bridge, at the same time and we subtracted 1 error point if they finished the course and came back passing the bridge again and subtracted 2 error points if they drove even longer. For example, if a participant's uncorrected score was 7 and she finished the course, the score was 7, but if she drove longer and come back 1 or 2 points were subtracted depending on whether she just passed the bridge again or drove even further, while 1 or 2 points were added if she passed the bridge but did not finish the course (1 point) or did not pass the bridge (2 points). The correlation between the two experimenters' scores was .97, $p < .001$. The range of uncorrected scores was 0–13 ($M = 4.08$, SD = 3.11), and of corrected scores was −1 to 15 ($M = 4.84$, SD = 3.82).

2.2. Results

2.2.1. Preliminary analyses

No difference in driving performance was found depending on the right or left position of the experimenters. The two groups (ST and control) did not differ on mean age, years since receiving their driving licence, or frequency of driving. The manipulation check revealed that one participant in the ST condition did not understand the instructions, so she was excluded from the analyses, and the final sample was of 39 ST and 41 control participants.

2.2.2. Effects of ST manipulation on performance

Overall participants made $M = 4.84$ (SD = 3.82) driving mistakes. As predicted, women under ST made more mistakes ($M = 6.50$, SD = 3.94) than those in the control condition ($M = 3.27$, SD = 2.97), $t(78) = 4.16$, $p < .001$. The Cohen d is .84: a large effect (Cohen, 1988). Among the mistakes, the most frequent were: passing the centerline (46 participants did so at least once), delay in starting (29 participants), passing over a red traffic light (27 participants), and unmotivated stop (26 participants). ST compared to control participants were more likely to pass over a red traffic light [22 participants under ST made this mistake one or more times ($M = 1.68$, SD = .84), whereas only 5 women in the control condition made this mistake once, $t(25) = 3.81$, $p = .001$] and to delay starting [22 women made this mistake one or more times, $M = 1.55$, SD = .67, whereas only 7 women in the control condition made this mistake once, $t(27) = 3.81$, $p < .001$]. All analyses were conducted on the corrected scores. Importantly, analyses using the uncorrected raw scores of performance led to similar results. In the latter case mean values of errors were 5.58 (SD = 3.26) in the ST condition and 2.66 (SD = 2.19) in the control, $t(78) = 4.72$, $p < .001$, Cohen $d = .94$, a large effect.

2.2.3. Effects of condition on self-report measures before and after driving task

As can be seen in Table 1, the experimental manipulation did not affect expected overall performance assessed prior to the driving simulation task (t-tests). Similarly, the post-experimental measures were unaffected by condition.

2.2.4. Correlations

The pattern of correlations did not change if the raw driving performance score was considered (r and p – if different – are reported after the slash). Results showed that, although experimental condition did not affect either expected performance or post-experimental self-report data, the correlational analyses showed that, while for both ST and control condition actual driving performance was related with post-task self-assessment ($r = -.74/-.72$ and $r = -.64/-.64$, $p < .001$, respectively for ST and control condition) and intrinsic motivation ($r = -.58/-.55$, $p < .001$, and $r = -.46/-.49$, $p = .002/.001$, respectively for ST and control condition), pre-task expectations were related with post-task measures only in the ST condition (self-assessment $r = .34$, $p = .03$, intrinsic motivation $r = .32$, $p = .04$). The pattern of correlations did not change if the raw driving performance score considered (r and p – if different – after the slash). These findings suggest that women under ST might have been more vigilant than women in the control condition so that their expectations and post-driving assessment were more closely linked.

2.3. Discussion

As expected, women made more mistakes in the ST than in the control condition. The effect was large, given that participants committed twice as many errors in the ST than in the control condition. Interestingly no difference in self-assessment of performance was found. Under ST women had the same expectations and rated the quality of their driving at the same level than in the control condition.

One limitation of this study is that women may have experienced the threat of underperforming even in the control condition, because the instructions referred to a driving task, and the label 'driving' could itself have activated the common stereotype of poor driving skills by women. Therefore we designed a second experiment comparing the ST condition with a stereotype boost condition. Stereotype boost is the increase in performance due to exposure of a positive stereotype about the group participants belong. Here, women were told that the aim of the study was to compare young and elderly people performance in driving. The hypothesis was to confirm the effects of the message inducing ST on performance and on self-assessment in comparison to the non-threatening condition, which referred to the positive stereotype of young people as good drivers.

3. Experiment 2

3.1. Method

3.1.1. Participants

Seventy women (M_{age} = 22.11, SD = 3.53) from the University of Padua, Italy, participated on a voluntary basis in this study. On average, they had a driving licence for 3.45 years (SD = 3.23) and self-rated their frequency of driving as M = 3.89 (SD = 1.10), on a scale 0 = never to 5 = everyday.

3.1.2. Materials

The materials below described were used.

3.1.2.1. Stereotype threat and stereotype boost instructions. The ST condition participants received the same instruction as in Experiment 1, delivered on a written sheet in Italian, twice, at arrival in the lab and just before the driving simulation task. The stereotype boost condition participants received the following message: 'You are going to engage in a driving simulation task. The aim of the study is to detect the differences in driving ability between young and elderly people. Follow the driving rules as much as possible'.

3.1.2.2. Driving simulation task. Same as in Experiment 1.

3.1.2.3. Expectations on driving performance. An important change was made from Experiment 1. After delivering the instructions, the following question was asked 'How many mistakes do you think you will make?' (write down the number).

3.1.2.4. Self-assessment of driving performance. After the driving simulation task the following two questions were asked: 'How many mistakes do you think you have made?' (write down the number) and 'How well do you think you have driven?' (anchoring points 1 = poorly to 7 = optimally).

3.1.2.5. Intrinsic motivation toward driving. Participants were asked the same four questions as in Experiment 1. Given the satisfactory internal consistency (Cronbach alpha = .78), the four items were averaged for each participant.

3.1.2.6. Stereotyped expectations. Right after the intrinsic motivation items, a fifth one asked to report participants opinions about the direction of expected differences: 'In your opinion, in this study, which was the experimenter's hypothesis?'. Participants had to chose one of the following two options: 'men perform better than women' versus 'women perform better than men' (ST condition), or 'young people perform better than the elderly' versus 'the elderly perform better than young people' (stereotype boost condition).

3.1.2.7. Procedure and design. The procedure was the same as in Experiment 1. Participants were randomly assigned to the ST condition (n = 35) or to the stereotype boost condition (n = 35), unknown to the experimenters (one male and one female) who scored the number of mistakes independently. Given that inter-rater agreement was r = .95, an average global performance score was computed, corrected for the length of the driven route, as in Experiment 1. The correlation between the uncorrected and corrected scores was .95, p < .001. The range of uncorrected scores was 0–14 (M = 5.28, SD = 3.87), of corrected 0–14.50 (M = 5.83, SD = 3.87).

3.2. Results

3.2.1. Preliminary analyses

The two groups (ST and stereotype boost) did not differ significantly with respect to age, years since obtaining the driving licence, or frequency of driving. All participants correctly remembered the instructions. Of the ST group 31 out of 35 participants reported to believe that men are better drivers than women, of the stereotype boost group 28 out of 35 reported that young people are better drivers than old people.

3.2.2. Effects of ST manipulation on performance

On average participants committed 5.53 (SD = 3.87) mistakes. As predicted, women in the ST condition made more mistakes (M = 8.29, SD = 3.17) than those in the stereotype boost condition (M = 2.77, SD = 2.16), $t(68)$ = 8.50, p < .001, again showing a large effect (Cohen d = 1.42). When considering only those participants who believed that men or young people are better drivers than women or old people, results were very similar, confirming that women made more mistakes in the ST condition (M = 8.53, SD = 3.13) than in the stereotype boost condition (M = 2.95, SD = 2.26), $t(57)$ = 7.78, p < .001, Cohen d = 1.52. Similar results were found considering the raw score of driving mistakes. Mean values of uncorrected errors were 7.93 (SD = 3.35) in the ST condition and 2.63 (SD = 2.16) in the stereotype boost, $t(68)$ = 7.86, p < .001, Cohen d = 1.37. When considering just those who believed in the stereotype mean errors were 8.24 (SD = 3.35) in the ST condition and 2.82 (SD = 2.31) in the stereotype boost condition, $t(57)$ = 7.15, p < .001, Cohen d = 1.36, a large effect.

3.2.3. Effects of condition on self-report measures before and after driving task

As can be seen in Table 1, the experimental manipulation did not affect expected errors assessed prior to the driving simulation task (t-tests). Similarly, the post-experimental measures, namely estimated mistakes, global assessment and intrinsic motivation, were unaffected by condition. This results' pattern did not change when only those who endorsed the respective stereotypes were considered.

3.2.4. Correlations

In the ST condition actual performance (number of mistakes) was related with expected driving mistakes (r = .46, p = .006), and self-assessed post-driving mistakes (r = .35/.34, p < .05). On the contrary, in the stereotype boost condition these correlations were far from significant, see also Fig. 2. Again, the pattern of correlations did not change when the raw number of mistakes was considered rather than the corrected score (r and p – if different – after the slash).

In addition, overall, the more pessimistic the pre-task expectation, the more negative was also the global post-self-assessment of performance r = −.37, p = .002, and the greater the estimated number of errors made during the driving task r = .39, p = .001. Both were significant in the ST condition (r = −.46, p = .005 and r = .42, p = .011, respectively) whereas in the stereotype boost condition only the

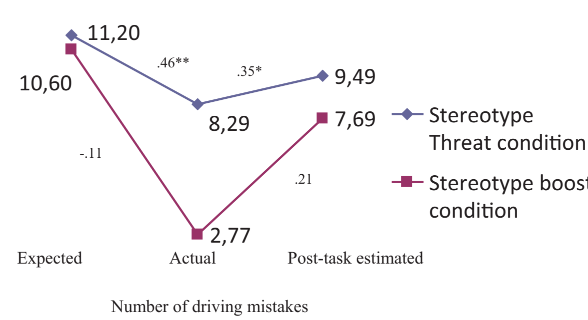

Fig. 2. Expected, actual and post-task estimated driving mistakes in the ST and stereotype boost conditions. Experiment 2. Along the arrows are the correlation indexes **p < .01.

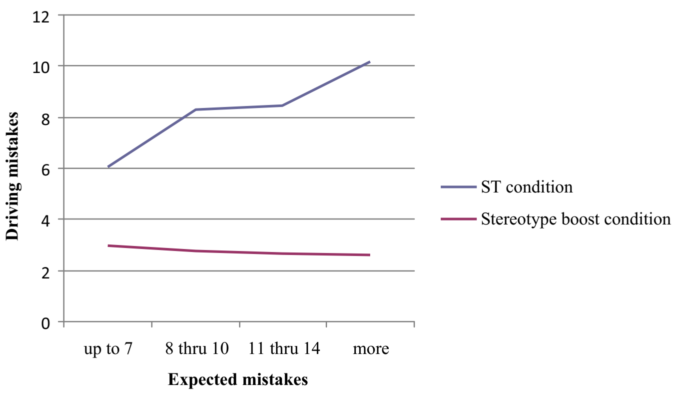

Fig. 3. Driving performance is affect by expected mistakes in ST condition. Experiment 2.

correlation between expected and post-task self-assessed number of mistakes was significant ($r = .34$, $p < .05$), thus confirming that also in Experiment 2 the evaluations expressed before and after the driving simulation task were more strongly associated with each other in the ST condition than in the stereotype boost condition.

3.2.5. Moderation analysis

Given that in this study we used the same measure (number of mistakes) for expected performance, actual performance, and post-driving self assessment, we ran a multiple regression analysis to test the hypothesis that lower expectations would lead to lower performance especially in the ST condition. At step 1 condition (0 = ST, 1 = stereotype boost) and expected mistakes were entered. Both affected driving mistakes: condition ($\beta = -.70$, $t = 8.52$, $p < .001$), expected mistakes ($\beta = .17$, $t = 2.07$, $p = .043$). At step 2 the interaction between condition and expectation was added. Expected mistakes was still significant ($\beta = .35$, $t = 3.33$, $p = .001$), condition was not, but interestingly the interaction was significant ($\beta = -.61$, $t = 2.59$, $p = .012$). Only in the ST condition did expected mistakes add to the threat and cause lowered performance (see Fig. 3 for mean values).

3.3. Discussion

As expected, women under ST made more mistakes than those in the stereotype boost condition, a result with a large effect size, as already observed in Experiment 1. Also, again, no difference was found in expected performance, self-assessment of performance, and motivation toward driving. Women appeared to self-assess their motivation and the quality of their driving performance independently of the instructions delivered. However, the expected mistakes played an important role in the ST (but not in the stereotype boost) condition. The better they expected their performance to be, the lower the number of mistakes they actually committed.

4. General discussion

Taking together, the results of the two experiments confirmed the prediction that women make more mistakes in performing a simulated driving task when exposed to ST (i.e., when led to think that the aim of the study was to verify the existence of gender differences in driving performance) as compared to a control group receiving no information about the aim of the study (Experiment 1) or one allegedly testing for differences between young and elderly individuals (Experiment 2), which should have favored young people. Interestingly, these mistakes were mainly omissions, exactly the type of error that is generally believed to be stereotypical of female driving behavior.

Although control participants made half as many mistakes as stereotyped women did, they self-assessed their performance no better. Indeed, women in both the control and the stereotype boost conditions expected to perform poorly and still believed to have performed badly after the driving simulation task. Apparently, they were unaware to have performed better than expected. Whereas expectations did not affect driving performance in the control condition, it did so in the ST condition so that expecting to perform poorly added to the threat arisen by the ST message. Implications of these effects of ST on driving performance and self-assessment are discussed below.

4.1. Female driving performance under ST

Previous studies had shown that under ST women are more likely to cause accidents (Yeung and von Hippel, 2008) and to fail questions about driving behaviors (Chateignier et al., 2011). Specifically in Yeung and von Hippel's study, the probability to hit a pedestrian doubled (59% vs. 25%) just as the number of mistakes doubled in the ST condition in both experiments reported here. No matter whether women were told nothing about the aim of the research (Experiment 1) or whether they were presented with an alleged aim that could have favored them (young people in Experiment 2), the results were quite similar. Specifically in Experiment 2, despite a potentially favorable stereotype boost condition, results simply confirmed the result pattern of Experiment 1, showing a strong and reliable increase in mistakes under ST. One explanation could be that, in an effort to disconfirm the stereotype, women in the ST condition drove more aggressively, failing to follow driving rules much like males do under normal driving conditions. Thus, future studies should investigate whether women under ST make more omission (typical female) or commission (typical male) errors. Also, the inclusion of a male sample would be useful to understand whether their performance improves (stereotype boost) or declines (choking under pressure) when gender stereotypes are activated. Such a study would also provide a direct comparison between the overall driving performance of men and women and, in particular, allow to test whether non-stereotyped women perform equal to or better than men (i.e. make fewer errors).

Overall, the findings suggest that when women are aware that their driving abilities are being evaluated and compared to men, they will make more mistakes. If we look at potential consequences in real life situations, when women feel observed by men while driving or parking, they are likely to experience ST and to underperform, thereby confirming the existing stereotype. Consistent with this interpretation, Chateignier et al. (2011) reported that women fail the driving license test more frequently than men; this probably occurs in part because the testers tend to be men, who may judge women as unskilled, but it may in part also depend on actual underperformance due to ST activation. Also, in real life, when couples or families travel by car it is more likely that men, rather than women, are driving, possibly as a way for women to avoid being judged and for men to show their ability as self-confident male drivers.

Given that the presence of men itself can elicit ST, and that it is practically impossible not to have men observing while women are driving (most drivers are men in normal traffic situations, and men tend to monitor women's driving behavior), women should be trained to fight ST effects in such situations. The ST literature provides a number of strategies how to reduce ST (Schmader and Croft, 2011). One such strategy is to prime a positive identity, but the results of Experiment 2 show that activating the stereotype that young people are better drivers is not sufficient, because it produces the same effects as a neutral control message. An alternative strategy could be to allow women to self-affirm. Self-affirmation has been shown to be an effective means to reduce ST effects in another stereotypically masculine task, namely mental rotation (Martens et al., 2006), in which men typically outperform women. It may be worthwhile to apply the same technique to driving. A further possibility to reduce the effects of ST is to lead women to think of examples of women who drive well (such as top ranking female racing car drivers in the United States) and to identify with these skilled women. Finally it has been shown that, for older drivers, downward social comparison reduces the number of adverse driving events (Levy et al., 2013), a principle that may also apply to women. Future studies should address the efficacy of these ways to alleviate the ST effects on female driving performance.

4.2. Women's expected and self-assessed performance

Considering the performance and the self-assessment data together, as in Table 1, three main conclusions can be drawn. First, women were not particularly optimistic about their driving abilities, given that in all conditions women expected to drive rather poorly. The situation did not change much after performing the driving simulation task. In Experiment 1, the post-test self-assessment scores were even lower than before the test. In Experiment 2, the post-test error estimates decreased only slightly and it did so mainly in the stereotype boost condition where the stereotype activation favored the (young) participants. But, strikingly, even in the stereotype boost condition women provided a post-test assessment of driving mistakes that much exceeded their actual errors. In fact they believed to have made more than twice as many mistakes than they had objectively made during the driving simulation. In other words, women performed well, but were unaware of this.

Stereotypes are rooted in the culture, maintained and transmitted through socializing processes. So one cannot exclude that the country where the study was run has played a role and that the actual level of expectations (a measure of the common-held stereotype) could be different, probably lower in other countries. Italy is known for its strong gender gap. It is ranked 74th according to the 2011 Global Gender Gap Report (Hausmann et al., 2011; see also Nadeau, 2011), and, hence, it remains to be seen whether stereotypes are equally likely to interfere with driving performance of women in highly egalitarian European (e.g., Norway, ranked 2nd) or not European countries (e.g., New Zealand, ranked 5th, South Africa, 12th, or United States, 19th) or, by contrast, in countries in which women have only recently be granted the right to drive cars (e.g., United Arabic Emirates, ranked 103).

This said, however, the critical point is that overall self-assessment did not vary as a function of conditions in our studies. In addition, the experience of driving appeared not to affect them, thus confirming that these beliefs are rooted in the culture and difficult to change, as outlined later in the discussion and in the conclusion. If this study were replicated in other countries where the gender gap is less pronounced, probably the overall self-assessment would be more positive, but we suspect that the lack of influence of condition/instructions on gender beliefs would be maintained. Turning to driving performance, in the absence of international comparison statistics regarding the diffusion of the female driving stereotype it is difficult to advance clear predictions, although we suspect that the stereotype is common in many countries and hence may exert similar effects across countries. This suspicion is, among others, supported by the fact that Yeung and von Hippel (2008) found a stereotype-induced performance decline of similar magnitude in Australia, ranked 23rd.

Second, performance expectancy played a complex role in our studies. On one side expectancy was not affected by our manipulation in either study, suggesting that it played no mediating role in the stereotype-induced performance decline observed here. Yet in line with previous research (Cadinu et al., 2003), under ST women who expected to do poorly showed a decrease in actual performance, whereas expectations were not predictive of actual performance in the stereotype boost condition (Experiment 2). This suggests a detrimental self-fulfilling cycle, in which women who expect to perform poorly confirm this expectancy, but only when the stereotype is made salient.

Third, we investigated whether ST and expectancies affected intrinsic driving motivation, defined as perceived difficulty, safety, satisfaction, and enjoyment. Somewhat surprisingly, stereotype activation did not reduce participant's motivation and enjoyment of the driving experience. However, intrinsic motivation was related both to participants' expectations and to their post-task self-assessment. In both studies, the more mistakes women expected to make and the more mistakes they thought they had made, the lower their intrinsic motivation. Thus, motivation and enjoyment seem to depend more on self-assessment than on actual performance.

Although other research reports that women consider themselves as good drivers overall (Chateignier et al., 2011, conducted in France) and wish to disconfirm the stereotype (Yeung and von Hippel, 2008, conducted in Australia), in Experiment 2, the large majority of women thought that the experimenter expected men to be better drivers than women. Although we did not assess their personal stereotype endorsement directly, women in our studies also thought that they would make a high number of mistakes and to have performed poorly. Interestingly, even when they did perform well, namely in the control or stereotype boost conditions, they still believed to have performed poorly. It seems that the experience of mastering the task, which is usually motivating and provides a sense of competence (Harter, 1978; Bandura, 1997), was not sufficient to modify the stereotyped beliefs of being poor drivers.

This last point is crucial for potential interventions aimed at helping women not only to drive better but also to have more confidence in their driving abilities. Given that the experience of successful driving is not enough to change beliefs, other strategies are warranted. Since stereotypical beliefs are deeply rooted in culture and resist even in the face of contradicting evidence, mass-media messages may provide a feasible tool for changing such beliefs; similarly, precocious interventions during the school years may be helpful. Future studies should assess their effects on both beliefs and driving performance.

5. Conclusions

In conclusion, women are able to drive well, as shown by the performance obtained in the present control and stereotype boost conditions, when they are not reminded of the stereotype concerning women's poor driving skills. However, they fail to recognize this. Despite their good performance in either the control and stereotype boost conditions, they seem to need more than experiencing success to believe that they can succeed. Therefore, interventions specifically aimed at reframing gender-related beliefs may be helpful in order to dismantle stereotypes, question incorrect expectations and self-assessment, and acquire higher driving confidence, above and beyond the acquisition of good driving skills.

References

Aronson, J., Quinn, D.M., Spencer, S.J., 1998. Stereotype threat and the academic underperformance of minorities and women. In: Swim, J.K., Stangor, C. (Eds.), Prejudice: The Target's Perspective. Academic Press, San Diego, CA, pp. 83–103.
Bandura, A., 1997. Self-Efficacy: The Exercise of Control. Freeman, New York.
Beilock, S.L., Rydell, R.J., McConnell, A.R., 2007. Stereotype threat and working memory: mechanisms, alleviation, and spillover. J. Exp. Psychol.: Gen. 136, 256–276.
Bergdahl, J., 2007. Ethnic and gender differences in attitudes toward driving. Soc. Sci. J. 44, 91–97, http://dx.doi.org/10.1016/j.soscij.2006.12.008.

Berger, M.L., 1986. Women drivers! The emergence of folklore and stereotypic opinions concerning feminine automotive behavior. Women's Stud. Int. Forum 9, 257–263, http://dx.doi.org/10.1016/0277-5395(86)90061-0.

Bina, M., Graziano, F., Bonino, S., 2006. Risky driving and lifestyles in adolescence. Accid. Anal. Prev. 38, 472–481, http://dx.doi.org/10.1016/j.aap.2005.11.003.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., Latinotti, S., 2003. Stereotype threat: the effect of expectancy on performance. Eur. J. Soc. Psychol. 33, 267–285.

Cadinu, M., Maass, A., Rosabianca, A., Kiesner, J., 2005. Why do women underperform under stereotype threat? Psychol. Sci. 16 (7), 572–578, http://dx.doi.org/10.1111/j.0956-7976.2005.01577.x.

Chateignier, C., Chekroun, P., Nugier, A., Dutrévis, M., 2011. Femme au volant. . .: effet de la menace du stéréotype et de la colère sur les performances des femmes à une tâche liée à la conduite automobile (Women driving: effects of stereotype threat and of anger on women performance in a driving task). L'Année Psychologique 111 (4), 673–700, http://dx.doi.org/10.4074/S0003503311004039.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, second ed. Lawrence Erlbaum Associates, Erlbaum, NJ.

Croizet, J.C., Claire, T., 1998. Extending the concept of stereotype and threat to social class: the intellectual underperformance of students from low socioeconomic backgrounds. Personal. Soc. Psychol. Bull. 24 (6), 588–594, http://dx.doi.org/10.1177/0146167298246003.

Derks, B., Scheepers, D., VanLaar, C., Ellemers, N., 2011. The threat vs. challenge of car parking for women: how self- and group affirmation affect cardiovascular responses. J. Exp. Soc. Psychol. 47, 178–183, http://dx.doi.org/10.1016/j.jesp.2010.08.016.

Good, C., Aronson, J., Harder, J.A., 2008. Problems in the pipeline: stereotype threat and women's achievement in high-level math courses. J. Appl. Dev. Psychol. 29 (1), 17–28, http://dx.doi.org/10.1016/j.appdev.2007.10.004.

Granié, M.A., Papafava, E., 2011. Gender stereotypes associated with vehicle driving among French preadolescents and adolescents. Transp. Res. Part F 14, 341–353, http://dx.doi.org/10.1016/j.trf.2011.04.002.

Hanna, C.L., Taylor, D.M., Sheppard, M.A., Laflamme, L., 2006. Fatal crashes involving young unlicensed drivers in the US. J. Saf. Res. 37, 385–393, http://dx.doi.org/10.1016/j.jsr.2006.05.007.

Harris, M.B., Miller, K.C., 2000. Gender and perception of gender. Sex Roles 43, 843–863, http://dx.doi.org/10.1023/A:1011036905770.

Harter, S., 1978. Effectance motivation reconsidered: toward a developmental model. Hum. Dev. 21, 34–64.

Hausmann, R., Tyson, L.D., Zahidi, S., 2011. The Global Gender Gap Report, Retrieved from http://reports.weforum.org/global-gender-gap-2011/ (accessed on 17.07.15).

ISTAT (National Institute of Statistics), 2013. Car Accident in Italy, Year 2012, Retrieved from http://www.istat.it/it/files/2013/11/incidenti_stradali_anno-2012_def.pdf (15.04.15).

Joanisse, M., Gagnon, S., Voloaca, M., 2013. The impact of stereotype threat on the simulated driving performance of older drivers. Accid. Anal. Prev. 50, 530–538, http://dx.doi.org/10.1016/j.aap.2012.05.032.

Johns, M., Inzlicht, M., Schmader, T., 2008. Stereotype threat and executive resource depletion: examining the influence of emotion regulation. J. Exp. Psychol.: Gen. 137, 691–705.

Lawrence, C., Richardson, J., 2005. Gender-based judgements of traffic violations: the moderating influence of car type. J. Appl. Soc. Psychol. 35, 1755–1774, http://dx.doi.org/10.1111/j.1559-1816.2005.tb02194.x.

Levy, B.R., Ng, R., Myers, L.M., Marottoli, R.A., 2013. A psychological predictor of elders' driving performance: social-comparisons on the road. J. Appl. Soc. Psychol. 43, 556–561, http://dx.doi.org/10.1111/j.1559-1816.2013.01035.x.

Lonczak, H.S., Neighbors, C., Donovan, D.M., 2007. Predicting risky and angry driving as a function of gender. Accid. Anal. Prev. 39, 536–545, http://dx.doi.org/10.1016/j.aap.2006.09.010.

Martens, A., Johns, M., Greenberg, J., Schimel, J., 2006. Combating stereotype threat: the effect of self-affirmation on women's intellectual performance. J. Exp. Soc. Psychol. 42, 236–243, http://dx.doi.org/10.1016/j.jesp.2005.04.010.

Mast, M.S., Sieverding, M., Esslen, M., Graber, K., Jäncke, L., 2008. Masculinity causes speeding in young men. Accid. Anal. Prev. 40, 840–842, http://dx.doi.org/10.1016/j.aap.2007.09.028.

Moè, A., 2009. Are males always better than females in mental rotation? Exploring a gender belief explanation. Learn. Individ. Differ. 19 (1), 21–27, http://dx.doi.org/10.1016/j.lindif.2008.02.002.

Moè, A., Pazzaglia, F., 2006. Following the instructions! Effects of gender beliefs in mental rotation. Learn. Individ. Differ. 16, 369–377, http://dx.doi.org/10.1016/j.lindif.2007.01.002.

Nadeau, B.L., 2011. Not Just Any Old Charlatan, Retrieved from http://www.thedailybeast.com/contributors/barbie-latza-nadeau.42.html (17.07.15).

NHTSA (National Highway Traffic Safety Administration), 2012. Fatality Analysis Reporting System, Retrieved from http://www.iihs.org/iihs/topics/t/general-statistics/fatalityfacts/gender (15.04.15).

Osborne, J.W., 2007. Linking stereotype threat and anxiety. Educ. Psychol. 27, 135–154, http://dx.doi.org/10.1080/01443410601069929.

Özkan, T., Lajunen, T., 2006. What causes the differences in driving between young men and women? The effects of gender roles and sex on young drivers' driving behaviour and self-assessment of skills. Transp. Res. Part F: Traffic Psychol. Behav. 9, 269–277, http://dx.doi.org/10.1016/j.trf.2006.01.005.

Schmader, T., Croft, A., 2011. How stereotypes stifle performance potential. Soc. Personal. Psychol. Compass 5 (10), 792–806, http://dx.doi.org/10.1111/j.1751-9004.2011.00390.

Schmader, T., Johns, M., 2003. Converging evidence that stereotype threat reduces working memory capacity. J. Personal. Soc. Psychol. 85 (3), 440–452.

Sibley, C.G., Harré, N., 2009. A gender role socialization model of explicit and implicit biases in driving self-enhancement. Transp. Res. Part F: Traffic Psychol. Behav. 12, 452–461, http://dx.doi.org/10.1016/j.trf.2009.08.006.

Steele, C.M., 1997. A threat in the air: how stereotypes shape intellectual identity and performance. Am. Psychol. 52, 613–629.

Steele, C.M., Aronson, J., 1995. Stereotype threat and the intellectual test performance of African Americans. J. Personal. Soc. Psychol. 69, 797–811.

Tronsmoen, T., 2008. Associations between self-assessment of driving ability, driver training and crash involvement among young drivers. Transp. Res. Part F: Traffic Psychol. Behav. 11, 334–346, http://dx.doi.org/10.1016/j.trf.2008.02.002.

Wolf, C.C., Ocklenburg, S., Ören, B., Becker, C., Hofstätter, A., Bös, C., Popken, M., Thorstensen, T., Güntürkün, O., 2010. Sex differences in parking are affected by biological and social factors. Psychol. Res. 74, 429–435, http://dx.doi.org/10.1016/j.lindif.2008.02.002.

Yagil, D., 1998. Gender and age-related differences in attitudes toward traffic laws and traffic violations. Transp. Res. Part F: Traffic Psychol. Behav. 1, 123–135.

Yeung, N.C.J., von Hippel, C., 2008. Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. Accid. Anal. Prev. 40 (2), 667–674, http://dx.doi.org/10.1016/j.aap.2007.09.003.