The effects of non-evaluative feedback on drivers’ self-evaluation and performance

Ebru Dogan a,*, Linda Steg a, Patricia Delhomme b, Talib Rothengatter 1

a Faculty of Behavioral and Social Sciences, University of Groningen, The Netherlands
b IFSTTAR, Laboratory of Driver Psychology, France

ARTICLE INFO

Article history:
Received 25 May 2011
Received in revised form 2 August 2011
Accepted 4 September 2011

Keywords:
Non-evaluative feedback
Self-evaluation
Self-regulation
Hazard perception
Self-enhancement bias

ABSTRACT

Drivers’ tend to overestimate their competences, which may result in risk taking behavior. Providing drivers with feedback has been suggested as one of the solutions to overcome drivers’ inaccurate self-evaluations. In practice, many tests and driving simulators provide drivers with non-evaluative feedback, which conveys information on the level of performance but not on what caused the performance. Is this type of feedback indeed effective in reducing self-enhancement biases? The current study aimed to investigate the effect of non-evaluative performance feedback on drivers’ self-evaluations using a computerized hazard perception test. A between-subjects design was used with one group receiving feedback on performance in the hazard perception test while the other group not receiving any feedback. The results indicated that drivers had a robust self-enhancement bias in their self-evaluations regardless of the presence of performance feedback and that they systematically estimated their performance to be higher than they actually achieved in the test. Furthermore, they devalued the credibility of the test instead of adjusting their self-evaluations in order to cope with the negative feelings following the failure feedback. We discuss the theoretical and practical implications of these counterproductive effects of non-evaluative feedback.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

We are motivated to see ourselves in a positive way in order to feel good about ourselves and to maintain a high self-esteem (Steele, 1988). This applies to drivers as well. Drivers very often believe that they drive better than other drivers or that they are more competent than they actually are, showing a self-enhancement bias in their self-evaluations (see Sundström, 2008). Generally, drivers consider themselves to be more skillful than other drivers (Svenson, 1981; Depoy, 1989; Delhomme, 1991; Gregersen, 1996; McKenna et al., 1991; Groeger and Grande, 1996), indicating that at least some of them overestimate their skills. Different motivational explanations have been offered for the mechanisms underlying the self-enhancement bias in drivers’ skill evaluations. McKenna et al. (1991) suggested that drivers inflate their own abilities instead of deflating those of other drivers. Walton (1999), on the other hand, found that truck drivers downgraded other drivers’ abilities rather than inflating their own abilities. Whichever motivational mechanism explains self-enhancement biases, such biases seem to be persistent for driving skills. In fact, this self-enhancement bias has been found to be even stronger when measured implicitly (Harré and Sibley, 2007), suggesting that drivers’ beliefs about the superiority of their driving competence are deeply rooted. Paradoxically, people also believed that they are less susceptible to judgmental biases than others (Pronin et al., 2004), which makes these biases even more robust.

The overestimation of skills and competence is associated with perceiving less risks, either by perceiving one’s self as a less risky driver (Svenson, 1981) or by perceiving one’s own crash risk as lower (Depoy, 1989; Deery, 1999; Harré and Sibley, 2007). Drivers generally take regulatory actions when they perceive that their competence falls short to meet the demands of the situation (Fuller, 2008). When drivers overestimate their competence, they may expect their performance to be better than it really is. Consequently, when drivers overestimate their skills and underestimate the risks involved, they may be more likely to take risks on the road, for instance, by driving faster. This leaves shorter time margins to detect hazardous situations in time, which in turn may hinder one’s ability to respond timely to dangers as to avoid negative consequences. It is therefore of great importance that drivers have accurate estimations of their competence and abilities (see Rothengatter, 2002).

* Corresponding author at: Faculty of Behavioral and Social Sciences, Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. Tel.: +31 50 3634722; fax: +31 50 3636304.
E-mail address: e.b.dogan@rug.nl (E. Dogan).
1 Talib Rothengatter passed away during the course of this work.

0001-4575/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
Kruger and Dunning (1999) suggested that inaccurate self-evaluations of competence, either overestimation or underestimation, is due to lack of metacognition about one's skills and competence. They suggest four possible feedback-related reasons for inaccurate self-evaluations: lack of feedback, attributing failure feedback to some other causes than lack of skills, not understanding why failure occurred, and not receiving self-corrective information. The driving task is subject to all the aforementioned deficiency factors because of the lack of systematic feedback in and the forgiving nature of the traffic environment for errors. Drivers may not develop a realistic representation of their abilities and competence because not every error or violation made while driving results in adverse consequences such as accidents, near accidents, or penalties, which implies that drivers do not receive explicit feedback on their performance. This is particularly problematic for learner and novice drivers because they are more in need of feedback in order to comprehend the effects of their behaviors on other road users, the road environment, what mistakes they do, and how to avoid such mistakes. Feedback from an instructor or from the environment may enable drivers to develop a sense of possible situations that they may encounter in the traffic environment and their abilities or lack thereof to deal with different traffic situations (Groeger, 2000; Kuiken and Twisk, 2001; Hatakka et al., 2002).

Kuiken and Twisk (2001) emphasized the importance of feedback for a safe calibration (i.e. self-regulation) of skills and driving task demands. In line with the self-regulation theory, they propose that adequate self-assessment of skills is crucial for a safe calibration of driving skills. They propose that provision of comprehensive feedback, by providing information on the way the task was performed and how it could be improved, is needed to enhance safe regulation of driving behavior because it enables driver to safely match their capabilities with the task demands. A safe match between the capabilities and task demands reflects on driver's goal setting at various stages of the driving task from route choice to the actions taken behind the wheel (cf. Roghengatter, 2002). Similarly, Hatakka et al. (2002) suggest that self-evaluation of one's driving skills should be integrated in the driver training in order to develop learner drivers' metacognitive skills for specific tasks of driving such as vehicle control or hazard perception, and that this can be realized by providing drivers with feedback on their performance. Such training is expected to promote learner drivers' self-regulatory behaviors in different road situations and task demands (Kuiken and Twisk, 2001; Hatakka et al., 2002).

In more and more European countries structured feedback that focuses on higher order safety skills and self-assessment of them are integrated in the driver training as part of the driver licensing systems (Twisk and Stacey, 2007), with promising effects in the short term. Research revealed that after this training, learners' assessments of their skills were positively correlated with their trainers' assessments of the same skills, suggesting that learners assessed their skills accurately (Bocca et al., 2011; Mynttinen et al., 2009a,b). The long-term effects of this training have not been studied yet, i.e. it is not clear whether accurate self-assessments observed during the training are retained after the training and whether the training indeed results in less risk taking behavior and accidents.

In the meantime, non-evaluative feedback is increasingly adopted in traffic for training purposes as well. This type of feedback is less comprehensive since typically, information is provided on actual performance levels only. Examples are the increased use of simulators and computer-based tests such as hazard perception tests, which provide non-evaluative feedback on one's driving skills. In essence, people taking these tests learn their absolute scores on a test or their scores relative to other test-takers on particular skills, but do not receive any information on why their score was low or high or on how scores may be improved. Despite being frequently used in driver training we do know little about the effectiveness of non-evaluative performance feedback as given in these instruments.

Research on air traffic control indicates that non-evaluative feedback on performance may be effective in promoting accurate self-evaluations (e.g. Mitchell et al., 1994). Mitchell et al. (1994) used a computerized test to simulate an air traffic controller's task which is a complex rule-driven task requiring participants to learn various rules about safe and efficient landing conditions. Participants received two sorts of non-evaluative feedback: a running feedback score on their performance after each landing and an overall performance score. Mitchell and colleagues found a strong positive correlation between the expected and actual performance scores of participants, suggesting that participants had an accurate view of their performance. Also, the relationship became stronger at the later trials, suggesting that the feedback enabled participants to further improve their self-evaluations in subsequent trials. Participants used two different strategies of self-regulation, that is, the non-evaluative feedback led to an adjustment of either their actual performance or their expected performance score. This suggests that the non-evaluative feedback resulted in a more accurate self-evaluation of performance, which improved the self-regulation of participants' expected performance throughout the skill acquisition. Could such feedback on performance be beneficial in overcoming the self-enhancement biases for certain driving skills related to drivers' hazard perception as well? Or is non-evaluative feedback not effective or even counterproductive because, for instance, such feedback does not provide any information on how people can improve their performance? In the current research, we will address this question via a hazard perception test.

Hazard perception is a higher-order safety skill which is used to anticipate the road environment and behavior of other road users (Horswill and McKenna, 2004). Specifically, hazard perception skills involve estimating what threats are present in the environment, as well as knowing what to do in order to avoid and handle those threats. Thus, hazard perception skills cover detection and anticipation of threats as well as one's assessment of abilities to handle those threats (Grayson et al., 2003). While hazard perception skills improve as drivers gain experience, hazard perception does not become automated, but rather becomes a less effortful process with practice (McKenna and Farrand, 1999 as cited in Horswill and McKenna, 2004). Therefore, drivers need to pay attention to information from constantly evolving situations and frequently take action in order to handle dangers safely and in time. As we have mentioned earlier, drivers' self-regulatory behaviors to avoid hazards may be influenced by overestimation of their competence. This is particularly the case among novice drivers because their higher order safety skills (such as hazard perception skills) to handle relatively complex traffic situations have probably not sufficiently developed yet (OECD-ECMT, 2006). Accurate self-evaluations in a hazard perception task are particularly important because computerized hazard perception tests are integrated as part of licensing system in several countries including the United Kingdom and the Netherlands. What happens when drivers receive non-evaluative feedback telling them that they are in fact not as good as they think they are, and learn that they are overestimating their competence and performance?

The perceived discrepancy between what drivers actually can do and what they believe they can do is assumed to trigger self-regulatory behaviors (cf. Carver and Scheier, 1998; Fuller, 2008). Specifically, feedback may elicit self-regulation by enabling a comparison between the expected and actual situation, and consequently making people aware of any discrepancy or balance between the expected and actual situation (Cervone and Wood, 1995; Carver and Scheier, 1998). An adaptive response to deal with a discrepancy would be to adjust the effort put in the task and try
to do better or to adjust the expectations about one's performance level, which would result in a closer match between one's expectations and reality. A non-adaptive response to deal with such a discrepancy, on the other hand, would imply not making these adjustments in effort or expectations. This could result in negligence of the feedback or detachment from the task goals such as disengaging from the effort and quitting the task after a few trials, or devaluation of the task if detachment is not possible (cf. Carver and Scheier, 1998; Kuiken and Twisk, 2001). Thus, there may be occasions where non-evaluative feedback does not promote adaptive self-regulation and thus does not reduce the self-enhancement biases. In the current study, we investigated the effect of non-evaluative performance feedback on self-evaluations and actual performance, and whether non-evaluative feedback reduces the self-enhancement bias among young, novice drivers.

We first examined the accuracy of driver's self-evaluations of their hazard perception skills via a computerized test. The performance criterion was the total score obtained at the end of the hazard perception test. Additionally, we examined the effect of non-evaluative feedback on these self-evaluations and on actual performance in subsequent trials of the test. In line with previous studies, we expected the majority of the participants to overestimate their expected performance in the hazard perception test initially in the first trial before they received feedback (hypothesis 1). This implies that we expect that the majority of the participants would receive negative feedback on their performance, and learn that they perform worse than expected. Based on the studies discussed above, we propose two competing hypotheses for the effects of this negative feedback on self-regulation in later trials. First, the negative feedback can elicit adaptive responses, in which participants adjust their self-evaluations or their performance2 in the subsequent trials in accordance with the negative feedback in the former trials. This implies that negative feedback results in a lower estimated test score in the subsequent trials compared to the estimated test score in the first trial, and a closer fit between estimated and actual performance in the subsequent trials (hypotheses 2). Alternatively, negative feedback may result in non-adaptive regulatory responses and not change their performance or their estimations of their performance. If participants do not adjust their performance estimations or their actual performance, the discrepancy between the estimated and actual performance will remain. In line with the self-regulation process (Carver and Scheier, 1998; Kuiken and Twisk, 2001), we expect this unresolved discrepancy between expected and actual performance to reflect on participants’ performance evaluations and task evaluations (hypothesis 3). Specifically, we expect that the feedback results in a more negative evaluation of one’s performance in previous trials, because the feedback makes participants aware of the discrepancy between their expected and actual performance. Also, we expect that the negative feedback will elicit negative feelings. More importantly, we expect that participants will try to restore a positive self image and feel good about themselves by devaluing the test.

2. Method

2.1. Participants

We tested our hypotheses in an experimental study. The participants were 36 students (11 male, 25 female) from the University of Groningen who held a valid drivers’ license for at least one year (M = 2.5 years; SD = 1.17) and had driven an average of 6505 km since licensure (SD = 4820). The mean age was 21 years (M = 21.22, SD = 1.44); age ranged from 19 years to 24 years. Participants were recruited via the student participant pool of the University of Groningen and received course credit in return for their participation.

2.2. Hazard perception test

To test our hypotheses, we developed a hazard perception test comprising natural traffic scenes that were recorded around the city of Groningen, the Netherlands. The recordings were taken from the drivers’ point of view during daylight in bright and dry weather conditions. A team of experts watched the recordings to mark the hazardous events following the definition and criteria set out by Grayson and Sexton (2002), who argued that a good hazard perception measure should capture respondent's scanning skills and anticipation of the developing situation rather than detecting only a quick reaction to the situation. Based on this, clips that would detect drivers’ anticipation of developing dangers were kept in, while the ones that would create difference only with respect to reaction time to suddenly occurring hazards were left out. The selected clips were then tested in a pilot study. At this stage, we applied two other selection criteria proposed by Grayson and Sexton (2002), namely a mean score of the hazard event between 2 and 3 for participants (the scoring system is explained below), and significantly different hazard perception scores between experienced and inexperienced participants. Then, all these clips were re-evaluated by the same team of experts. Five of the clips were omitted at this stage because the hazards in those clips did not differentiate between the experienced and inexperienced groups or they were not detected. The remaining 36 clips were used for the test, after the start and end scenes of the hazard situations were adjusted based on the reaction times obtained in the pilot test.

The shortest clip was 22 s while the longest one was 69 s. The number of hazards in each clip ranged from one to three. Examples of the hazards included were a car emerging from the right, pedestrians getting out of parked cars, a lane reduction in a construction area, and a cyclist crossing the road. We developed three different versions of the hazard perception test to be used in three trials, all of which had 12 comparable hazard situations, and took approximately 10 min to complete.

The clips were shown to participants on a 19” computer screen with a 3-s interval between the clips. Participants used an external button to respond to the hazard situations. They were instructed to press the button as soon as they thought they should modify their behavior to avoid a potential danger. Participants were informed that their response would not be valid should they press the button more than five times for each hazard situation.

The hazard perception test enabled us to provide participants' with feedback in terms of a non-evaluative test scores. To do so, responses were scored following the method developed by Grayson and Sexton (2002), that is, the starting and ending frames of the hazard situations were marked and the time range was divided into 5 equal intervals. As different hazards had different durations and require different response times, the time interval was idiosyncratic for every item. The closer the response to the starting frame of the hazard situation, the higher the points gained for each hazard (5 points for the fastest response and 1 point for the slowest response). We recorded the time frame in which participants responded. The maximum possible score for each version of the hazard perception test was 60 meaning that the participant detected each hazard and reacted to all hazards in a very short time, while the minimum score was 0 meaning that the participant failed to detect any hazard in time.

3 Although theoretically it is plausible to expect an adjustment in performance, we think this it is not likely that participants improve their performance because the non-evaluative feedback does not convey information about what they did wrong or what they should do differently to increase their performance.
2.3. Measures

2.3.1. Self-evaluations for the expected performance in the hazard perception test

We measured self-evaluations of the expected performance in the hazard perception test based on Bandura’s (1997) self-efficacy scale, because the structure of the scale enabled us to measure expected performance for different levels, i.e. scores, of the task. Participants were presented a table with possible scores on the hazard perception test, ranging from 20 to 60 with 5-point intervals (9 levels in total). They indicated whether or not they could reach any of the given test scores. The number of level that participants reported that they could execute was summed and used as an indicator of the estimated performance score. For instance, if the participant reported he or she could reach the test scores mentioned in the first four levels, the estimated performance would be 35 in the hazard perception test. Scores on self-evaluations for the estimated performance could range from 20 (i.e. I can detect few hazards in time) to 60 (i.e. I can detect all the hazards present in the video clips in time).

2.3.2. Experience of correspondence between estimated and actual performance

After completing each trial, we measured how participants experienced their test performance. Specifically, we asked participants how successful they felt about their performance in the hazard perception test, and how satisfied and frustrated they were with their performance, respectively. Participants in the experimental group answered these questions after receiving their true performance score, whereas participants in the control group did not receive this feedback. Additionally, participants’ rated how effortful it was for them to perform at the level they did. Next, the experimental group received a final question asking whether they thought their test score reflected their true performance. This question was not presented to the control group as they did not receive feedback on their test score. Participants answered these four questions on a 7-point scale ranging from 1 – not at all to 7 – very much.

2.4. Procedure and design

The experiment consisted of three trials of the computerized hazard perception test. This enabled us to examine the effects of the non-evaluative feedback on self-evaluations over time. In each trial, participants completed a different version of the hazard perception test. The three versions of the hazard perception test were counterbalanced. Before starting the experiment, participants were given the instructions on the hazard perception test identical to those Grayson and Sexton (2002). Thus, a hazard was defined as “something that a driver should keep an eye on because it could lead to an accident situation” (Grayson and Sexton, 2002, p. 6). Participants were informed that they should press the external button as soon as they recognize a hazard developing. We also mentioned that higher scores on this test would mean safer reactions to a hazard while lower scores would mean unsafe reactions. Afterwards, participants received a brief training on the hazard perception test with three sample clips. During the training we explained how the test worked and how their responses were scored. They were allowed to repeat the training if they needed to. During the training, all participants received non-evaluative feedback on their scores following each clip and then received a total test score once all the clips in the relevant test had been shown. This was done to inform participants about their performance. After the training, participants filled in a questionnaire. The questionnaire included a scale for their self-evaluations of their performance in the hazard perception test in the coming trial as well as a few demographics questions.

Participants were then randomly assigned to the experimental or control group. Participants in the experimental group received performance feedback on their score on the hazard perception task similar to the feedback they received during the training, while participants in the control group did not longer receive feedback. Running feedback was presented after each video clip that informed participants about the score obtained for the particular clip. At the end of the hazard perception test, overall performance feedback was provided reflecting the total score obtained at the end of the test. The feedback (i.e. experimental group, E) and control (i.e. C) groups were similar in license duration, \( t(34) = -0.52, p = 0.610, M_E = 2.47, SD_E = 1.23; M_C = 2.68, SD_C = 1.13 \) and total mileage, \( t(34) = 1.60, p = .117, (M_E = 7700, SD_E = 5095; M_C = 5170, SD_C = 4246) \). We also checked whether the two groups were similar in their performance estimations at the beginning of the study. As expected, the experimental and control groups did not differ in terms of their performance estimations in the first trial, \( F(1, 34) = 2.68, p = .111 \).

Next, participants performed the first hazard perception test and either did or did not receive performance feedback depending on the group in which they were. After completing the first test (for the experimental group after learning their test score), all participants filled in a short scale measuring their experience of the contingency or the discrepancy between their performance and estimated performance in the previous trial. Next, they filled in the self-evaluation scale for their estimated performance in the subsequent trial. The same procedure was repeated after the second and third trials, except that the self-evaluation scale was not administered after the third trial because there were no more trials.

3. Results

In order to test our first hypothesis on the accuracy of the performance estimations on the hazard perception test, we compared the estimated and actual test scores for each trial via paired sample t-test analysis.3 As expected, performance estimations for the first trial and the actual test scores on this trial significantly differed: participants in both groups overestimated their hazard perception test score \( t(18) = -7.88, p < .001 \) for the experimental group and \( t(18) = -6.53, p < .001 \) for the control group, see Fig. 1.

Our second and third hypotheses predicted competing regulatory responses for the discrepancy between the estimated and actual performance. The second hypothesis predicted an adaptive regulatory response, in which the experimental group would adjust their performance expectations in the subsequent trials on the basis of the feedback they received. We expected the majority of the participants in the experimental group to receive negative feedback. Indeed, no participant in the experimental group received positive feedback, i.e. no participant performed better than their estimated performance. On the contrary, all participants received negative feedback, i.e. they performed worse than they expected. Fig. 1 shows that participants in both the experimental and control groups maintained their overestimated performance estimations in all three trials. Indeed, a mixed model Anova with group as a between factor and trial as a within factor showed that the interaction between group and trial did not have a significant effect on the performance estimations, \( F(2, 68) = 3.1, p = .037 \), or on the actual performance, \( F(1, 68) = .40, p = .670 \). For both the experimental and control group, neither estimated nor actual performance

---

3 We computed the difference score between the estimated and actual performance. The difference score was used as the dependent variable in a mixed ANOVA with group as the between measure and trial as the within measure. This analysis revealed no significant main and interaction effects indicating that the discrepancy between the estimated and actual performance was not different between the experimental and control groups or across trials.
changed over time. Thus, contrary to the second hypothesis, non-evaluative feedback did not result in adaptive self-regulatory responses.

So, we found that the participants showed non-adaptive regulatory responses. How does this reflect on the way they experienced their performance? How do they cope with the discrepancy between their estimated and actual performance? The third hypothesis was about the way participants felt about their performance in the previous trials in case they did not adjust their estimated or actual performance. Since the control group did not receive the performance feedback, we expected the experimental group, who was made aware of a discrepancy between their expected and actual performance, to be less satisfied and more frustrated about their performance and to feel less successful than the control group. We conducted independent sample t-test analysis to compare the satisfaction, frustration, and the perceived success of the performance level in the previous trials for the experimental and control group. After the first trial, the participants in the experimental group were significantly less satisfied (t (34) = −3.94, p < .001) and marginally more frustrated (t (34) = 1.80, p = .080) with their performance compared to the control group participants who had not received feedback on their performance on the hazard perception test in the previous trials. Similar results were observed for the responses following the second and the third trial (see Fig. 2). In addition, the control group participants considered themselves to be more successful in each preceding trial than the experimental group participants did after the first trial, t (34) = −4.58, p < .001. An independent sample t-test revealed that perceived effort did not differ between the two groups in any trial, t (34) = −.69, p > .10 in the first trial, t (34) =.56, p > .10 in the second trial, and t (34) = .44, p > .10 in the third trial. So the test was perceived as equally effortful by both groups, and differences in experienced effort cannot explain our results.

In line with our third hypothesis, the maladaptive self-regulatory responses resulted in devaluing the hazard perception test. In all three trials, participants in the experimental group believed that their test scores did not reflect their actual performance well (M1 = 2.37, SD1 = 1.11; M2 = 2.21, SD2 = 0.98; M3 = 2.42, SD3 = 1.07). So, in line with our hypothesis, the experimental group downgraded the credibility of the test, presumably to handle the discrepancy between their estimated and actual performance.

4. Discussion

Drivers' self-serving evaluations of their driving competence and abilities may negatively affect their risk taking behavior and road safety. Self-enhancement biases have been found for various driving skills and appear to be robust. We first tested whether the self-enhancement bias was also apparent in hazard perception. In line with our expectation, we found a strong self-enhancement bias: participants systematically overestimated their hazard perception skills and thought they would perform better than they actually did.

Lack of feedback may be one of the reasons why people do not have accurate beliefs about their skills and competence. In the current study we investigated whether inaccurate self-evaluations can be overcome by giving participants' non-evaluative performance feedback. More specifically, we tested whether non-evaluative feedback would result in adjustment of the self-evaluation of one's performance. We did not expect the non-evaluative feedback to result in changes in one's performance, as the feedback did not indicate why participants’ performance was lower than they expected (so participants’ did not learn how to improve their performance), and indeed, actual performance did not change over time. Interestingly, we did not find any effects of the non-evaluative feedback on one's self-evaluations in subsequent trials. On the contrary, our
results indicate a profound self-enhancement effect in terms of performance estimations among both the feedback (experimental) and no-feedback (control) groups. Furthermore, although the test was perceived as similarly effortful by both groups, the experimental group who learned about the discrepancy between their expected and actual performance experienced a higher level of frustration and feeling of failure, and a lower level of satisfaction with their performance than the control group who did not explicitly learn about this discrepancy. Thus, the experimental group experienced negative feelings about their performance, probably as a result of the discrepancy between their estimated and actual performance. Nonetheless, they did not respond to this discrepancy in an adaptive way by adjusting their estimated performance scores or actual performance. Rather, they downgraded the credibility of the hazard perception test by indicating that they believed that their test scores did not reflect their true performance, supporting our third hypothesis.

Why did the experimental group participants, who were explicitly informed about their scores on the hazard perception test and therefore understood that they failed to perform at the level they expected, keep providing higher performance estimations than they actually achieved? Why did they change their performance assessments on the basis of the feedback on their test score but rather disqualified the test?

One explanation is related to the nature of hazard perception skills. Hazard perception requires integration of several driving skills, which makes it rather ambiguous to assess one’s performance on hazard perception skills. Indeed, when a self-evaluation dimension is ambiguous, it may result in self-enhancement biases per se (Dunning et al., 1989; Ackerman et al., 2002). A recent study demonstrated that somewhat experienced young drivers showed self-enhancement biases for their skills and risk assessments even when they were told that their driving skills would be assessed by objective measures, namely a driving simulator (White et al., 2011). Furthermore, White et al. (2011) argued that computerized tests may not represent the true difficulty of the task involved in various driving situations. Therefore, it is plausible to expect that participants underestimated the difficulty of the computerized task and overestimated their performance to some extent, particularly in the initial trial. We expected this initial self-enhancement effect to diminish in the subsequent trials for the experimental group. However, the robustness of the self-enhancement bias observed in the current study in all three trials study calls for further explanation.

The number of hazards detected in all trials, although participants did not know this, was rather high (around 9 out of 12 in each trial), especially considering the experience level of our sample. So, lower-than-expected scores must be due to reaction time latency. Thus, either the importance of reaction time latency was not clear to the participants, despite the information provided beforehand and the training with the initial clips, or participants could not accurately estimate the timing of their reactions to hazardous situations. In fact, Chapman and Underwood (1998) found that novice drivers had a longer fixation time than experienced drivers. Furthermore, inexperienced drivers had less accurate estimations of time-to-collision (Cavallo and Laurent, 1988). So, it is not surprising that our participants, who are relatively inexperienced drivers, had a higher reaction time latency, despite detecting a rather high number of hazards.

Another explanation to address the robust self-enhancement bias may be that the non-evaluative feedback on the test performance was not effective in changing drivers’ self-evaluations and actual performance because it did not convey information on why participants’ scores were lower than they expected. Thus, non-evaluative feedback was not specific to elicit an adaptive self-regulation (Bandura, 1997). Not knowing why their expectations were not accurate may not be sufficiently motivating for drivers to change the self-assessments of their competences, but it apparently did induce negative feelings (such as higher frustration and dissatisfaction) among the experimental group that participants could not ignore. Thus, affective self-regulatory process was activated (Carver and Scheier, 1998). When faced with negative feedback without knowing the exact cause of it, the feedback group devalued the task by degrading its credibility instead of adjusting their estimated performance. This may be because they could not disengage from the task due to their commitment to complete the task, despite being frustrated and dissatisfied with their performance (Carver and Scheier, 1998; Kuijen and Twisk, 2001). They may have dealt with these unpleasant feelings due to the disconfirming feedback by downgrading the credibility of the test. Furthermore, non-evaluative feedback of this kind might have created differences between the experimental and control group in how they felt about their performance by triggering two mechanisms: creating a threat to participants’ self-view and giving a competitive nature to the hazard perception test. These points will be further elaborated below.

The feedback made participants’ aware of the discrepancy between their expected and actual performance, which may have threatened their self-perception. We expected this to be the case if participants would not adjust their self-evaluations or performance, that is, we expected that negative feedback would yield negative feelings and external attributions. Apparently, participants coped with these negative feelings by downplaying the credibility of the task (cf. Carver and Scheier, 1998; Kuijen and Twisk, 2001). Hepper et al. (2010) argued that defensiveness is triggered by a threat to one’s self-concept and such defensiveness is notable by attributing failure to the situation or the task rather than to one’s ability. Research showing the strength of implicit self-enhancement bias among drivers indicates that driving abilities are core for drivers’ self-concept (Harré and Sibley, 2007). Indeed, research indicates that threatening information is especially processed defensively if the information is related to issues that are important for individuals (Pietersma and Dijkstra, 2011). Therefore, non-evaluative feedback may have prompted a self-threat among the feedback group.

Considering the fact that our sample consisted of university students, who are presumably very much used to performance evaluations via grades, the feedback we used might have added a competitive nature to an ordinary risk perception task. Delhomme and Meyer (1998) showed that failure in a speed regulation task resulted in a worse speed regulation among novice drivers especially when the task was presented as a competitive one than a cooperative task. People exhibit a self-serving bias by making external attributions to temporary and specific sources for their failures in order to protect their ego and self-image. However, discounting own responsibility for negative outcomes weakens the motivation and ability to take necessary regulatory actions to change those outcomes (see Mezulis et al., 2004). Therefore, the type of feedback and how it should be given are crucial especially for inexperienced drivers, who are more likely to be influenced by external motivational factors (Delhomme and Meyer, 1998). Using computerized tests or simulators for training driving skills may backfire if trainees do not learn the consequences of their behaviors, and do not learn why their performance was high or low, and how to improve their performance.

In the current study, we were mainly interested in the effect of non-evaluative feedback on self-enhancement biases. In doing so, we focused on participants’ estimated performance rather than their performance goals. It is likely that actual performance is not only related to one’s estimated performance, but also to one’s goals. We did not measure participants’ goals related to their performance in the hazard perception task; and consequently, we do not know whether participants’ motivation to not adjust their estimated or
actual performance was influenced by their goals. Furthermore, correspondence between specificity of goals and feedback, and the usefulness of feedback for self-regulation are interrelated (Cervone and Wood, 1995). Future research could incorporate goal setting and self-evaluations in order to disentangle their effect on self-regulation. This could improve our understanding of the way affect influences regulatory responses to feedback and how drivers cope with negative feedback.

In conclusion, our results suggest that non-evaluative performance feedback may not be suitable for improving self-evaluations and performance that are measuring (complex) skills because it can alter the way the task is perceived. Most importantly, non-evaluative feedback may not change self-evaluations or performance but rather result in devaluation of the task. Non-evaluative feedback was effective in regulating self-evaluations and expected scores in an air traffic controller task (Mitchell et al., 1994), but our results suggest it is not effective for a driving task. This might be due to the differences between driving tasks and air traffic controller tasks in terms of the amount of information available to the operators and the level of automation assisting the operator. Lack of feedback contributes to self-enhancement biases; nonetheless, the mere presence of feedback is not sufficient for overcoming such biases for driving competence. The content of the feedback seems to be more crucial than the presence of feedback (Kruger and Dunning, 1999). We think that providing drivers with detailed information on what caused their failure or success and what they should do in order to improve their performance is needed for feedback to be effective in reducing self-enhancement biases.

References


