Title

Information, Assessment, or Decision: A Driving Simulator Study on the Effect of Real-Time Feedback Based on Information-Processing Stages

Authors

Angèle Picco [a,*], Arjan Stuiver [a], Joost de Winter [b], Dick de Waard [a]

[a] Faculty of Behavioural and Social Sciences, University of Groningen, the Netherlands

[b] Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, the Netherlands

*Corresponding author. *E-mail address:* a.picco@rug.nl

ORCiDs:
Angèle Picco 0000-0001-5155-1062
Arjan Stuiver 0000-0003-3823-0242
Joost de Winter 0000-0002-1281-8200
Dick de Waard 0000-0003-0262-4573

Funding details

This research is funded by Transitions and Behaviour grant 403.19.243 ("Towards Safe Mobility for All: A Data-Driven Approach"), provided by the Netherlands Organisation for Scientific Research (NWO).

Disclosure statement

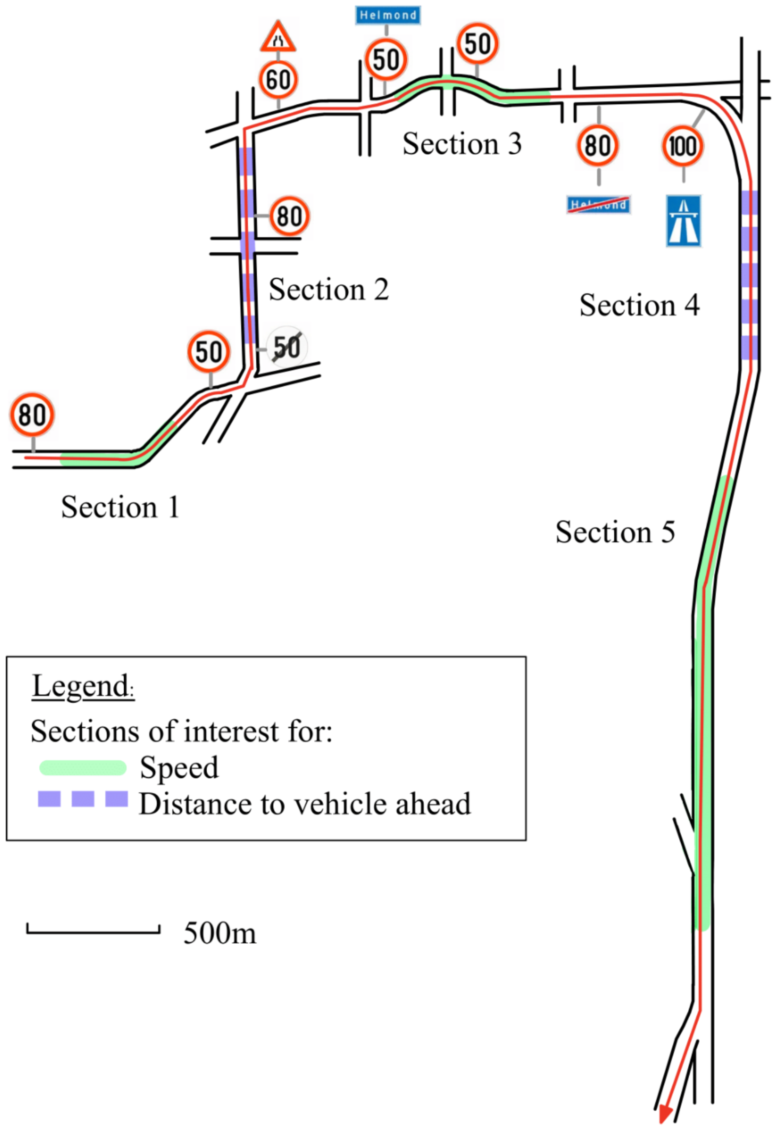
The authors report there are no competing interests to declare.

populated with traffic: oncoming traffic was created pseudo-randomly during the whole experiment and specifically at intersections. In some segments, there was also a car driving in front of the ego-vehicle, with the purpose of performing a car-following task and thus being able to determine time headway (THW). Traffic was programmed to drive adhering to speed limits for the whole drive, except on the motorway where traffic was less homogeneous, with some cars driving slower and others faster than the speed limit. On average, participants completed the drive in about 10 minutes.

This route is illustrated in Figure 1. The figure also depicts sections of interest: while the participants' behaviours were recorded throughout the whole drive, particular sections were identified and aimed at analysing behaviours depending on the context, as described in Section 2.4.1.

Figure 1

Map of the Experiment's Route



2.2.3 Conditions

A within-subjects design with three experimental conditions, Information, Assessment and Decision, and one control condition was used. In all experimental conditions, feedback was provided, differing in the stage of information processed. In the control condition, no feedback was given. The feedback pertained to the participants' speed and the distance between their vehicle and the vehicle in front of them. For this purpose, a "feedback area" was positioned in the right-hand section of the dashboard, in the lower part of the central screen, showing two black squares where information could be displayed (as displayed in Figure 2); one for each type of behaviour of interest.

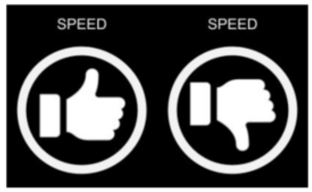
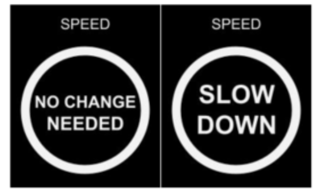
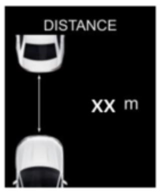
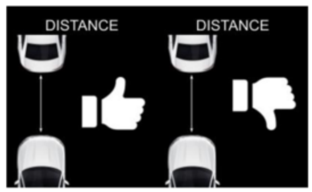
Figure 2

Simulator Set-Up Depicting the "Feedback Zone" on the Screen

In the Information condition, the feedback presented a numerical value, specifically, their speed in km/h and the distance to the lead vehicle in metres. In the Assessment condition, the feedback provided an assessment of these behaviours, indicating whether they were violating traffic rules or not, indicated by either a thumbs up or thumbs down icon. And in the Decision condition, the feedback suggested a behaviour to adopt, by either showing the text "no change needed", "slow down" or "increase distance". The specific illustrations and choice of words were chosen after collecting the opinions of 28 people in a short questionnaire study, to assess the understandability of the feedback and the participants' preferences. These are presented in Table 2.

Table 2

Feedback Images

	Condition		
	Information	Assessment	Decision
Feedback on Speed			
Feedback on Distance			

Note. A numerical factual value was presented instead of 'xx' for the Information condition

For the Information condition, the feedback continuously provided the driver with their current speed (in km/h) and the distance to the car ahead (in metres), when applicable. For conditions Assessment and Decision, the positive feedback was displayed at the start of the drive and would switch to the negative feedback if a threshold was exceeded. For speed, this threshold was when the speed was over the speed limit + 5.4 km/h (1.5 m/s), and for distance, the threshold was when the time headway (i.e., the distance headway divided by the ego vehicle's speed) was less than 1.2 s. To prevent abrupt changes between positive and negative feedback, possibly leading to flashing images, a dead band was introduced by a second threshold. The speed had to be below the speed limit + 0.5 m/s, and the time headway (THW) greater than 1.3 s to resume the positive feedback.

2.2.4 Recorded driving performance data

The two main variables of this experiment are speed and time headway (THW), presented as 'distance' to the participants. Both variables were recorded continuously throughout the drive. Based on that data, several dependent measures were computed, including average THW and speed, standard deviations of THW and speed, maximum and minimum THW and speed, and percentage of time below the recommended THW of 2 s or above the speed limit. To only keep relevant data, moments when the speed was < 3 km/h were removed, and the THW was analysed when it was below 100 s.

Furthermore, based on the speed and THW, feedback-related measures were computed: the percentage of time the positive feedback was displayed (i.e. when the behaviour was on the "correct" side of the threshold) and the number of times negative feedback was triggered. These measures were also computed for the Control and Information conditions, although the presence of a threshold in those conditions was not made explicit to the participant.

2.2.5 Self-reported data

In addition to performance data, self-reports were used to collect demographic information, attitudes concerning their driving and the feedback, and the mental effort and the situation awareness experienced during the drives.

Demographics included age, age of obtaining the first driving licence, nationality, primary mode of transport, and average annual mileage. Opinions on driving were collected with 7-point Likert-scale items from 1 (strongly agree) to 7 (strongly agree), assessing: whether participants enjoy driving a car, whether they think they are good drivers, whether they would judge themselves as better drivers than the average driver, and if they believe there is room for improvement when it comes to their driving ability. The questionnaire also included two Likert-scale items probing respondents' propensity of risk-taking, specifically, the acceptability of driving a little faster if one considers themselves a good driver, and the appropriateness of continuing to drive when traffic lights change from green to yellow. Technology affinity was also assessed with an adapted, shortened version of the Affinity for Technology Interaction (ATI) scale (Franke, Attig & Wessel, 2019). After the experiment, participants were asked whether they sometimes have difficulties assessing their speed, and the distance to a car in front of them.

Opinions on the speed and distance feedback were collected after each drive: whether participants found the feedback useful, whether they found it easy to understand, and whether they would like to have this kind of feedback in their own car. For these six items, a slider scale was used, ranging from 0 (not at all) to 100 (very much). Opinions on the three types of feedback were also collected at the end of the experiment: usefulness on a slider scale from 0 (not useful at all) to 100 (very useful), as well as the intent to use the feedback if it was available in the participants' cars, using a slider scale from 0 (no chance of using the feedback) to 100 (definite intention to use the feedback). Participants were also asked to rank the three types of feedback, based on their preference, at the end of the experiment.

The subjective mental effort for each condition was assessed using the Rating Scale Mental Effort (RSME; Zijlstra & Doorn, 1985) after each ride: participants could indicate their mental effort on a scale from 0 (absolutely no effort) to 150 (more than extreme effort). After all the rides, the subjective mental effort was assessed again per condition, using a slider scale ranging from 0 (no effort at all) to 100 (the biggest effort). Similarly, the situation awareness was assessed with three questions inspired by the Situational Awareness Rating Technique (SART; Taylor, 2017) and adapted to fit the driving activity. For example, one question was "How much were you concentrating on the road situations? Were you concentrating a lot (High) or a little (Low)?", and was rated by participants from 1 (low) to 7 (high).

The questionnaire in its entirety is included in the Supplementary Material.

2.3 Procedure

After receiving information about the study, specifying that the study concerned feedback on driving, and after giving their informed consent, participants started the experiment with a practice drive, in which they had to complete the entire route presented in Figure 1. They were instructed to follow the signs to go to the city "Venekerk", and were told that this route would take them through a rural area, a small village and then on a motorway and that they could stop the car at a parking area, after the exit lane of the motorway. They were also instructed to drive as they would drive in the real world. The feedback was not mentioned, nor were participants asked to follow the traffic rules. Upon completing the practice drive, and if they agreed to pursue the experiment, they completed a 15-minute questionnaire and proceeded with the experimental part of the study: four drives were completed on the same route again. Participants had to fill out a short questionnaire between each drive and after the four drives.

The order of the four conditions was randomised using a balanced Latin-square design (e.g., Edwards, 1951). A Latin square design was originally generated, but to avoid carryover effects (Kim & Stein, 2009), the conditions were also randomised by row and by column, obtaining a balanced Latin-square design. Thirty different conditions orders were then computed.

2.4 Data analysis

2.4.1 Driving performance data

Two participants were excluded from the driving performance analysis: the first participant had missed a turn in one of the conditions, before making a U-turn to rejoin the route, making the ride not comparable to the other three. The second participant showed aberrant and inconsistent performance across conditions, with for example a standard deviation of their average speed on the motorway of 25.1 km/h, against an average of 2.0 km/h for the other participants, and was therefore excluded. All the other participants were included in the analysis, even for example if their speed far exceeded the limits, as their behaviour remained consistent throughout the conditions. The final sample size for the driving performance data was $n = 27$.

The variables presented above (Section 2.2.4) were analysed using repeated measures analysis of variance (ANOVA) with feedback conditions as a factor. A *p*-value of .05 or under was considered significant, and effect sizes were interpreted according to Cohen's guidelines (1988). When the repeated measures ANOVA yielded a significant result, further analyses were conducted using planned contrasts: indeed, when specific research questions and hypotheses are formulated *a priori,* contrasts can be used to answer these questions (Bewick et al., 2004; Furr & Rosenthal, 2003) and avoid Type III errors (Haans, 2018). The first question was whether one of the feedback conditions affected the variable compared to the baseline level, and was answered using simple contrast analysis (i.e., each feedback condition was compared to the control condition). The second question was whether, compared to the average behaviour, a condition yielded different results, and was answered using deviation contrasts (i.e., each condition was compared to the average of all four conditions).

These analyses were conducted for both speed and THW on the whole drive. Additionally, as depicted in Figure 1, specific sections of interest were identified. Speed was also analysed (1) in a rural area limited at 80 km/h (from 150 to 850 m after the start of the ride, titled 'Section 1'), (2) in an urban area limited at 50 km/h (from 3600 to 4500 m, titled 'Section 3'), and (3) on a motorway limited at 100 km/h (from 8500 to 10500 m, titled 'Section 5'). THW was also analysed

(1) in an urban area (from 1500 to 2700 m, titled 'Section 2'), and

(2) on a motorway (from 6000 to 8000 m, titled 'Section 4'), two sections where a car drove in front of the participant.

2.4.2 Self-reported experiential data

All 29 participants were included in the analysis of the self-reports. Their opinions and attitudes were collected as described above, but the method of self-reporting led to 8.5% of missing data (167 missing values, out of a total of 1972). Missing data were imputed using the KNN imputation for the nearest neighbour row, allowing repeated measures ANOVA to be conducted on these results.

An acceptability score was computed for each feedback type based on the attitudes questions. This acceptability score is an average of the three scores to the questions asked immediately after each ride (usefulness, understandability, and whether participants would like to have this feedback) and the two scores to the questions asked at the end of the experiment (usefulness and intent to use). The acceptability score had a potential range of 0 to 100. To compare the acceptability of the different types of feedback, a two-way repeated measures ANOVA was used with feedback conditions (information, assessment, decision) and type of behaviour (speed, THW) as factors.

3. Results

3.1 Driving performance results

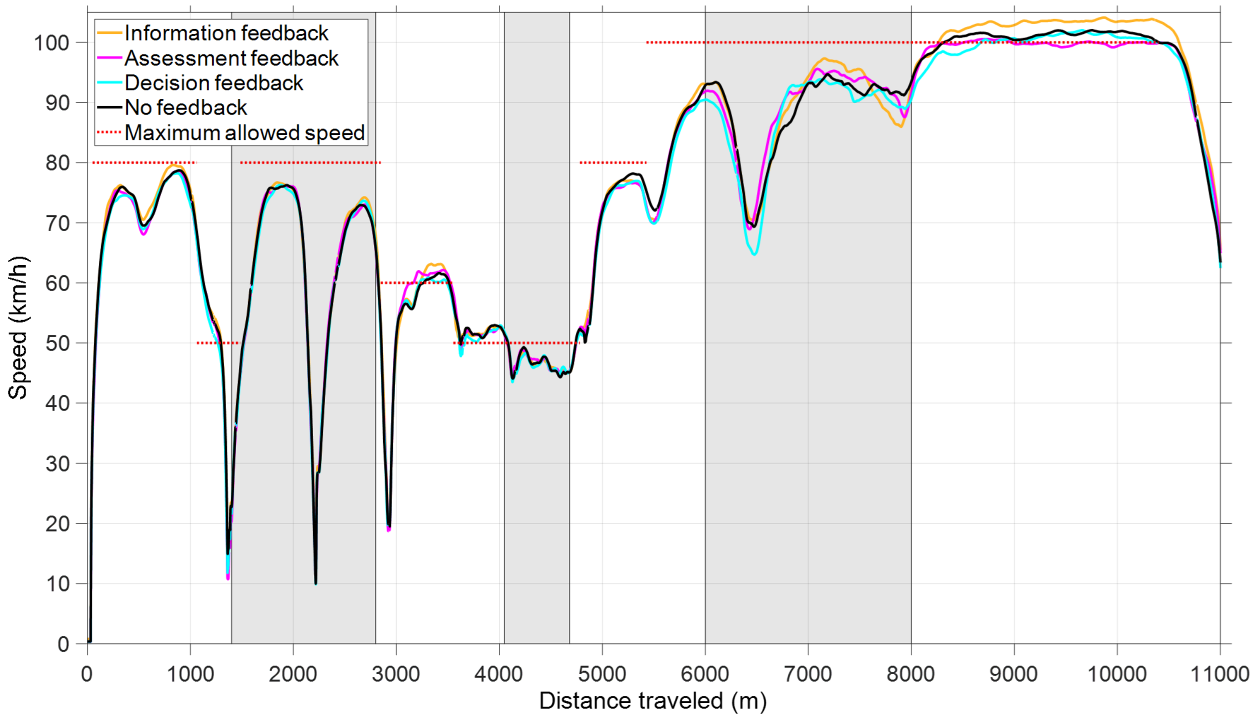
3.1.1 Speed and THW on the whole drive

Average Speed and THW

On average, participants drove 67.5 km/h (*SD* = 4.6), and more specifically, on average 67.7 km/h (*SD* = 4.4) in the Control condition, 67.9 km/h (*SD* = 4.6) in the Information condition, 67.7 km/h (*SD* = 4.9) in the Assessment condition and 66.8 km/h (*SD* = 4.3) in the Decision condition. The repeated measure ANOVA revealed no significant difference ($F(3,78) = 1.60$, $p = .196$, $\eta^2 = .058$) between conditions, yet indicating a small to medium effect size for a repeated-measure ANOVA. The speed in all conditions is plotted against the travelled distance driven in Figure 3.

Figure 3

Mean speed Across Participants as a Function of Travelled Distance, for Each of the Four Experimental Conditions



Note. The horizontal dotted lines represent the speed limits on that portion of the route. The grey-shaded backgrounds represent portions of the route where car-following took place.

The average THW was not significantly different across conditions ($F(3,78) = 1.68$, $p = .178$, $\eta^2 = .061$): 4.6 s (*SD* = 1.0) in the Control condition, 4.7 s (*SD* = 1.5) in the Information condition, 4.8 s (*SD* = 1.3) in the Assessment condition and 5.0 s (*SD* = 1.2) in the Decision condition.

Standard Deviation of Speed and THW

The participants' average *SD* of speed was significantly different across conditions ($F(3,78) = 4.19$, $p = .008$, $\eta^2 = .139$). The average *SD* of speed was 24.6 km/h (*SD* = 1.5) in the Control condition, 25.0 km/h (*SD* = 1.2) in the Information condition, 24.8 km/h (*SD* = 1.3) in the Assessment condition and 24.2 km/h (*SD* = 1.2) in the Decision condition. The simple contrast analysis revealed no significant differences, while the deviation contrasts analysis revealed that the *SD* of speed was significantly higher (0.37 km/h) in the Information condition than in the overall average *SD* ($t(78) = 2.57$, $p = .012$) and that the *SD* of speed was lower (0.43 km/h) in the Decision condition ($t(78) = -3.03$, $p = .003$).

The standard deviations of THW were not statistically significantly different across conditions ($F(3,78) = 0.06$, $p = .980$, $\eta^2 = .002$), with an average *SD* of 5.1 s in the Control, Information and Assessment conditions and 5.2 s in the Decision condition.

Maximum Speed and Minimum THW

Differences were found in the maximum speed ($F(3,78) = 3.87$, $p = .012$, $\eta^2 = .130$). The simple contrast analysis revealed no significant difference, and the deviation contrast analysis revealed that the maximum speed was higher in the Information condition ($t(78) = 2.91$, $p = .005$) with an average maximum speed of 106.9 km/h ($SD = 7.2$), and lower in the Decision condition ($t(78) = -2.63$, $p = .010$) with an average maximum speed of 104.2 km/h ($SD = 5.0$).

No significant differences were found in minimum THW ($F(3,78) = 0.14$, $p = .936$, $\eta^2 = .005$) across conditions.

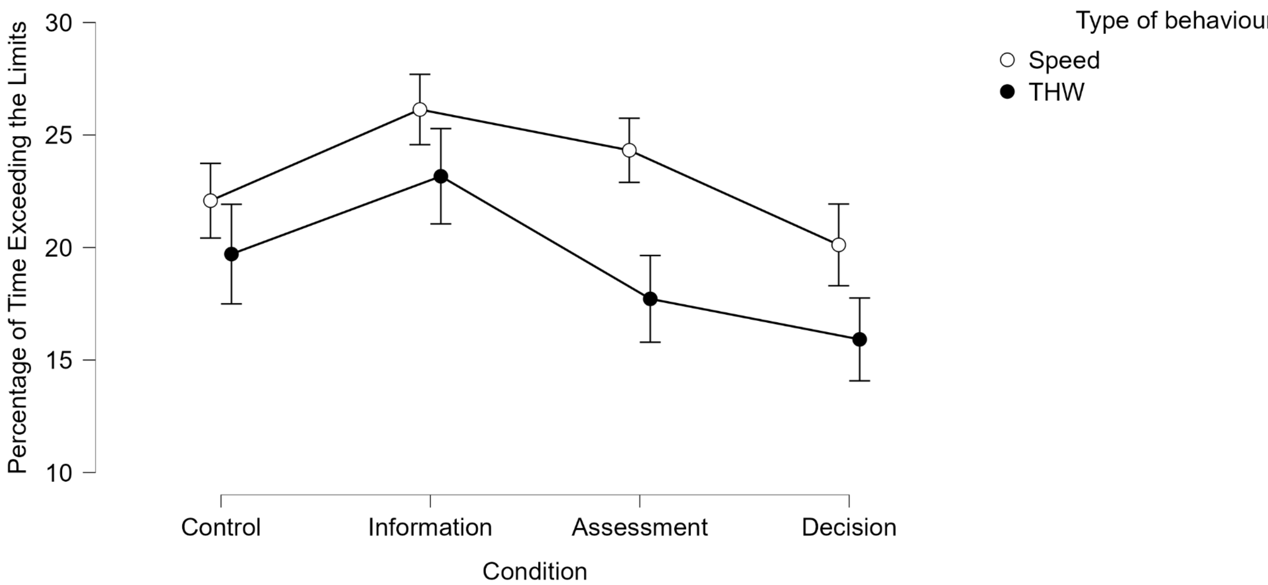
Comparison to the limits

The comparison between the percentage of time participants drove above the speed limit differed significantly, $F(3,78) = 3.96$, $p = .011$, $\eta^2 = .132$. Participants were above the speed limit for 22.1% (SD = 15.4) of the drive in the Control condition, 26.1% (SD = 16.0) in the Information condition, 24.3% (SD = 15.9) in the Assessment condition and 20.1% (SD = 13.0) in the Decision condition. The simple contrast analysis indicated one significant difference: the time spent above the speed limit was significantly higher (4.1%) in the Information condition than in the Control condition ($t(78) = 2.18$, $p = .033$). Deviation contrast analysis revealed that the percentage of time above the speed limit was higher in the Information condition ($t(78) = 2.60$, $p = .011$), and lower in the Decision condition ($t(78) = -2.67$, $p = .009$).

The percentage of time under the recommended THW (i.e., under 2 s) yielded a significant difference in the repeated measures ANOVA ($F(3,78) = 2.78$, $p = .047$, $\eta^2 = .097$); participants were under the THW limit for 19.7% (SD = 16.0) of the drive in the Control condition, 23.2% (SD = 19.7) in the Information condition, 17.7% (SD = 15.7) in the Assessment condition and 15.9% (SD = 11.5) in the Decision condition. The simple contrast analysis revealed no statistical differences, and the deviation contrast analysis revealed that the percentage of time under the limit was significantly higher (4.0%) in the Information condition ($t(78) = 2.50$, $p = .014$) than the average of all conditions. These results are illustrated in Figure 2, with standard deviation error bars.

Figure 2

Percentage of Time Exceeding the Limits Depending on the Feedback Condition and the Behaviour Type



3.1.2 Speed and THW in specific sections

Speed was investigated in three sections: a rural area ('Section 1'), which yielded no significant differences (F(3,78) = 1.14, p = .337, η^2 = .042), an urban area ('Section 3'), which also yielded no significant differences (F(3,78) = 1.52, p = .215, η^2 = .055), and a motorway area ('Section 5'), which led to significant differences in the average speed (F(3,78) = 5.15, p = .003, η^2 = .165). On this particular section of the motorway, the average speed was 100.2 km/h (*SD* = 4.9) in the Control condition, 101.5 km/h (*SD* = 3.8) in the Information condition, 99.9 km/h (*SD* = 4.1) in the Assessment condition, and 99.0 km/h (*SD* = 2.9) in the Decision condition. The simple contrast analysis indicated one significant difference: the average speed was significantly higher (1.3 km/h) in the Information condition than in the Control condition (t(78) = 2.04, p = .044). Deviation contrast analysis indicated that the average speed was significantly higher (1.4 km/h) in the Information condition than the average (t(78) = 3.42, p = .001) and 1.2 km/h lower in the Decision condition (t(78) = -2.93, p = .005).

The analysis of the average THW revealed no significant differences, neither in the rural area ('Section 2'; F(3,78) = 1.43, p = .239, η^2 = .052) nor in the motorway area ('Section 4'; F(3,78) = 0.49, p = .687, η^2 = .019). The average THW was 6.4 s for Section 2, and 3.7 s for Section 4.

3.1.3 Other variables

Percentage of negative feedback

The comparison of the percentages of time participants received negative feedback on speed revealed no significant differences ($F(3,78) = 2.57$, $p = .060$, $\eta^2 = .090$). Note that although participants in the Control condition received no feedback and those in the Information condition received only neutral feedback, we estimated the virtual negative feedback that hypothetically would have been provided in these conditions using the same thresholds as used for the negative feedback in the Assessment and Decision conditions. Participants received, on average, negative feedback for 8.2% (*SD* = 11.0) of the drive in the Control condition, 9.0% (*SD* = 11.1) in the Information condition, 6.3% (*SD* = 9.1) in the Assessment condition and 4.9% (*SD* = 4.7) in the Decision condition.

The percentage of negative feedback for the THW was also found to be not statistically different across conditions ($F(3,78) = 2.01$, $p = .119$, $\eta^2 = .072$). Participants received negative feedback for 6.8% (*SD* = 12.7) of the drive in the Control condition, 7.3% (*SD* = 13.4) in the Information condition, 4.6% (*SD* = 9.7) in the Assessment condition and 3.5% km/h (*SD* = 3.3) in the Decision condition.

Number of triggers of negative feedback

The number of times the speed feedback turned negative was not significantly different between conditions ($F(3,78) = 0.33$, $p = .802$, $\eta^2 = .013$). On average, the feedback turned negative 4.0 times (*SD* = 3.0) in the Control condition, 4.4 times (*SD* = 3.3) in the Information condition, 4.3 times (*SD* = 3.1) in the Assessment condition and 4.1 times (*SD* = 2.8) in the Decision condition.

Similarly, the number of times the THW feedback turned negative was not significantly different across conditions ($F(3,78) = 0.50$, $p = .685$, $\eta^2 = .019$). On average, the feedback turned negative 2.1 times (*SD* = 1.6) in the Control condition, 2.2 times (*SD* = 1.3) in the Information condition, 2.0 times (*SD* = 1.6) in the Assessment condition and 2.3 times (*SD* = 1.3) in the Decision condition.

3.2 Self-reported results

3.2.1 Attitudes and acceptability of the feedback

Attitudes about the feedback were collected both immediately after participants experienced the feedback and after the experiment, on slider scales ranging from 0 (not at all) to 100 (very much). The mean scores and standard deviations of these items are presented in Table 3 for the feedback on Speed, and in Table 4 for the feedback on THW.

In terms of speed-related feedback, the Information feedback was preferred overall, with an average acceptability score of 70.4, against 60.4 and 52.1 for the Assessment and Decision feedback. The

Information feedback received the highest rating on understandability, desire to have the feedback, intent to use if all options were available and usefulness when asked at the end of the experiment.

Regarding distance-related feedback, no such pattern was found as no condition stood out as the preferred. Acceptability scores averaged around 50, in the middle of the scale from 0 to 100; indicating in this context that participants are neither positive or negative about the THW feedback. The Decision feedback was rated slightly higher on its usefulness and desire to have the feedback.

Table 3

Mean Scores and Standard Deviations of Attitudes Items (Range 0–100) on the Speed Feedback

Feedback condition	Information	Assessment	Decision
Immediate attitudes (on Speed Feedback)			
The information was useful	58.3 (31.7)	65.3 (21.3)	50.1 (27.3)
The information was easy to understand*	92.8^{a*} (9.7)	88.0 (18.5)	81.2^{a*} (23.3)
I would like to have this kind of information available in my own car**	64.3 $^{a*\ b**}$ (34.0)	44.4 ^{a*} (30.2)	39.9 ^{b**} (28.0)
Post-experiment attitudes (on Speed Feedback)			
The information was useful*	67.4 ^{a*} (31.9)	58.8 (25.4)	49.5 ^{a*} (26.2)
Assuming that all feedback options were available in your car, intent to use feedback***	69.2 $^{a***\ b***}$ (31.4)	45.6 ^{a***} (29.8)	39.9 ^{b***} (27.4)
Acceptability score (on Speed Feedback)			
Average of the five prior items**	70.4 ^{a**} (22.0)	60.4 (20.3)	52.1 ^{a**} (17.7)

Note. Statistical differences were investigated with RM ANOVA and are indicated with * when $p < .05$, ** when $p < .01$ and *** when $p < .001$.

Table 4

Mean Scores and Standard Deviations of Attitudes Items (Range 0–100) on the THW Feedback

Feedback condition	Information	Assessment	Decision
Immediate attitudes (on THW Feedback)			
The information was useful	41.8 (28.5)	43.6 (21.6)	50.2 (32.6)
The information was easy to understand	82.1 (28.1)	92.1 (10.9)	80.1 (25.0)
I would like to have this kind of information available in my own car	40.2 (29.5)	40.4 (29.8)	50.5 (32.7)
Post-experiment attitudes (on THW Feedback)			
The information was useful	45.6 (33.7)	47.9 (27.5)	52.7 (30.2)
Assuming that all feedback options were available in your car, intent to use feedback	38.4 (34.2)	41.6 (28.0)	42.8 (33.3)
Acceptability score (on THW Feedback)			
Average of the five prior items	49.6 (25.3)	53.1 (17.5)	55.3 (21.5)

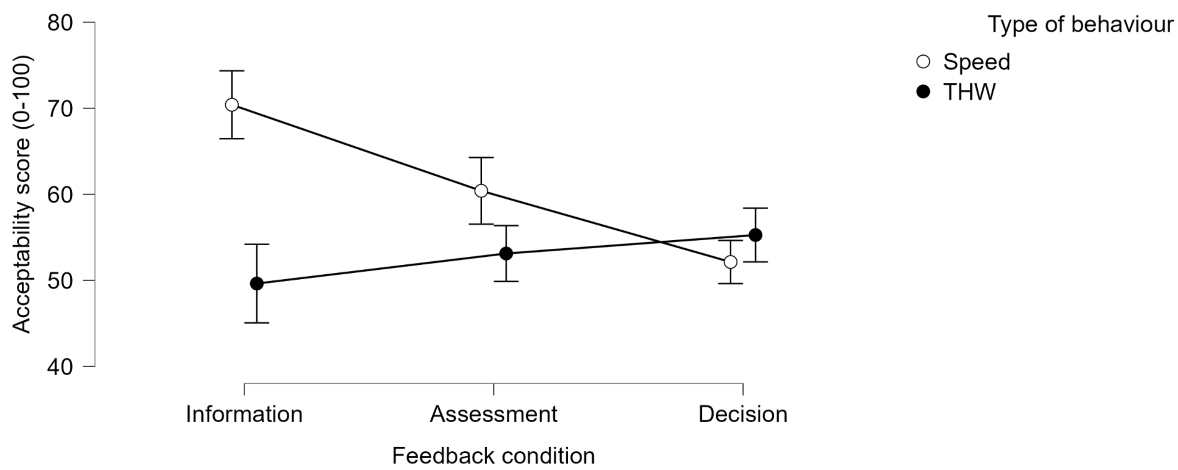
Note. No significant differences were found between conditions.

Acceptability scores were further analysed with a two-way repeated measures ANOVA, which indicated a significant effect for one of the factors: the type of behaviour targeted yielded differences ($F(1,28) = 12.9$, $p = .001$, $\eta^2 = .049$), but not the condition ($F(2,56) = 0.89$, $p = .418$, $\eta^2 = .019$). Feedback, regardless of the condition, was preferred when it related to speed than THW ($p = .001$, 95% C.I. = [3.6; 13.1]).

The interaction effect was also significant: $F(2,56) = 12.1$, $p < .001$, $\eta^2 = .068$. The post hoc analysis of the interaction effect, using the Bonferroni correction, revealed that three pairs are significantly different: (1) the Information feedback was preferred when it related to speed that to THW ($p < .001$, 95% C.I. = [9.8; 31.8]), (2) the Information feedback on speed was preferred to the Assessment feedback on THW ($p = .032$, 95% C.I. = [0.8; 33.8]) and (3) when it related to speed, the Information feedback was preferred to the Decision feedback ($p = .014$, 95% C.I. = [2.1; 34.4]). These results are illustrated in Figure 3, with standard deviation error bars.

Figure 3

Acceptability Scores (range 0-100) Depending on the Condition and the Type of Behaviour



The average rank, from 1 (most preferred) to 3 (least preferred), was 1.57 (*SD* = 0.68) for the Speed feedback and 2.48 (*SD* = 1.33) for the THW feedback in the Information condition. In the Assessment condition, it was 2.24 (*SD* = 0.94) for the Speed feedback and 2.38 (*SD* = 1.36) for THW feedback, and in the Decision condition, the ranks were 2.19 (*SD* = 0.68) for Speed feedback and 2.67 (*SD* = 0.73) for THW feedback.

3.2.2 Impact of feedback on mental effort and situation awareness

The mental effort, evaluated on the RSME ranging from 0 to 150 directly after each ride, yielded scores of 45.7 (*SD* = 25.0) in the Control condition, 45.9 (*SD* = 23.9) in the Information condition, 45.5 (*SD* = 23.6) in the Assessment condition and 46.7 (*SD* = 23.5) in the Decision condition. These results are not statistically different ($F(3,84) = 0.08$, $p = .971$, $\eta^2 = .003$). At the end of the experiment, the mental effort was assessed again, on a scale from 0 to 100: 43.2 (*SD* = 22.8) in the Control condition, 45.2 (*SD* = 23.1) in the Information condition, 40.7 (*SD* = 24.2) in the Assessment condition and 41.0 (*SD* = 26.2) in the Decision condition, all found to not be statistically different ($F(3,84) = 0.84$, $p = .476$, $\eta^2 = .029$).

The situation awareness was first assessed directly after each ride, on a 7-point Likert scale from 1 (least aware) to 7 (most aware). The results are not statistically different ($F(3,84) = 0.34$, $p = .798$, $\eta^2 = .012$): the average score was 4.7 (*SD* = 0.9) in the Control condition, 4.8 (*SD* = 0.8) in the Information condition, 4.7 (*SD* = 0.7) in the Assessment condition, and 4.7 (*SD* = 0.8) in the Decision condition. Similarly to the mental effort variable, the situation awareness was assessed again at the end of the experiment with a scale from 0 to 100: 64.2 (*SD* = 24.7) in the Control condition, 58.4 (*SD* = 24.2) in the Information condition, 54.1 (*SD* = 23.4) in the Assessment condition and 54.5 (*SD* = 24.7) in the Decision condition; which also yielded no statistical difference ($F(3,81) = 2.71$, $p = .051$, $\eta^2 = .091$).

4. Discussion

This driving simulator study aimed to investigate the stage of information processing required when receiving feedback on driver behaviour. Specifically, we sought to determine whether providing a piece of information alone, an assessment of that information, or a decision based on that assessment would yield different driving outcomes, and which approach would better support drivers in their tasks. This research interest was grounded in the assumption that providing feedback affects behaviour, as was observed for example by Abrahamse et al. (2005) on household energy conservation and De Waard et al. (1999) on traffic violations.

The first result of the study is the similarity of behaviours in all four conditions in the control as well as feedback conditions. Providing feedback had no apparent effect on the two behaviours of interest, speed and time headway (THW). Still, other behavioural variables yielded differences between feedback conditions. This was the case, for example, for the standard deviation of speed, the maximum speed, the time above speed limits, time under the recommended THW, and the speed on the motorway section. For most of these measures, it turned out that the Information condition yielded, in terms of safety, worse behaviour and the Decision condition yielded better behaviour. These two main findings—the non-significant effect of feedback conditions on speed and THW and the significant effect on other behavioural variables—will be discussed here, as well as the results from the self-reports.

The first result of the study is the apparent null effect of feedback on speed and THW. This finding points to the difficulty of inducing observable behaviour change with feedback. Although both target behaviours did not differ significantly across conditions, the eta squared values ($\eta^2 = .058$ for speed and $\eta^2 = .061$ for THW) indicate medium effects (Cohen, 1988), suggesting that a larger sample size may be needed. Additionally, a significant speed change was observed on the motorway with a large effect size ($\eta^2 = .165$) and a maximum difference in mean speed of 2.5 km/h between conditions. Thus, providing feedback had a limited but non-negligible effect: a change of 2.5 km/h, especially in a zone limited to 100 km/h, has an impact on crash reduction (approximately 7.5% reduction according to Mazureck & Van Hattem, 2006). Our relatively small effects may be unexpected considering a recent questionnaire study by Ge et al. (2023), in which a significant difference in self-reported behavioural intentions was observed among 110 participants between feedback conditions similar to those used in this study. Thus, it seems that neither behavioural intentions nor feedback alone may reliably lead to meaningful behavioural change.

The limited effect of feedback alone, i.e., without specific instructions or incentives, has been noted before. For example, Marciano et al. (2015) investigated the effect of overt and covert speed cameras and of immediate feedback and observed no speed improvement due solely to immediate feedback. Furthermore, Mullen et al. (2015) conducted an on-road study in which they examined the effect of feedback and of a token economy (i.e., monetary incentive) on speeding behaviour. The results not only showed no difference between the feedback-only and control conditions but also indicated similar and better behaviours in the token economy coupled with feedback and token-only conditions. In other words, providing an incentive, even without feedback, yielded behaviour change, contrary to providing feedback without an incentive. The same conclusion was reached by Bolderdijk et al. (2011) in a study investigating the effect of a Pay-As-You-Drive (PAYD) scheme on speed choice. They found that a combination of feedback and incentive was effective in reducing speeds. However, they reported that the effect was most likely due to the incentive alone, as the majority of their participants did not log in to the website to access their feedback. It could be inferred that participants received virtually no feedback and their knowledge of the incentive alone was enough to change their behaviour. Considering that many studies on the effect of (real-time) feedback on driving performance also include incentives or rewards (e.g., Dijksterhuis et al., 2015, 2016; Masello et al., 2023; Mazureck & Van Hattem, 2006), it seems reasonable to infer that the effect of incentives on behaviour could have been misattributed to the potential effect of feedback. Whether the lack of studies involving real-time feedback alone is due to publication bias or due to a limited research interest remains to be determined.

The limited effect of feedback, taken into consideration together with the more flagrant effect of incentives, indicates how drivers can be guided toward behaviour change. Indeed, based on the studies from Mullen et al., (2015) and Bolderdijk et al. (2011), as well as the primary result of the current simulator experiment, it seems that providing external motivation leads to behavioural change while providing feedback that could help drivers, at the very least with self-assessment or assessing the situation, only has limited effect. This statement is also supported by the disappearing long-term effect of most safety interventions: for as long as incentives are present, behaviours typically improve, but the effects withdraw as the external motivation does as if no learning has been done (see e.g., Mazureck & Van Hattem, 2006). Speeding should therefore be considered more of a motivational issue than a capability one, and the same logic could be extended to other risky driving behaviours. As an example, Pampel et al. (2018) brought to light that people know how to drive eco-friendly and can opt for it when requested, but that they would return back to 'everyday' driving after some time or when their mental workload increases. It could be argued that a similar pattern exists for safe driving: people know how to drive safely, but without incentives or external motivation, it is not a long-lasting mental model.

When considered through this lens, observing only limited behavioural change due to feedback could have been expected, especially with regard to consciously controllable behaviours such as speeding and THW, for which participants did not indicate having difficulties assessing their own behaviour. It can be hypothesised that on such behaviours, participants felt knowledgeable and able enough not to necessitate support, and there is only little margin for the feedback to have an effect.

When feedback conditions yielded different behaviour, it most often followed a certain pattern: the Information condition yielded worse behaviour (specifically on the standard deviation of speed, maximum speed, time spent above the speed limits and under the THW recommendation), the Assessment condition did not differ even once from the Control condition, and the Decision condition yielded better behaviour (specifically on the standard deviation of speed, maximum speed and the time spent above the speed limits). Better and worse behaviours are determined as follows: a higher speed SD is considered more dangerous as speed fluctuations can reduce safety margins, a higher maximum speed is also considered more dangerous, as is more time spent above speed or THW limits. The differences found between feedback conditions, and moreover, the pattern observed, indicate an effect of the stages of information processing.

The nuances of this pattern of results are as follows. First, with respect to the results of the Assessment condition, providing drivers with an assessment of their behaviour did not have any effect compared to a control condition. While speed limits were easily inferred from numerous speed signs, the Assessment condition continuously provided drivers with more information than their speedometer. This result seems to point to the fact that their assessment (i.e., comparing their speed to the speed limits) was already functional. In fact, this result is in line with elements developed earlier in this discussion: it indicates that the participants already knew whether their behaviour was *good* or *bad*, perhaps in the same way people know how to drive in an eco-friendly manner (Pampel, 2018). Second, the Information condition yielded worse behaviour, which could be an example of behavioural adaptation (Smiley & Rudin-Brown, 2020): it can be hypothesised that providing a numerical value for the speed changed the basis on which participants made their assessment, possibly making them feel safer, and causing an increase in more dangerous behaviour. Third and last, the Decision condition was the most effective in inducing positive behaviour change and could be explained by two mechanisms. It could be, first, evidence of a failure in the association between dangerous behaviour and a need for behaviour change: informing participants that their actual behaviour is not conforming did not have any effect compared to not informing them (as was done in the Assessment condition), but instructing them to follow a specific behaviour when their behaviour was not conforming had little effect. It could also be evidence of more social phenomena, such as social desirability or instructions compliance, as it was made more clear what was expected from the participants in the Decision condition.

Interestingly, the condition that yielded the best behavioural results also yielded the worst attitudinal results: while all feedback conditions were rated the same for the THW variable, for the speed variable, the results differed with the Information condition being the preferred feedback and the Decision condition the least liked. This suggests a trade-off between experiential and instrumental outcomes, similar to Degirmenci and Breitner (2023). During the informal debriefing of the experiment, participants primarily expressed their dislike of being told what to do, with some mentioning that they felt more capable than their car in deciding the best behaviour to adopt. This element could be a shortcoming of Decision-oriented feedback: the credibility of the feedback and of its issuer are main predictors of effectiveness (Poulos & Mahony, 2008). Another potential shortcoming could be the exclusion of the driver from the decision-making process, which could possibly reduce situation awareness (De Winter et al., 2014), although the feedback in this experiment did not statistically significantly impact the participants' self-reported situation awareness. In fact, the feedback seemed to not have negative - nor positive - consequences either in terms of mental effort or situation awareness.

5. Conclusion

This experiment entails two main findings: the limited effect that feedback alone has on speed and THW, resulting from risky behaviour being a motivational issue rather than a capability one, and the importance of the stage of information processing of the feedback. From an ergonomics viewpoint, it seemed that suggesting a behaviour was the best approach (Forzy, 2004) as a way to keep the driver in the control loop. However, the main drawback of this approach is that it does not directly affect motivation. Future research should focus on identifying ways to appeal to safer behaviour, either through changing risk perception, possibly using feedback, or through incentives, as it has been proven more effective. Identifying risks and risky behaviour seems to not be an issue for drivers, but risks are too often considered acceptable on the roads (see e.g., Molin & Brookhuis, 2006, on the influence of beliefs regarding speeding on the acceptability of Intelligent Speed Adaptation).

As a result, the recommendations for future safety interventions are: (1) to assist drivers in associating road risks with a need for behavioural change, either by altering risk perception and beliefs or by providing external motivation for this behavioural change; and (2) to suggest a behaviour to adopt when providing feedback, as it has proven to be the most effective stage of information processing to provide drivers with.

References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*, *25*(3), 273–291. https://doi.org/10.1016/J.JENVP.2005.08.002

Bell, J. L., Taylor, M. A., Chen, G. X., Kirk, R. D., & Leatherman, E. R. (2017). Evaluation of an in-vehicle monitoring system (IVMS) to reduce risky driving behaviors in commercial drivers: Comparison of in-cab warning lights and supervisory coaching with videos of driving behavior. *Journal of Safety Research*, *60*, 125–136. https://doi.org/10.1016/J.JSR.2016.12.008

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 9: One-way analysis of variance. *Critical Care*, *8*(2), 130–136. https://doi.org/10.1186/cc2836

Bolderdijk, J. W., Knockaert, J., Steg, E. M., & Verhoef, E. T. (2011). Effects of Pay-As-You-Drive vehicle insurance on young drivers' speed choice: Results of a Dutch field experiment. *Accident Analysis & Prevention*, *43*(3), 1181–1186. https://doi.org/10.1016/J.AAP.2010.12.032

CBS (Dutch Central Bureau of Statistics). (2019). *80 percent of adults have a driving licence*. https://www.cbs.nl/en-gb/news/2019/09/80-percent-of-adults-have-a-driving-licence

CBS (Dutch Central Bureau of Statistics). (2022). *Mobility; per person, personal characteristics, modes of travel and regions*. https://www.cbs.nl/en-gb/figures/detail/84709ENG?q=kilometer driven#DistanceTravelled_2

CBS (Dutch Central Bureau of Statistics). (2022). *Traffic performance motor vehicles; kilometres, type of vehicle, territory*. https://www.cbs.nl/en-gb/figures/detail/85395ENG?q=traffic accidents

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). NJ: Erlbaum.

Degirmenci, K., & Breitner, M. H. (2023). Gamification and sensory stimuli in eco-driving research: A field experiment to reduce energy consumption in electric vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *92*, 266–282. https://doi.org/10.1016/J.TRF.2022.10.014

De Waard, D., Van der Hulst, M., & Brookhuis, K. A. (1999). Elderly and young drivers' reaction to an in-car enforcement and tutoring system. *Applied Ergonomics*, *30*(2), 147–157. https://doi.org/10.1016/S0003-6870(98)00002-7

De Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 196–217. https://doi.org/10.1016/J.TRF.2014.06.016

Dijksterhuis, C., Lewis-Evans, B., Jelijs, B., De Waard, D., Brookhuis, K., & Tucha, O. (2015). The impact of immediate or delayed feedback on driving behaviour in a simulated Pay-As-You-Drive system. *Accident Analysis & Prevention*, *75*, 93–104. https://doi.org/10.1016/J.AAP.2014.11.017

Dijksterhuis, C., Lewis-Evans, B., Jelijs, B., Tucha, O., De Waard, D., & Brookhuis, K. (2016). In-car usage-based insurance feedback strategies. A comparative driving simulator study. *Ergonomics, 59*(9), 1158–1170. https://doi.org/10.1080/00140139.2015.1127428

Driessen, T., Picco, A., Dodou, D., De Waard, D., & De Winter, J. (2021). Driving examiners' views on data-driven assessment of test candidates: An interview study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *83*. https://doi.org/10.1016/j.trf.2021.09.021

Edwards, A. L. (1951). Balanced Latin-square designs in psychological research. *The American Journal of Psychology*, *64*(4), 598–603. https://doi.org/10.2307/1418200

Elvik, R. (2005). Speed and road safety: Synthesis of evidence from evaluation studies. *Transportation Research Record*, *1908*, 59–69. https://doi.org/10.3141/1908-08

Eriksson, A., Petermeijer, S. M., Zimmermann, M., De Winter, J. C. F., Bengler, K. J., & Stanton, N. A. (2019). Rolling out the red (and Green) Carpet: Supporting driver decision making in automation-to-manual transitions. *IEEE Transactions on Human-Machine Systems*, *49*(1), 20–31. https://doi.org/10.1109/THMS.2018.2883862

Forzy, J.-F. (2004). Conduite automobile et conception ergonomique. In P. Falzon (Ed.), *Ergonomie* (pp. 647–654). PUF.

Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction*, *35*(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Furr, R. M., & Rosenthal, R. (2003). Evaluating Theories Efficiently: The Nuts and Bolts of Contrast Analysis. *Understanding Statistics*, *2*(1), 33–67. https://doi.org/10.1207/S15328031US0201_03

Ge, Y., Luo, X., & Qu, W. (2023). Impact of feedback content on dangerous driving behaviours based on the triangle model of responsibility. *Transportation Research Part F: Traffic Psychology and Behaviour*, *93*, 1–10. https://doi.org/10.1016/J.TRF.2022.12.003

Goodrich, M. A., & Quigley, M. (2004). Learning haptic feedback for guiding driver behavioral. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, *3*, 2507–2512. https://doi.org/10.1109/ICSMC.2004.1400706

Haans, A. (2018). Contrast Analysis: A Tutorial. *Practical Assessment, Research, and Evaluation*, *23*(1), 9. https://doi.org/https://doi.org/10.7275/7dey-zd62

Kerwin, T., & Bushman, B. J. (2020). Measuring the perception of aggression in driving behavior. *Accident Analysis & Prevention*, *145*, 105709. https://doi.org/10.1016/J.AAP.2020.105709

Kim, B. G., & Stein, H. H. (2009). A spreadsheet program for making a balanced Latin Square design. *Revista Colombiana de Ciencias Pecuarias*. http://rccp.udea.edu.co

Lewis-Evans, B., De Waard, D., & Brookhuis, K. A. (2010). That's close enough—A threshold effect of time headway on the experience of risk, task difficulty, effort, and comfort. *Accident Analysis & Prevention*, *42*(6), 1926–1933. https://doi.org/10.1016/J.AAP.2010.05.014

Marciano, H., Setter, P., & Norman, J. (2015). Overt vs. covert speed cameras in combination with delayed vs. immediate feedback to the offender. *Accident Analysis & Prevention*, *79*, 231–240. https://doi.org/10.1016/J.AAP.2015.03.028

Mase, J. M., Majid, S., Mesgarpour, M., Torres, M. T., Figueredo, G. P., & Chapman, P. (2020). Evaluating the impact of Heavy Goods Vehicle driver monitoring and coaching to reduce risky behaviour. *Accident Analysis & Prevention*, *146*, 105754. https://doi.org/10.1016/J.AAP.2020.105754

Masello, L., Sheehan, B., Castignani, G., Shannon, D., & Murphy, F. (2023). On the impact of advanced driver assistance systems on driving distraction and risky behaviour: An empirical analysis of irish commercial drivers. *Accident Analysis & Prevention*, *183*, 106969. https://doi.org/10.1016/J.AAP.2023.106969

Mazureck, U., & Van Hattem, J. (2006). Rewards for safe driving behavior: Influence on following distance and speed. *Transportation Research Record*, *1980*, 31–38. https://doi.org/10.3141/1980-07

Mcruer, D., & Weir, D. H. (2007). Theory of manual vehicular control. *Ergonomics*, *12*(4), 599–633. https://doi.org/10.1080/00140136908931082

Molin, E. J., & Brookhuis, K. A. (2007). Modelling acceptability of the intelligent speed adapter. *Transportation research part F: traffic psychology and behaviour*, *10*(2), 99-108.

Mullen, N. W., Maxwell, H., & Bédard, M. (2015). Decreasing driver speeding with feedback and a token economy. *Transportation Research Part F: Traffic Psychology and Behaviour*, *28*, 77–85. https://doi.org/10.1016/J.TRF.2014.11.008

Navarro, J. (2017). Human-machine interaction theories and lane departure warnings. *Theoretical issues in ergonomics science*, *18*(6), 519–547. https://doi.org/10.1080/1463922X.2016.1243274

Pampel, S. M., Jamson, S. L., Hibberd, D. L., & Barnard, Y. (2018). Old habits die hard? The fragility of eco-driving mental models and why green driving behaviour is difficult to sustain. *Transportation Research Part F: Traffic Psychology and Behaviour*, *57*, 139–150. https://doi.org/10.1016/J.TRF.2018.01.005

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans.*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Picco, A., Stuiver, A., De Winter, J., & de Waard, D. (2023). The use of monitoring and feedback devices in driving: An assessment of acceptability and its key determinants. *Transportation Research Part F: Traffic Psychology and Behaviour*, *92*. https://doi.org/10.1016/j.trf.2022.10.021

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment and Evaluation in Higher Education*, *33*(2), 143–154. https://doi.org/10.1080/02602930601127869

Singh, S. (2018). Critical reasons for crashes investigated in the national motor vehicle crash causation survey.(Traffic safety facts crash stats). *National Highway Traffic Safety Administration (NHTSA), 1200*.

Smiley, A., & Rudin-Brown, C. (2020). Drivers adapt–Be prepared for It! *Accident Analysis & Prevention*, 135, 105370.

Stanton, N. A., & Salmon, P. M. (2009). Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Safety Science*, *47*(2), 227–237. https://doi.org/10.1016/J.SSCI.2008.03.006

Taylor, R. M. (2017). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational Awareness* (pp. 111–128). Routledge.

Toledo, T., & Lotan, T. (2006). In-Vehicle Data Recorder for Evaluation of Driving Behavior and Safety. *Transportation Research Record*, *1953*(1), 112–119. https://doi.org/10.1177/0361198106195300113

Toledo, T., Musicant, O., & Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transportation Research Part C: Emerging Technologies*, *16*(3), 320–331. https://doi.org/10.1016/J.TRC.2008.01.001

Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., Hume, R. D., Mayer, R. E., Stansifer, R. L., & Castellan, N. J. (1979). *Tri-level study of the causes of traffic accidents: final report. Executive summary.* http://deepblue.lib.umich.edu/handle/2027.42/64993

Van der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

Zijlstra, F. R. H., & Van Doorn, L. (1985). *The construction of a scale to measure perceived effort.*