Inaccuracies in energy efficiency perception based on instantaneous consumption displays – Implications for interface design

Vivien Moll[a], Thomas Franke[a]

[a] Institute for Multimedia and Interactive Systems, Engineering Psychology and Cognitive Ergonomics, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany, moll@imis.uni-luebeck.de, franke@imis.uni-luebeck.de (corresponding author)

ABSTRACT

Instantaneous consumption displays (ICDs) can be used as central information source to perceive the energy efficiency of their manoeuvre-level driving. A key question is whether drivers who use ICDs can accurately derive efficiency value differences across driving strategies based on this information presented by the ICD. There is reason to assume that drivers' consumption judgements may be biased, similar to related phenomena like the time-saving bias. Therefore, the aim of the present research was to examine drivers' accuracy in deriving average consumption from dynamic ICD sequences. Participants viewed videos of a schematic ICD in a controlled experiment where the maximum instantaneous consumption systematically varied over time. Participants' (*N* = 55) overestimated the average consumption values. The empirical ranking of the sequences also significantly differed from the correct efficiency ranks. The current study incorporated multilevel modelling due to the nested structure of the data. The estimation difference was greater with higher peak height and shorter peak duration. The effect of peak height on estimation difference weakens with longer peak duration. In sum, the results indicate that ICDs can create biased perceptions of energy efficiency. Knowledge and affinity for technology interaction appear to impact estimation biases, whereas experience with consumption displays seems irrelevant. Further studies should test less biased interface designs such as manoeuvre-based aggregation or fading-trace approaches. Moreover, studies are needed that enable modelling of the effects of more natural temporal-spatial visual attention distribution (e.g., via an occlusion paradigm applied in a driving simulator setting).

Keywords: Biased perception, instantaneous consumption displays, dynamic data visualisation, energy efficiency, electric vehicle

1 INTRODUCTION

Drivers' are a critical factor regarding whether a vehicle's energy efficiency potential can be optimized in real-world usage (Barkenbus, 2010). Electric drivetrains in particular are highly susceptible to variations in everyday operational driving behaviour (Sivak & Schoettle, 2012). Drivers continuously adapt their behavior to balance different driving goals (Dogan, Steg, & Delhomme, 2011). From an analytical standpoint, one can assume that, at the most basic level, drivers want to reach the destination safely, on time, without excessive cognitive workload or stress, without significant financial waste (e.g., speeding tickets, vehicle damage, or excessive fuel/energy consumption), and without disturbing other road users or passengers with the resulting driving style. While avoiding crashes are assumed to be the primary goal, a secondary goal might be to drive as energy-efficient as possible for environmental (e.g., reducing personal CO2 footprint) or financial considerations (e.g., because of increasing energy costs).

Drivers' key challenge in controlling their operational eco-driving behaviour to maximize energy efficiency is to determine which driving behaviour is the most energy-efficient at different times and then control the vehicle accordingly. Control-theoretical models of driver behaviour assume that continuous monitoring of goal-oriented behaviour is required to successfully manage the driving task (Fuller, 2005; Summala, 2007). This also applies to the driving objective of increasing energy efficiency (Franke, Arend, McIlroy, & Stanton, 2016). The essential basis for this dynamic control process is the perception of relevant environment variables, and, as the perceptibility of energy efficiency via visual or noise cues from the environment is limited, energy displays play a key role for this first step in controlling of energy-related driving behaviour. Based on perception of the energy-related facets while driving via consumption displays, the driving behaviour has to adjust within each driving manoeuvre. Indeed, 52% of interviewed hybrid electric vehicle drivers (Franke, Arend, & Stanton, 2017) monitor consumption displays to derive the energy efficiency level from different driving strategies.

Generally, consumption displays can convey abstract or concrete feedback. In the context of action regulation, these different types of feedback can be attributed to various aspects of the adaptive control of eco-driving selection (see Franke et al., 2016). Given the present research focuses on acquiring eco-driving knowledge and identifying applicable energy-efficient strategies, concrete numeric feedback (e.g. instantaneous consumption displays) is the central aspect. It aims to teach individuals how to accelerate efficiently (manoeuvre-level) because it is constantly obvious which behaviour promotes eco-driving (Dahlinger, Wortmann, Ryder, & Gahr, 2018). In contrast, abstract feedback uses rather unclear, symbolic representations of aggregated information to make the reason salient why someone should drive energy efficiently (Dahlinger et al., 2018). Instantaneous consumption displays (ICDs), in particular, represent the central, immediate and salient information source or system variable within the framework of the monitoring process. By this process, the driver can determine energy efficiency of individual driving manoeuvres. The most basic driving manoeuvre in this respect is speed control through accelerations. Compared to other displays, ICDs provide a large bandwidth of information, a high salience, and a high value due to immediacy. Only ICDs represent the actual acceleration and situation-related influences, while a further aggregation (e.g., average or total consumption) also typically includes manoeuvre-irrelevant information.

However, whether the driver can ultimately derive the situation-specific influence of his or her driving style through ICDs depends on the accuracy of the subjective temporal integration of the consumption parameter presented. Of course, the ultimate objective measure of energy efficiency is a manoeuvre-based average consumption for the particular accelerations. Therefore the key question is how the driver perceives the magnitude dynamics of an ICD and how accurate the aggregated estimates of energy efficiency are based on these dynamics. Surprisingly, little research exists regarding dynamic magnitude perception in the context of eco-driving with a few exceptions in other driving contexts such as speed perception (e.g., Svenson, 1976; Svenson & Salo, 2010). While a snapshot is sufficient to check if a speed limit is adhered to, the dynamic course must be taken into account concerning consumption regulation. Therefore, a complex research agenda is needed to better understand how consumption displays can optimally support the perception of energy-efficient driving styles. Thus, a controlled experiment is needed initially to identify perceptual phenomena or biases that an appropriate display design must address. To investigate the possible influence of dynamic components (e.g. magnitude and time), a study design with sufficient variation of these components is required. The resulting study design must yield sufficient power to detect small to large effects in a multilevel approach. Hence, the initial experiment needs to reflect a kind of ideal state (continuous view) as a baseline, which is comparable to driving on level 2 (driving autonomously with enough time to look at the display). Of course, drivers cannot constantly focus on consumption displays and always perceive them peripherally. Further experiments must address how the estimation error systematically increases when different display variants are embedded in an occlusion paradigm (Gelau, Henning, & Krems, 2009; Gelau & Krems, 2004), for example. This follow-up study should then include a reduced stimuli set with less dynamic component variation but more occlusion parameters and displays. As a next step, displays developed on this basis should be compared in a driving simulator study to investigate the effects on energy-efficient driving behavior. The final goal should be to test the resulting display(s) in the field.

Consequently, by first understanding the possibly biased perception of ICDs, we can identify possible approaches for display design and finally better support drivers and their correct development of mental models (see also Pampel, Jamson, Hibberd, & Barnard, 2015). As drivers' mental models of energy efficient strategies like accelerations differ (e.g., Franke et al., 2016), it is likely that drivers' perception of energy dynamics based on inter-individual difference variables such as knowledge, experience and general cognitive and behavioural styles in interaction with technology differs as well.

Hence, the objectives of the present research are to examine (1) whether drivers can correctly rank the magnitude dynamics in ICDs with regard to energy efficiency, (2) to what extent distortions exist in deriving the manoeuvre-based average consumption from different ICD dynamics, and (3) whether inter-individual difference variables play a role regarding the derivations' quality.

2 BACKGROUND

2.1 Biased magnitude perception in driving

To accurately derive an overall energy efficiency value for a specific driving manoeuvre through monitoring ICDs, a completely rational evaluation would require the integration of instantaneous consumption over a given time. All instantaneous consumption values must be totalled only over acceleration time (manoeuvre-based total consumption) and then divided by the number of values (manoeuvre-based average consumption). It is certainly more likely that human judgements in everyday dynamic environments such as driving will rely more on heuristics (i.e., bounded rationality; Simon, 1957, 1982). Since people have a limited capacity to evaluate and process available information and usually rely on simplifying heuristics when making intuitive judgements, everyday judgements are often biased in various ways (e.g., Tversky & Kahneman, 1973).

One exemplary heuristic in the broader context of magnitude perception is the tendency to automatically assume a linear relationship in various situations when detecting a functional relationship between different values is required. However, this effect can be mediated for example by additional graphic material (Van Dooren, De Bock, Janssens, & Verschaffel, 2008), previous experience with non-linear relationships (Christandl & Fetchenhauer, 2009; Keren, 1983) or changes in displayed units (Eriksson, Patten, Svenson, & Eriksson, 2015; Larrick & Soll, 2008; Peer & Gamliel, 2013). As an example from a driving context, people adopt linear strategies as opposed to considering the curvilinear relationship between speed and travel time (e.g., Peer & Gamliel, 2012, 2013; Svenson, 2008, 2009). Specifically, this means that time saved is underestimated when increasing speed from a relatively low starting point. At a relatively high speed, however, the time saved by a speed increase is overestimated. Similarly, people often assume that the amount of gas consumed relative to the fuel efficiency, when expressed as miles per gallon (MPG), will decrease as a linear function (Larrick & Soll, 2008). However, the actual relationship between the amount of gas consumed and a vehicle's MPG value is curvilinear. This may cause undervaluation of small MPG improvements and overvaluation of higher MPG improvements. Hence, there is a pattern that judgements based on non-linear relationships and on values with different magnitudes are simplified and biased.

Nevertheless, studies examining the integration of dynamic magnitude information over time would be more relevant for the objective of the present research. Unfortunately, to the best of the authors' knowledge, no studies currently exist that examine such dynamics in the context of consumption displays. A first line of research in this direction addresses temporal integration of magnitude information in speed perception. Here, the perception of a vehicle's average speed seems dependant on speed magnitude (Svenson, 1976). Furthermore, a higher speed is given too much weight when judging average speed with different combinations of original and reduced speeds over a certain distance (Svenson & Salo, 2010). However these studies used static scenarios with written descriptions of scenarios/stimuli instead of letting participants experience dynamic situations. Considering the biased calculation of average speeds (Svenson, 1976; Svenson & Salo, 2010), the displayed amount and duration of consumption may also play a role when judging the average over a given time in a dynamic scenario.

As in many domains of human performance and cognitive biases (e.g., Carnevale, Inbar, & Lerner, 2011; Hoppe & Kusterer, 2011; Ingre, Akerstedt, Peters, Anund, & Kecklund, 2006; Tett, Jackson, & Rothstein, 1991) it can further be assumed that considerable inter-individual differences regarding magnitude perception will exist. Unfortunately, inconsistencies exist in the research as some evidence supports (Eriksson & Svenson, 2012) an influence of inter-individual differences in (magnitude) perception biases in driving whereas some does not (Peer & Solomon, 2012; Svenson, 2009). Eriksson and Svenson (2012) asked students and truck drivers to intuitively estimate the average fuel consumption (l/100km) when increasing (70, 80, 90, 100, 110 and 120 km/h) or decreasing (110, 100, 90, 80, 70 and 60 km/h) speed (one speed change per problem). Participants also received the average consumption of the particular reference speed (60 km/h or 120 km/h). Truck drivers underestimated the fuel saving effect of decreasing speeds. While their judgments regarding increasing speeds were more accurate, students overestimated the fuel consumption. Regardless, changing speeds were presented statically and the influence of acceleration on energy consumption was not considered. Nevertheless, it can be assumed that specific experience and knowledge could influence accurate perception of consumption. However, Peer and Solomon (2012) showed different results: Taxi drivers were just as biased as non-professional drivers regarding speed judgements, journey time and time saving. Furthermore, education and training (physics, engineering) also do not seem to reduce bias regarding time savings, accident risk, and the speed while braking when hitting an object (Svenson, 2009). In sum, inter-individual differences may influence several simplified and biased judgements in the context of driving (e.g. speed, consumption in static scenarios, travel time).

2.2 Graphical magnitude perception

Although little research on magnitude perception in driving exists, there is ample research in psychophysics and graphical perception regarding static data visualisation (e.g., Cleveland & McGill, 1984; Falmagne, 1971; Hollands & Spence, 1992; Stevens, 1957). For example, research has shown that height differences are more likely detected in “framed rectangles” compared to when non-framed “bars” are used for visualisation (Weber‘s law; Baird & Noma, 1978; Cleveland & McGill, 1984). Moreover, change can be perceived faster and more accurately in bar and line charts than in pie charts or tiered bar graphs (Hollands & Spence, 1992). However, since ICDs are dynamic rather than static visualisations of data, these findings are not applicable to the perception of ICDs.

To the best of the authors’ knowledge, there seems to be little research regarding animated (i.e. dynamic) data visualisation comparable to static data visualisation. Wu, Jiang, Xu, and Nandi (2016) first examined perceptual accuracy in animated data visualisations and emphasized the need for further controlled studies. They showed that the maximum height of an animated bar (peak height) improved estimation accuracy about the time position of the peak and rate of change of the animation (varying slope). However, higher peaks seemed to increase bar height estimation error (Wu et al., 2016). Previous research on animated data visualisation have focused on different topics such as possible benefits provided by animated data (e.g., Heer & Robertson, 2007), different animation styles (e.g., Merz, Tuch, & Opwis, 2016), animated graphics to teach complex systems (B. Tversky, Morrison, & Betrancourt, 2002) or animated images (Gonzalez, 1996). These topics have been studied in various contexts such as trends (Robertson, Fernandez, Fisher, Lee, & Stasko, 2008), influence on decision making (Gonzalez, 1996) and user experience (Merz et al., 2016) or transitions of static data (Heer & Robertson, 2007). For example, Herr and Robertson (2007) showed that animated transitions of static data visualizations enhance graphical change perception. Thus, the dynamic magnitude perception seems biased. However, the concrete transfer is lacking in the context of driving (or rather energy perception).

3 PRESENT RESEARCH

The present research is part of a more complex research agenda to better understand how the display can optimally support the perception of energy-efficient driving styles. The objective was to examine how varying ICD characteristics influence the derivation of manoeuvre-based average consumption as measure for energy efficiency. In this context, the perception accuracy (estimation difference of average consumption) is considered. Given the limited amount of data and considering the previous research approach regarding graphical perception in animated data visualizations (see Wu et al., 2016), the present research was largely exploratory. The following research questions were examined:

RQ1 – Can drivers rank the various dynamics of a schematic ICD in the correct order of energy efficiency for acceleration manoeuvres?

RQ2 – Do any biases exist in the perception accuracy of energy efficiency based on the dynamic magnitude characteristics?

RQ3 – Do inter-individual difference variables impact the perception accuracy of energy efficiency?

Because it is atypical in driving to accurately estimate the exact manoeuvre-based average consumption when comparing different driving manoeuvres, RQ1 asks for the correct ranking order. RQ2 addresses the possible transfer of previous findings regarding judgement biases in the driving context (see section 2.1) on the perception of consumption. As people showed a peak-height-bias in animated bar charts (Wu et al., 2016) and gave faster speeds too much weight (Svenson, 1976; Svenson & Salo, 2010), a rational assumption is that the peak height influences the estimation difference. In addition, the peak duration at higher peaks is lower to achieve a certain average consumption. Therefore, it could be assumed that an effect of peak duration and an interaction between peak height and peak duration exists.

In RQ3, three inter-individual variables that presumably impact the perception of and interaction with technology are considered: general knowledge, concrete practical experience and stable interaction styles (i.e. personality). Since the influence of experience on judgements of average consumption is unclear (Christandl & Fetchenhauer, 2009; Eriksson & Svenson, 2012; Svenson, 1976; Svenson & Salo, 2010), a direct and specific facet is considered with respect to system interaction: experience with consumption displays. Technical-mathematical knowledge is also considered a variable since system knowledge influences the perceived strategy effectiveness (i.e. driving strategy, Franke et al., 2016) and is acquired with system interaction (Franke, Arend, & Stanton, 2017). Furthermore, it is particularly important if someone is a novice (see Rasmussen, 1983) regarding energy efficient driving. This is true for novice drivers as well as people new to certain types (e.g. electric vehicles). Technical-mathematical knowledge is considered a potential factor because it could promote the comprehension of data aggregation over time. This comprehension is essential in estimating average consumption. Besides experience and knowledge, the affinity for technology interaction (ATI) scale serves as an exemplary facet of general cognitive and behavioural styles in technology interaction. The scale is developed based on a review of literature between 1982 and 2016 on individual personality traits that influence interacting with new technical systems (Attig, Wessel, & Franke, 2017). In the broader context of biases regarding non-linear relationships, need for cognition (Cacioppo & Petty, 1982) and therefore a stronger motivation to solve estimation tasks for example seem to positively influence the underestimation of exponential growth (Christandl & Fetchenhauer, 2009). A high ATI as a dimension of need for cognition and therefore more actively exploring systems (in this case, consumption displays) may lead to better estimations based on information aggregation.

4 METHOD

4.1 Participants

A total of 58 participants were recruited via a psychology student e-mail distribution list and via social media. Three participants were excluded because they mistook the ICD for a display of speed. Participants in the resulting sample (N = 55) were an average age of M = 22.55 years (SD = 3.84). Within the sample, 86% were female and 87% were psychology students. Their average driving experience was 4.68 years (SD = 3.36 years) and 52.98 km (SD = 97.21 km) per week.

4.2 Stimulus material

To examine the perception of ICDs during acceleration, schematic sequences of energy consumption during periods of acceleration were created. The acceleration's basic elements include an initial consumption rate at a constant start speed, an end consumption rate at a constant target speed and a sharp increase of consumption rate depending on acceleration intensity and speed. Despite the schematic construction of the sequences, considerable efforts were invested to create similar energy dynamics to real-world consumption in electric vehicles. Specifically, we used Galvin's (2017) work to define realistic values. According to Galvin (2017), the consumption of a typical compact electric vehicle (i.e. Kia Soul Electric 2015) is 11.1 kWh/100km at a constant speed of 20km/h, 13.4 kWh/100km at a constant speed of 80 km/h, and 73 kWh/100km within a typical acceleration (1.35 m/s^2-1.4 m/s^2 within a speed range of around 50-80km/h; Galvin, 2017). Hence, the values' ratio used in the schematic sequences created for the present experiment is comparable to real-life energy consumption of battery electric vehicles.

Twenty-five sequences lasting 5 seconds and with a temporal resolution of 100 Hz were created, forming five groups of different average consumption values (32, 34, 36, 38, 40) with five different peak values each (100, 90, 80, 70, 60). The consumption unit was fictional to enable broader generalisation. All sequences began with the same starting consumption value (6) and ended on the same value as well (10). Peaks always began at 1.1 seconds and continued for the needed duration to achieve the targeted average consumption of the sequence given the defined peak height. Hence, the higher the peak, the shorter the sequence of the peak had to be to result in the same average consumption. To increase realism and to ensure a comparable rate of change, a Lowess smoothing (Cleveland, 1979) with a smoother span of f = .15 was conducted for all sequences. Furthermore, noise (±0 or ±0.5 with a respective probability of 50%) was added to the constant consumption phases to underline the ongoing dynamic process and increase the simulation authenticity (see Figure 1). All sequences varied in peak duration (see Table 1). The peak durations include the display time of the respective average peak value (±0.5).

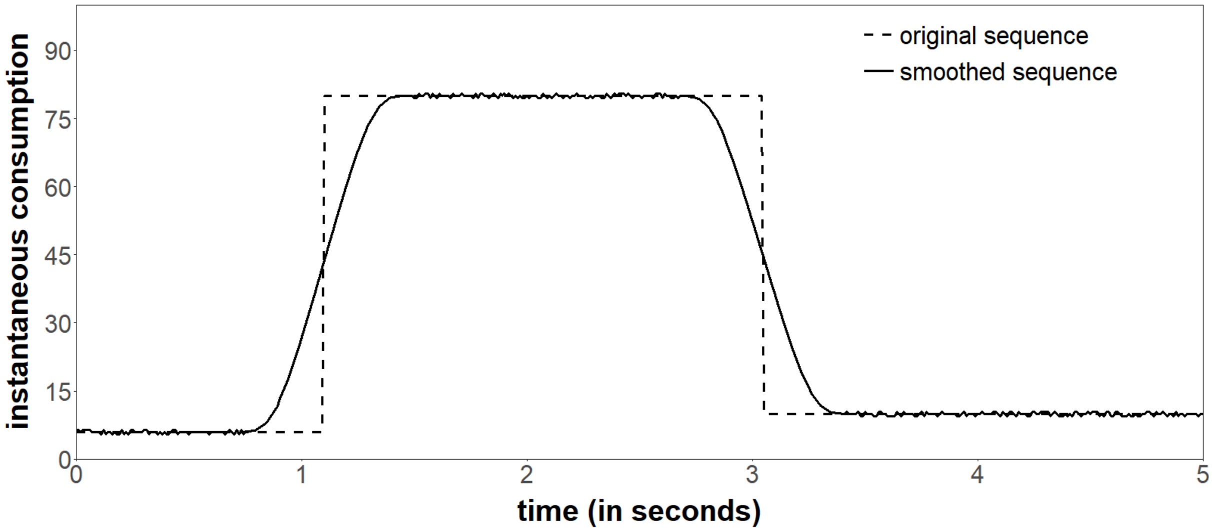

Figure 1. Example of one smoothed sequence after adding noise (average peak height = 80, average consumption = 36).

Table 1.
Sequence design

Sequence ID	Average consumption	Average peak height	Peak duration (in s) *
22	32	60	1.75
18	32	70	1.35
14	32	80	1.08
10	32	90	0.86
6	32	100	0.69
23	34	60	1.94
19	34	70	1.52
15	34	80	1.21
11	34	90	0.99
7	34	100	0.80
5	36	60	2.15
4	36	70	1.69
3	36	80	1.36
2	36	90	1.12
1	36	100	0.91
24	38	60	2.35
20	38	70	1.86
16	38	80	1.51
12	38	90	1.25
8	38	100	1.03
25	40	60	2.55
21	40	70	2.03
17	40	80	1.65
13	40	90	1.37
9	40	100	1.14

Notes. *peak duration refers to the smoothed sequences and an average peak (±0.5)

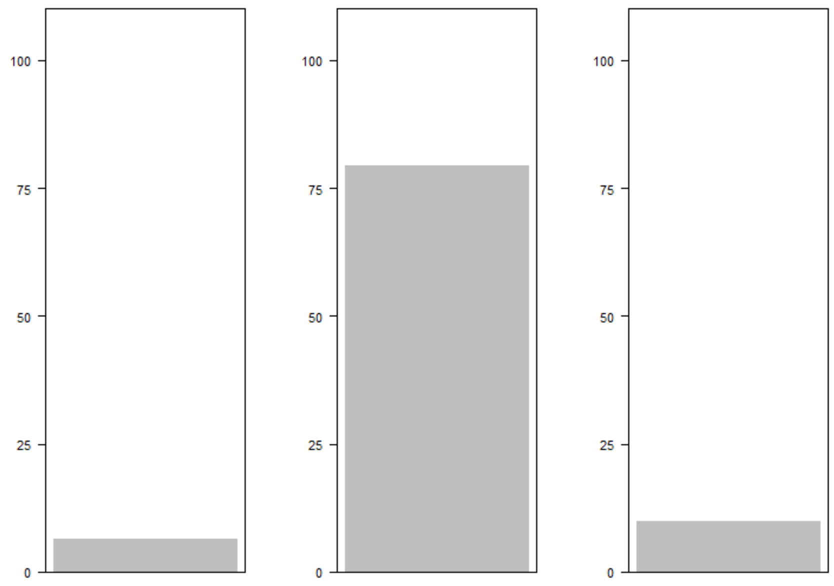

Figure 2. Video sections for sequence ID 3 with 0.01 seconds, 1.60 seconds and 4.99 seconds (see https://youtu.be/SDsMO85Ejwg for the animated ICDs).

RStudio and ffmpeg were used to create the schematic ICD videos in the form of an animated bar plot (see Figure 2). The bar plot's frame size in the presented videos on LimeSurvey was 342 x 114 pixel (height x width). Just prior to the actual sequence, a fixation cross was displayed in the middle of the video screen for 1.5 seconds. After the sequence occurred, a white screen was shown.

4.3 Procedure

The experiment was conducted online with LimeSurvey in an uncontrolled setup. As a general scenario, participants were asked to imagine a traffic situation in which they accelerate the vehicle after the end of a previous speed limit. An example sequence including the average parameters of all sequences (average consumption = 36, peak = 80) was presented to explain the dynamics of consumption (start consumption level, maximum consumption level, end consumption level) and to familiarize participants with the ICD videos. The participants were told that the consumption unit was fictional. This was due to avoid comparisons (e.g. to a more familiar combustion engine) and irritations through deviations from real and more familiar sequences.

The participants were instructed to estimate the average consumption level for each sequence: “*Please estimate the average consumption level on a scale from 0 to 100 consumption units as accurately as possible*”. Afterwards, 2 blocks of 25 sequences each were presented, each block starting with the same first sequence (ID = 1). The further 24 sequences were randomized within the blocks.

After completing the experimental trials, participants answered an open-ended question in the following post-experimental questionnaire to determine which cues they based their estimations on and if they understood the task correctly. Furthermore, participants gave socio-demographic information, their experience with consumption displays, their technical-mathematical knowledge and their ATI. Participation in the entire online study lasted an average of 30 minutes.

4.4 Scales and measures

According to common practice (e.g., Cripps, 2017), Cronbach’s alpha was interpreted as poor (.5 ≤ α < .6), questionable (.6 ≤ α < .7), acceptable (.7 ≤ α < .8), good (.8 ≤ α < .9), or excellent (≥ .9) for all measures. Scales in the post-experimental questionnaire contained a 6-point Likert scale ranging from 1 = " completely disagree” to 6 = "completely agree", if not stated otherwise.

4.4.1 Estimation difference

For the dependent variable, the difference between the participants’ empirical estimates and the correct average consumption levels was calculated as a measure of perception accuracy/estimation error (0 = perfect estimate, >0 overestimation, <0 underestimation of average consumption levels). Test-retest reliability of the mean estimation difference between block 1 and 2 was excellent (α = .97). Even when considering individual sequences, test-retest reliabilities were all acceptable (α > .7) with an average α of .89. The mean estimation difference was *M* = 8.51 (*SD* = 14.48, range = 93).

4.4.2 Experience with consumption displays

Experience with consumption displays was assessed via eight items (see Table A1) focusing on attention directed at displays, helpfulness, and specific as well as general relevance of displays to achieve fuel (i.e. energy and gas) efficiency. The reliability of the means of all item values was excellent ($\alpha = .92$). The average experience with consumption displays was $M = 3.85$ ($SD = 1.05$, range = 4.75), indicating that the participants were neither very experienced nor very unexperienced. Therefore, the conditions were optimal, as participants were not exclusively expert or novice (i.e. variance of experience is not limited).

4.4.3 Affinity for technology interaction

The ATI scale (Franke et al., 2018) consists of nine items. The ATI score is computed as the mean score of all nine items (Items 3, 6, 8 reversed). The internal consistency of the ATI scale was excellent ($\alpha = .90$). The average ATI in the sample was $M = 3.21$ ($SD = 1$, range = 4.22).

4.4.4 Technical and mathematical knowledge

The participant's self-rated technical and mathematical knowledge was assessed by five items (see Table A1). Internal consistency of the knowledge scale was acceptable ($\alpha = .77$). The average score of knowledge was $M = 2.74$ ($SD = 0.97$, range = 4.83).

5 RESULTS

The significance level was set to $\alpha = .05$ for all analyses. Based on Arend and Schäfer (2018), we estimated that the present study design yielded sufficient power (i.e., >= .80; cf. Cohen, 1988) to detect small to medium L1 direct effects in a multilevel approach. Similarly, we assumed medium to large effects for the L2 direct and cross-level interaction effects.

5.1 RQ1 - Can drivers rank the various dynamics of a schematic ICD in the correct order of energy efficiency for acceleration manoeuvres?

Estimation differences were not normally distributed as assessed by the Shapiro-Wilk-Test ($p < .05$). Therefore, a Kendall rank correlation between the empirical and the correct ranks were computed for each block to test RQ1. Significant correlations were non-existent in both block 1 and block 2 ($r_τ = .16$, $p = .293$ and $r_τ = .13$, $p = .389$). Thus, it seems participants did not correctly order the ICD scenarios.

In addition, open-ended comments from 15 participants indicate that heuristics were used without accounting for the dynamic process (e.g., "Comparison of the starting consumption value with the difference of the consumption value during acceleration", "Smallest and highest value, calculating the mean", "Half of the consumption value during acceleration (when the bar moved up)"). Although 35 participants attempted to account for time (e.g. "Smallest and highest value, calculating the mean"), it seems they failed to consider the dynamic rise (e.g. "highest level duration, lowest level duration, highest level value, lowest level value"). One participant mentioned considering primarily the value ("Value (primary) and duration"). Another mentioned changing his strategy over time: first taking value and duration into account, then using a simplifying heuristic ("At first, I paid attention to duration and amount of consumption. After some videos, […] I divided the highest value in half […]"). Comments from three participants could not be clearly assigned to a category. The mentioned heuristic calculations (e.g., peak/2, (minimum consumption value + peak)/2, (start consumption value + end consumption value + peak)/2) result in a different ranking than the correct ranking. We computed a Kendall rank correlation between the empirical and heuristic ranks for each block, showing significance in both block 1 and block 2 ($r_τ = .91$, $p < .001$).

In sum, although the empirical ranking does not reflect the correct ranking, it better fits a heuristic ranking. A possible conclusion is that ICDs do not allow for correct recognition of the optimal strategy regarding a specific driving manoeuvre.

5.2 RQ2 - Do any biases exist in the perception accuracy of energy efficiency based on the dynamic magnitude characteristics?

To test RQ2, the mean estimation difference was compared to perfect estimation (estimation difference = 0). Prior testing determined whether a difference between the two blocks existed because of possible variations due to motivational changes, learning effects or other reasons. A Wilcoxon signed-rank test indicated a significantly smaller mean estimation difference in block 1 (*Mdn* = 3.76) than in block 2 (*Mdn* = 5.44), W = 1034.00, z = 2.21, p = .027, with a small effect size (r = .21; Rosenthal, 1994). Therefore the participants' estimation was more accurate in block 1 than in block 2. Furthermore, a one-sample Wilcoxon signed-rank test showed that the mean estimation differences for block 1 and block 2 were significantly higher than 0 (W = 1425.50, z = 5.49 and W = 1368.00, z = 5.01, p < .001). The effect sizes were r = .52 and r = .48 respectively, corresponding to large and medium effects (Rosenthal, 1994). This shows that the average consumption was overestimated in both blocks.

Following Leckie (2013) as well as Snijders and Bosker (2012), a multilevel model was created to examine possible effects (peak height, peak duration, interaction between peak height and peak duration). Thus far, multilevel models have seldom been employed in traffic psychology, human factors and ergonomics with a few exceptions. A modest number of published articles between 2013 and November 2019 incorporate multilevel models (search for "multilevel model" and screen relevant articles) when compared to the number of all published articles in the particular journal during this time period: < 1.42 % in "Ergonomics" (e.g., Jung, Kaß, Schramm, & Zapf, 2017), "Applied Ergonomics" (e.g., Hiemstra-van Mastrigt, Kamp, van Veen, Vink, & Bosch, 2015), "Computers in Human Behavior" (e.g., Kushlev, Hunter, Proulx, Pressman, & Dunn, 2019) and in "Transportation Research Part F: Traffic Psychology and Behaviour" (e.g., Molnar et al., 2018). The central advantage of multilevel approaches is that they consider hierarchical or clustered structures in the data, thereby avoiding underestimation of standard errors of regression coefficients in contrast to multiple regression, for example. Ignoring the multilevel data structure may also lead to employing statistical procedures with violated independence assumptions (e.g., dependent observations from the same individual) and unfounded conclusions (for further details, see Snijders & Bosker, 2012).

With a three-level model, the clustered data structure, comprising participants at the highest level (L3), peak height groups at the medium level (L2) and measures at the lowest level (L1) was used to analyse the data. There are 55 participants at L3, (5 peaks x 55 participants =) 275 peak height groups at L2 and (5 observations per peak height group x 2 blocks x 275 peak height groups =) 2750 measures at L1.

First, it is important to test if the empty three-level model fits the data significantly better than simpler two-level models and the single-level model (Leckie, 2013). A likelihood ratio test showed that the empty three-level model M0 (see table 2) was preferred to its single-level counterpart L0 (χ^2_2 = 3959.36, p < .001). To test the null hypothesis that no effects of peak height group exist, M0 was compared to the simpler two-level measures-within-participants model T1. M0 was preferred to T1 (χ^2_1 = 1719.17, p < .001).

Table 2.
Models

Model	Equation
L0	$Y_i = \beta_0 + e_i$
T1	$Y_{ij} = \beta_{0j} + V_{0j} + e_{ij}$
T2	$Y_{ip} = \beta_{0p} + U_{0p} + e_{ip}$
M0	$Y_{ipj} = \beta_{0pj} + V_{0pj} + U_{0pj} + e_{ipj}$
M1	L1: $Y_{ipj} = \alpha_{0jk} + \alpha_{1jk}\, duration + e_{ipj}$
	L2: $\alpha_{0jk} = \gamma_{00j} + \gamma_{01j}\, peak + U_{0pj}$
	$\alpha_{1jk} = \gamma_{10j} + U_{1pj}$
	L3: $\gamma_{00j} = \beta_{000} + V_{00j} + \beta_{110}\, duration \times peak$
	$\gamma_{01j} = \beta_{010} + V_{01j}$
	$\gamma_{10j} = \beta_{100} + V_{10j}$
	$Y_{ipj} = \gamma_{000} + \beta_{100}\, duration + \beta_{010}\, peak + \beta_{110}\, duration \times peak + V_{01j}\, peak + V_{10j}\, duration + U_{1pj}\, duration + V_{00j} + U_{0pj} + e_{ipj}$

To test the null hypothesis that there are no participant effects, M0 was compared to the simpler two-level measures-within-peak-height-group model T2. M0 was also preferred to T2 (χ^2_1 = 3452.39, p < .001). A multilevel analysis approach was clearly favoured over a single-level (i.e. simple regression analysis) and two-level approach.

The empty three-level model reveals the raw within-group and between-group variances, which are useful as a general description and a starting point for further model fitting (Snijders & Bosker, 2012). Calculating the variance partition coefficients and the intraclass correlation helps to determine whether a multilevel model is necessary and to show the degree of data clustering (Leckie, 2013). A total of 54.5% of the variation in estimation difference lay between participants, 28.7% lay within participants between the different peak height groups, while the remaining 16.8% lay within peak groups between the estimation differences (see variance partition coefficient in Table 3). Thus, there was substantial variation between participants and a variation between peak height groups. Finally, the participants ICC is .545, while the peak height group ICC is .832. Although ICC values are usually small in practice (Musca et al., 2011), rather large ICC values are expected in a repeated measures design (Arend & Schäfer, 2018). Overall, an ICC of .50 can be considered large (Arend & Schäfer, 2018). As even very small ICCs can dramatically influence Type-I error rate in standard one-level analyses for independent data, a multilevel approach is favoured, where the variation between groups is part of the model (Musca et al., 2011).

Table 3.
VPC and ICC statistics for the three-level variance components model for estimation difference measures

	VPC	ICC
Participants	.545	.545
Peak height group	.287	.832
Measures	.168	

Notes. VPC = variance partition coefficient, ICC = intraclass correlation coefficient.

The predictor variables (peak height, peak duration), their cross-level interaction as well as their random slopes were added stepwise to model M1 (see Table 2), corresponding to the respective likelihood ratio tests. At L2, there were random slopes for peak height and duration, whereas at L3, they only existed for peak height. M1 showed a significantly better fit to the data than the empty three-level model (χ^2_{10} = 816.10, p < .001, R^2_{marginal} = .19 (proportion of variance explained by the fixed factors alone), $R^2_{\text{conditional}}$ = .86 (proportion of variance explained by both fixed and random factors)). The R^2 values were computed according to Nakagawa and Schielzeth (2013). The fixed effects of peak height (β = 0.39, *SE* = 0.04) and peak duration (β = -9.92, *SE* = 0.73) as well as the interaction (β = -0.18, *SE* = 0.02; see figure 3) were significant (t(137.11) = 8.82, t(64.95) = -13.57 and t(614.82) = -8.12, p < .001). The estimation difference is larger with greater peak height and shorter peak duration. The effect of peak height on estimation difference becomes weaker with longer peak duration. In sum, perception accuracy of energy efficiency is biased based on the dynamic magnitude characteristics (peak height and peak duration).

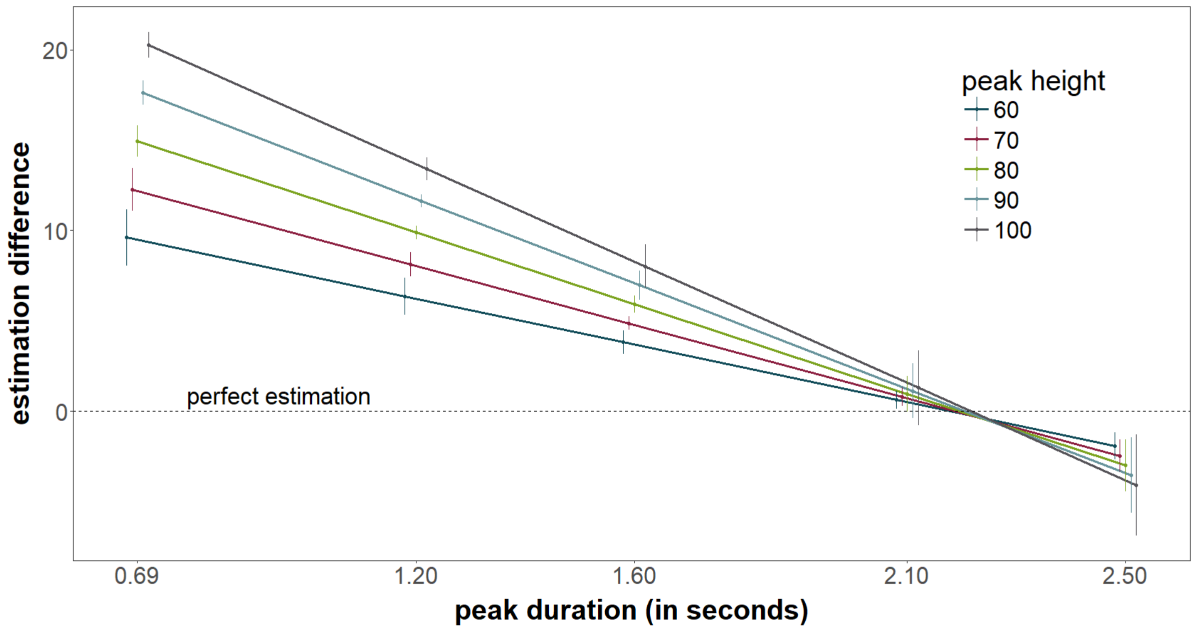

Figure 3. Cross-level interaction between peak duration and peak height. Standard errors are offset.

5.3 RQ3 - Do inter-individual difference variables impact perception accuracy of energy efficiency?

To identify potential inter-individual influencing factors in an exploratory manner, the cross-level-interactions of each person-level predictor with peak height and peak duration were considered individually in 9 total models (see Table A2 for the detailed results). To avoid erroneous interpretations, the fixed effects were also included in each case regardless of significance (Snijders & Bosker, 2012). The cross-level-interactions between ATI and the L2 predictor peak height as well as the L1 predictor peak duration were significant ($p < .05$). This means that the effect of peak height and peak duration on estimation difference become weaker with higher ATI. Additionally, the interaction between peak height, peak duration and ATI was significant ($p < .05$). ATI influences the peak-height-bias (greater estimation difference with greater peak heights) which decreases with higher peak duration (see Figure 4). With high ATI, the effect of peak height on the estimation difference decreases faster with increasing peak duration when compared to a small ATI. Further, it can be surmised that there is a peak-height-bias towards underestimation depending on peak duration (i.e. acceleration duration) after reaching a certain peak duration (greater underestimation with taller peak heights).

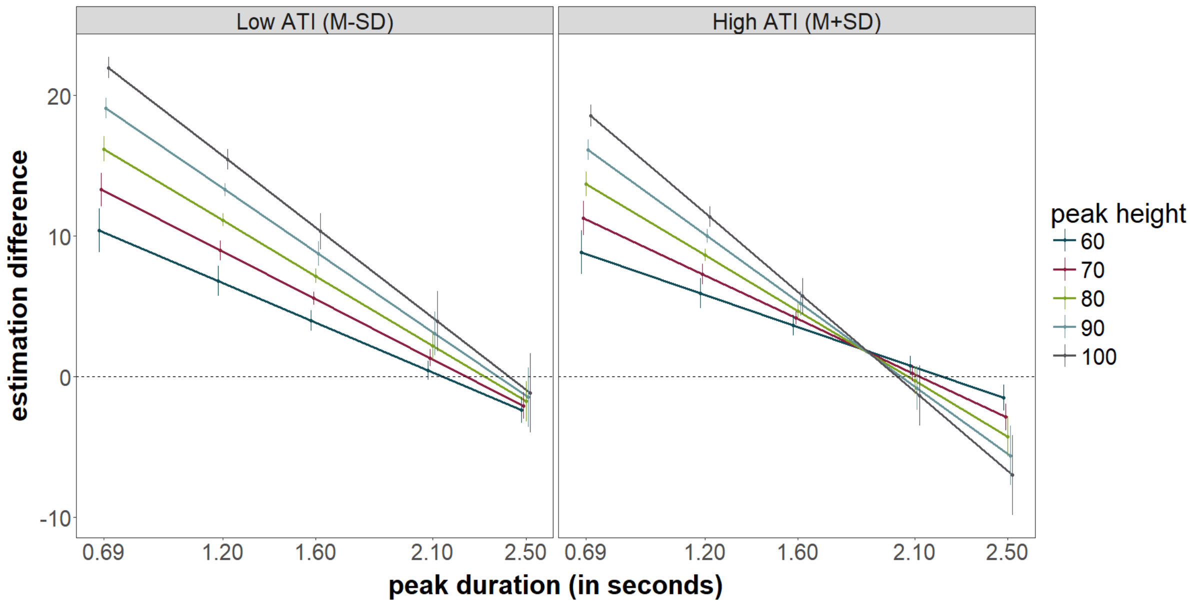

Figure 4. Cross-level interaction between peak duration and peak height facetted by ATI ($M \pm SD = 3.21 \pm 1$). Standard errors are offset.

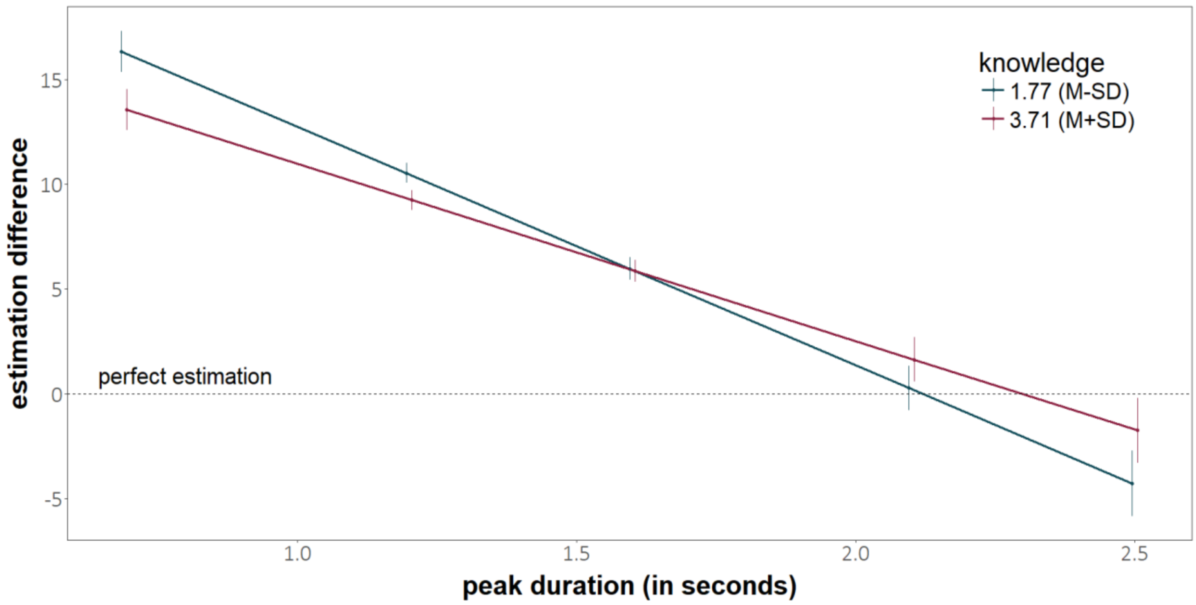

Figure 5. Cross-level interaction between peak duration and knowledge. Standard errors are offset.

In addition to the significant effect of ATI, the interaction between knowledge and peak duration was significant (p < .05, see Figure 5). The effect of peak duration on estimation difference is weaker with enhanced knowledge. Overall, knowledge and ATI seem to impact estimation biases, whereas experience with consumption displays seem irrelevant.

6 DISCUSSION

6.1 Summary of results and theoretical implications

The objective of the present study was to examine whether drivers can accurately determine efficiency differences of accelerations based on perceiving dynamic ICD sequences. People tend to overestimate the average consumption level when the maximum consumption value is higher and is displayed for less time (i.e. the shorter the acceleration).

The overestimation is comparable to judgements of average speed that depend on speed amount (Svenson, 1976; Svenson & Salo, 2010). The present results also expand Wu et al.'s (2016) findings regarding the lack of accuracy in estimating higher peaks. Thus, drivers cannot correctly integrate information on time and magnitude into their average consumption judgements and therefore into their perception of energy efficient driving strategies. A possible explanation for the peak-height-bias is that the higher values could be more easily available (availability heuristic; A. Tversky & Kahneman, 1973). In addition, open-ended comments from participants indicate that they incorporated simplifying heuristics without accounting for the dynamic process (e.g., adding peak and minimum weight divided by 2) or without considering the dynamic rise. The mentioned heuristic calculations (e.g., peak/2, (minimum consumption value + peak)/2, (start consumption value + end consumption value + peak)/2) result in a ranking that differs from the correct ranking and significantly correlates with the empirical ranking. The heuristics also fit a real driving situation well, in which only "snapshots" of the dynamic process are perceived.

Furthermore, the overestimation was higher in block 2 than block 1. This could have been due to several different reasons. The participants may have become less motivated over the course of the experiment. It is also possible that participants tended to increasingly use simplifying heuristics. One open-ended comment supported this explanation ("At first, I paid attention to duration and amount of consumption. After some videos, […] I divided the highest value in half […]").

Further, the peak-height and peak-duration-bias apparently depend on inter-individual difference variables such as ATI (Franke et al., 2018). As ATI is a personal resource for coping with technology as well as an interaction style rooted in the construct need for cognition (Cacioppo & Petty, 1982), the latter may fundamentally influence the biased dynamic perception of consumption displays. Therefore, drivers with higher tendencies to cognitively engage with systems might be less biased because of actively exploring consumption displays.

Knowledge influences the perceived strategy effectiveness (Franke et al., 2016) and reduces the peak-duration-bias. However, experience with consumption displays does not impact any bias. This corroborates findings regarding other estimation biases in the driving context (Peer & Solomon, 2012; Svenson, 2009). However, knowledge and experience with consumption displays were self-rated in the present study instead of objectively measured.

In addition to the overestimation, monitoring of simple ICDs does not lead to a correct rankings of different acceleration sequences regarding their energy efficiency. The incorrect ordering might be due to heuristic estimations and the non-consideration of time or rather the dynamic process. Estimations based on only peak, peak and start consumption values or peak, start and end consumption values better fit the empirical ranking. If drivers are unable to identify the most energy efficient strategies, they obviously cannot select and apply them. Future experiments could include a pairwise comparison or consecutively presented sequences with an ordering task ("Which sequence was the most energy efficient?") as indirect alternatives to consumption estimations.

6.2 Design Implications

Since judgment biases can be countered (e.g., Eriksson, Patten, Svenson, & Eriksson, 2015; Larrick & Soll, 2008; Peer & Gamliel, 2013) and interfaces can support drivers during trips (e.g., Lundström & Bogdan, 2017), a fruitful strategy based on the present research is to develop and examine different variants of consumption displays to improve perception of energy efficient driving manoeuvres. It remains unclear if differences in ICD shape (e.g., bar, radial) and other display modes (e.g., digit vs. graphical, scale legend vs. no scale legend, normal scale vs. shrinking scale) could influence how magnitude changes are perceived. Shrinking small values (compared to larger values) could emphasize their proportion within one acceleration manoeuvre and therefore reduce overestimation. Nevertheless, it suggests that simple ICDs are insufficient to optimize energy efficiency. Besides the high perceptual workload and distraction potential, aggregating magnitude information with duration remains the main problem regarding eco-driving strategy selection.

Information availability could be improved by providing previous information via a fading trace reflecting magnitude changes through different transparencies and width. Some interfaces already combine graphs of remaining range (in km), consumption over the last (x) km driven, or average consumption during the last (x) minutes, providing more information than a real-time snapshot (e.g., Tesla Model S). This approach can provide a clear reference period and encourage the driver to consider magnitude information over time. The most obvious solution for the integration problem is to compute average consumption for each identified driving manoeuvre or situation. This could improve a manoeuvre-specific selection of strategies and focus on the increased energy loss during accelerations due to the energy/distance metric. Using a metric other than energy/distance might be also useful as accelerations should occur at the highest possible conversion efficiency of electric energy invested per kinetic energy gained (for a possible design and further discussion, see Franke, Görges, & Arend, 2019).

6.3 Limitations and further research

Different from Wu et al.'s (2016) more mathematical-quantitative approach, the present study instead compared a defined set of visualizations. Effects for smaller (< 60) or higher (> 100) peaks as well as longer peak durations remain unclear. Also, the given scale (0 to 100) may have created confounded estimations.

Furthermore, it must be mentioned that the videos were larger than comparable displays in vehicles. Follow-up studies should closely examine the judgement biases influenced by visual angle. Moreover, the ecological validity of the current study is rather low given it occurred in an artificial setting. Participants may have employed the simplifying approach due to low motivation or increasing fatigue. As the setting was uncontrolled, it is also possible that participants were increasingly distracted. Thus the setting should be controlled in further experiments. Besides, this cause could also suggest that in a more complex and real driving situation simplifying heuristics are increasingly used to better balance the limited resources of attention. Further studies should consider actual driving behaviour as a dependent variable by testing different displays in the driving simulator. Likewise, it is important to obtain a more representative sample (age, driving experience, ATI) to make definitive statements about the impact of knowledge and experience. The present sample does not perfectly match the general population regarding ATI (*M* = 3.21) as the average ATI score in the population has been reported to be around 3.5 (Franke et al., 2018). The present study centred on general psychological (perceptual) phenomena and thus systematic gaze strategies were not examined. Thus, driving experience was not a factor in the present study. Future research should examine gaze strategies as they likely differ between experts and novices. Nevertheless, it must be noted that expertise with consumption displays does not necessarily correlate with driving experience as some experienced drivers may rarely use the consumption display. Likewise, some with little driving experience may often use the display. In addition, experience with consumption displays had no influence on the cognitive biases in the present study.

It is not the subject of our research, but abstract feedback could be more appropriate with higher expertise in energy-efficient driving styles, because motivation is then at the crucial level. In this case, an ecodriving learning mode (concrete feedback) and an ecodriving expert mode (abstract feedback) could be integrated. Nevertheless, this assumptions are too dissociated from the present research results. But it might be interesting to investigate the effect of different levels of ecodriving expertise as well as different feedback modes on perception biases, driving behaviour and gaze strategies

The present study is only an initial step towards a more complex research agenda. The following step would be embedding the estimation task that included different display variants within an occlusion paradigm (Gelau et al., 2009; Gelau & Krems, 2004). It is possible to compare various factors on presentation level (shape, size, colour) or completely different display variants (e.g. fading yes / no, additional aggregated indicator yes / no, etc.). It is expected that in a real and more demanding driving situation, the identified biases will have an even stronger influence as heuristics were already used in the present experiment. More precisely, only perceiving display "snapshots" (in the direct or indirect field of view) might encourage heuristic estimations without accounting for the dynamic process.

6.4 Conclusion

To conclude, a simple visualization of instantaneous consumption may be insufficient to determine the actual efficiency of different acceleration strategies. The present study revealed biases in perception of energy efficiency driving manoeuvers and potential for improvement. Therefore, the present research laid the foundation and can serve as a basis for further studies investigating the perception of dynamic data in general, as well as the effect of different dynamics of consumption sequences in particular.

7 REFERENCES

Arend, M. G., & Schäfer, T. (2018). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods*. https://doi.org/10.1037/met0000195

Attig, C., Wessel, D., & Franke, T. (2017). Assessing personality differences in human-technology interaction: An overview of key self-report scales to predict successful interaction. In C. Stephanidis (Ed.), *HCI International 2017 – Posters' Extended Abstracts, Part I, CCIS 713* (pp. 19–29). Cham, Switzerland: Springer International Publishing AG.

Baird, J. C., & Noma, E. J. (1978). *Fundamentals of scaling and psychophysics*. New York: Wiley.

Barkenbus, J. N. (2010). Eco-driving: An overlooked climate change initiative. *Energy Policy*, *38*(2), 762–769. https://doi.org/10.1016/j.enpol.2009.10.021

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Carnevale, J. J., Inbar, Y., & Lerner, J. S. (2011). Individual differences in need for cognition and decision-making competence among leaders. *Personality and Individual Differences*, *51*, 274–278.

Christandl, F., & Fetchenhauer, D. (2009). How laypeople and experts misperceive the effect of economic growth. *Journal of Economic Psychology*, *30*(3), 381–392. https://doi.org/10.1016/j.joep.2009.01.002

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 9.

Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, *79*(387), 531–554. https://doi.org/10.1080/01621459.1984.10478080

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: L. Erlbaum Associates.

Dahlinger, A., Wortmann, F., Ryder, B., & Gahr, B. (2018). The Impact of Abstract vs. Concrete Feedback Design on Behavior Insights from a Large Eco-Driving Field Experiment. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 379:1–379:11. https://doi.org/10.1145/3173574.3173953

Dogan, E., Steg, L., & Delhomme, P. (2011). The influence of multiple goals on driving behavior: The case of safety, time saving, and fuel saving. *Accident Analysis & Prevention*, *43*(5), 1635–1643. https://doi.org/10.1016/j.aap.2011.03.002

Eriksson, G., Patten, C. J. D., Svenson, O., & Eriksson, L. (2015). Estimated time of arrival and debiasing the time saving bias. *Ergonomics*, *58*(12), 1939–1946. https://doi.org/10.1080/00140139.2015.1051592

Eriksson, G., & Svenson, O. (2012). Driving faster or slower? Biased judgments of fuel consumption at changing speeds. In N. A. Stanton (Ed.), *Advances in Human Aspects of Road and Rail Transportation* (pp. 293–297). CRC Press.

Falmagne, J. C. (1971). The generalized Fechner problem and discrimination. *Journal of Mathematical Psychology*, *8*(1), 22–43. https://doi.org/10.1016/0022-2496(71)90021-6

Franke, T., Arend, M. G., McIlroy, R. C., & Stanton, N. A. (2016). Ecodriving in hybrid electric vehicles – Exploring challenges for user-energy interaction. *Applied Ergonomics*, *55*, 33–45. https://doi.org/10.1016/j.apergo.2016.01.007

Franke, T., Arend, M. G., & Stanton, N. A. (2017). How Do Hybrid Electric Vehicle Drivers Acquire Ecodriving Strategy Knowledge? *Engineering Psychology and Cognitive Ergonomics: Cognition and Design*, 363–374. https://doi.org/10.1007/978-3-319-58475-1_27

Franke, T., Attig, C., & Wessel, D. (2018). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction*, *0*(0), 1–12. https://doi.org/10.1080/10447318.2018.1456150

Franke, T., Görges, D., & Arend, M. G. (2019). The Energy Interface Challenge. Towards Designing Effective Energy Efficiency Interfaces for Electric Vehicles. *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 35–48. https://doi.org/10.1145/3342197.3344526

Fuller, R. (2005). Towards a general theory of driver behavior. *Accident Analysis and Prevention*, *37*(3), 461–472.

Galvin, R. (2017). Energy consumption effects of speed and acceleration in electric vehicles: Laboratory case studies and implications for drivers and policymakers. *Transportation Research Part D: Transport and Environment*, *53*, 234–248. https://doi.org/10.1016/j.trd.2017.04.020

Gelau, C., Henning, M. J., & Krems, J. F. (2009). On the reliability of the occlusion technique as a tool for the assessment of the HMI of in-vehicle information and communication systems. *Applied Ergonomics*, *40*(2), 181–184. https://doi.org/10.1016/j.apergo.2008.04.008

Gelau, C., & Krems, J. F. (2004). The occlusion technique: a procedure to assess the HMI of in-vehicle information and communication systems. *Applied Ergonomics*, *35*(3), 185–187. https://doi.org/10.1016/j.apergo.2003.11.009

Gonzalez, C. (1996). Does Animation in User Interfaces Improve Decision Making? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 27–34. https://doi.org/10.1145/238386.238396

Heer, J., & Robertson, G. (2007). Animated Transitions in Statistical Data Graphics. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1240–1247. https://doi.org/10.1109/TVCG.2007.70539

Hiemstra-van Mastrigt, S., Kamp, I., van Veen, S. A. T., Vink, P., & Bosch, T. (2015). The influence of active seating on car passengers' perceived comfort and activity levels. *Applied Ergonomics*, *47*, 211–219. https://doi.org/10.1016/j.apergo.2014.10.004

Hollands, J. G., & Spence, I. (1992). Judgments of change and proportion in graphical perception. *Human Factors*, *34*(3), 313–334.

Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, *110*(2), 97–100. https://doi.org/10.1016/j.econlet.2010.11.015

Ingre, M., Akerstedt, T., Peters, B., Anund, A., & Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. *Journal of Sleep Research*, *15*(1), 47–53. https://doi.org/10.1111/j.1365-2869.2006.00504.x

Jung, T., Kaß, C., Schramm, T., & Zapf, D. (2017). So what really is user experience? An experimental study of user needs and emotional responses as underlying constructs. *Ergonomics*, *60*(12), 1601–1620. https://doi.org/10.1080/00140139.2017.1341555

Keren, G. (1983). Cultural differences in the misperception of exponential growth. *Perception & Psychophysics*, *34*(3), 289–293. https://doi.org/10.3758/BF03202958

Kushlev, K., Hunter, J. F., Proulx, J., Pressman, S. D., & Dunn, E. (2019). Smartphones reduce smiles between strangers. *Computers in Human Behavior*, *91*, 12–16. https://doi.org/10.1016/j.chb.2018.09.023

Larrick, R. P., & Soll, J. B. (2008). The MPG Illusion. *Science*, *320*(5883), 1593–1594. https://doi.org/10.1126/science.1154983

Leckie, G. (2013). Three-Level Multilevel Models. LEMMA VLE Module 11, University of Bristol, Centre for Multilevel modeling. Retrieved 21 August 2018, from http://www.bristol.ac.uk/cmm/learning/course.html

Lundström, A., & Bogdan, C. (2017). Design to Support Energy Management for Electric Car Drivers. In G. Meixner & C. Müller (Eds.), *Automotive User Interfaces: Creating Interactive Experiences in the Car* (pp. 121–141). https://doi.org/10.1007/978-3-319-49448-7_5

Merz, B., Tuch, A. N., & Opwis, K. (2016). Perceived User Experience of Animated Transitions in Mobile User Interfaces. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 3152–3158. https://doi.org/10.1145/2851581.2892489

Molnar, L. J., Ryan, L. H., Pradhan, A. K., Eby, D. W., St. Louis, R. M., & Zakrajsek, J. S. (2018). Understanding trust and acceptance of automated vehicles: An exploratory simulator study of transfer of control between automated and manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 319–328. https://doi.org/10.1016/j.trf.2018.06.004

Musca, S. C., Kamiejski, R., Nugier, A., Méot, A., Abdelatif, E.-R., & Brauer, M. (2011). Data with hierarchical structure: impact of intraclass correlation and sample size on Type-I error. *Frontiers in Psychology*, *2*, 1–6.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Pampel, S. M., Jamson, S. L., Hibberd, D. L., & Barnard, Y. (2015). How I reduce fuel consumption: An experimental study on mental models of eco-driving. *Transportation Research Part C: Emerging Technologies*, *58*, 669–680. https://doi.org/10.1016/j.trc.2015.02.005

Peer, E., & Gamliel, E. (2012). Estimating time savings: The use of the proportion and percentage heuristics and the role of need for cognition. *Acta Psychologica*, *141*(3), 352–359. https://doi.org/10.1016/j.actpsy.2012.09.005

Peer, E., & Gamliel, E. (2013). Pace yourself: Improving time-saving judgments when increasing activity speed. *Judgment and Decision Making*, *8*(2), 10.

Peer, E., & Solomon, L. (2012). Professionally biased: misestimations of driving speed, journey time and time-savings among taxi and car drivers. *Judgement and Decision Making*, *7*(2), 165–172.

Robertson, G., Fernandez, R., Fisher, D., Lee, B., & Stasko, J. (2008). Effectiveness of Animation in Trend Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *14*(6), 1325–1332. https://doi.org/10.1109/TVCG.2008.125

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Simon, H. A. (1957). *Models of man*. Oxford, England: Wiley.

Simon, H. A. (1982). *Models of bounded Rationality*. Cambridge, MA: MIT Press.

Sivak, M., & Schoettle, B. (2012). Eco-driving: Strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy. *Transport Policy*, *22*, 96–99. https://doi.org/10.1016/j.tranpol.2012.05.010

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (second edition). London: Sage Publishers.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181. https://doi.org/10.1037/h0046162

Summala, H. (2007). Towards Understanding Motivational and Emotional Factors in Driver Behaviour: Comfort Through Satisficin. In C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments critical issues in driver interactions with intelligent transport systems* (pp. 189–207). Retrieved from http://springerlink.metapress.com/openurl.asp?genre=book&isbn=978-1-84628-617-9

Svenson, O. (1976). Experience of Mean Speed Related to Speeds over Parts of a Trip. *Ergonomics*, *19*(1), 11–20. https://doi.org/10.1080/00140137608931510

Svenson, O. (2008). Decisions among time saving options: When intuition is strong and wrong. *Acta Psychologica*, *127*(2), 501–509. https://doi.org/10.1016/j.actpsy.2007.09.003

Svenson, O. (2009). Driving Speed Changes and Subjective Estimates of Time Savings, Accident Risks and Braking. *Applied Cognitive Psychology*, *23*, 543–560. https://doi.org/10.1002/acp.1471

Svenson, O., & Salo, I. (2010). Effects of speed limit variation on judged mean speed of a trip. *Accident Analysis & Prevention*, *42*(2), 704–708. https://doi.org/10.1016/j.aap.2009.10.019

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality Measures as Predictors of Job Performance: A Meta-Analytic Review. *Personnel Psychology*, *44*(4), 703–742. https://doi.org/10.1111/j.1744-6570.1991.tb00696.x

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies*, *57*(4), 247–262. https://doi.org/10.1006/ijhc.2002.1017

Van Dooren, W., De Bock, D., Janssens, D., & Verschaffel, L. (2008). The Linear Imperative: An Inventory and Conceptual Analysis of Students' Overuse of Linearity. *Journal for Research in Mathematics Education*, *39*(3), 311–342. https://doi.org/10.2307/30034972

Wu, E., Jiang, L., Xu, L., & Nandi, A. (2016). Graphical Perception in Animated Bar Charts. *ArXiv:1604.00080 [Cs]*. Retrieved from http://arxiv.org/abs/1604.00080

8 APPENDIX

Table A1.
Translated items of inter-individual difference variables

Scale name	Item
Experience with consumption displays	When driving, I typically pay close attention to the display of the instantaneous consumption in order to save as much fuel as possible.
	The display of the instantaneous consumption hardly matters to me when I try to save fuel.
	The display of the instantaneous consumption helps me to assess which driving behaviour is fuel-efficient.
	When driving, I typically pay close attention to the display of the average consumption in order to save as much fuel as possible.
	The display of the average consumption hardly matters to me when I try to save fuel.
	The display of the average consumption helps me to assess which driving behaviour is fuel-efficient.
	Overall, I intentionally use displays while driving to save fuel.
	Displays do not really play a role for me when I try to drive energy efficiently.
Technical and mathematical knowledge	I have gained a good technical understanding of fuel-efficient acceleration through specific qualifications or personal activities.
	Concepts such as efficiency and conversion losses are familiar to me.
	I am familiar with integral calculus.
	I often solve mathematical-logical problems in my everyday life.
	Solving mathematical-logical problems does not pose a difficulty to me.

Table A2.

Cross-level interactions with inter-individual difference variables

No.	interactions	estimate	SE	df	t	p	
1	Peak-height*ATI	-0.09	0.03	52.83	-3.04	.004	**
2	Peak-duration*ATI	1.72	0.62	52.98	2.79	.007	**
3	Peak-duration*peak-height*ATI	-0.03	0.01	152.26	-2.33	.021	*
4	Peak-height*experience with consumption displays	-0.03	0.03	48.99	-0.83	.410	
5	Peak-duration*experience with consumption displays	0.67	0.64	48.99	1.04	.305	
6	Peak-duration*peak-height*experience with consumption displays	-0.02	0.01	170.46	-1.07	.288	
7	Peak-height* knowledge	-0.06	0.03	52.83	-1.85	.071	
8	Peak-duration* knowledge	1.51	0.65	52.97	2.33	.024	*
9	Peak-duration*peak-height*knowledge	-0.03	0.02	168.21	-1.76	.081	

Notes. ** p < .01; * p < .05.