External Human-Machine-Interfaces on Automated Vehicles: Which message and perspective do pedestrians in crossing situations understand best?

Sebastian Ludwig Weiß[1], Daniel Eisele[1], Tibor Petzoldt[1]

[1]Chair of Traffic and Transportation Psychology
Technische Universität Dresden, Germany

ABSTRACT

In the future, external-Human-Machine-Interfaces (eHMIs) may facilitate the communication between automated vehicles (AVs) and nearby pedestrians. The aim of this study was to investigate which messages (AVs' intention to yield or not to yield) and perspective (does the message refer to the behavior of the vehicle or the behavior that is expected of the pedestrian) of eHMIs are understood best in terms of objective comprehension, subjective comprehensibility, and speed. Participants in an online study

Intelligent Human Systems Integration (IHSI 2022): Integrating People and Intelligent Systems
https://openaccess.cms-conferences.org/#/publications/book/978-1-7923-8988-7

($N = 85$) indicated whether they can safely cross or not in reaction to six different eHMI icons. Messages that tell them to cross the street were understood better and faster compared to those that instructed them not to do so. Those referring to the pedestrian were best understood objectively as well as subjectively. We advise caution regarding eHMIs that communicate that the AV is not yielding / that the pedestrian cannot cross

Keywords: Automated Vehicles, eHMIs, Pedestrians, Crossing Situations

INTRODUCTION

Future automated vehicles (AVs) could be equipped with external human-machine-interfaces (eHMIs) that are supposed to facilitate the communication of AVs with surrounding road users. Prior research documents that such interfaces might support pedestrians' crossing decisions (Dey et al., 2020). To maximize their benefits, eHMIs need to be easily understandable to all road users, under all circumstances. In that regard, there is a discussion about *what* message should be communicated in such situations: Should the eHMI communicate both, the AV’s intention to yield as well as the intention not to yield to the pedestrian, or just one of these messages? Another question is *how* these messages should be communicated: Should the message refer to the pedestrian (i.e., egocentric: “You can(not) go”) or the AV (i.e., allocentric “I (do not) intend to yield”)? And, of course, it is vital that eHMIs are understood even under high cognitive load, as pedestrians might, e.g., be cognitively distracted (Dommes, 2019).

Some authors advise against the communication of non-yielding behaviour by AVs as this signal may be confusing rather than beneficial (Weber et al., 2019). However, communicating the intent not to yield can be critical information because it could have fatal consequences if a pedestrian crosses the road although an AV does not intend to yield. In this context, it is important to find a consensus about what to communicate. Both the messages and designs should be standardized. Otherwise, too many different signals and communication cues may confuse the recipients (Dey et al., 2020).

Apart from different messages, two perspectives have been distinguished (Bazilinskyy et al., 2019). On the one hand, an egocentric eHMI directly refers to the pedestrian. So, pedestrians are directly addressed and do not need to switch their perspective. In a crossing situation, the egocentric signal instructs the pedestrian in a form like “You can go”. Egocentric signals are often regarded as clear and unambiguous (Bazilinskyy et al., 2019). On the other hand, an eHMI signals from an allocentric perspective if the message refers to the AV. In this case, the signal communicates the AV’s status or intent (i.e., “I yield” or “I drive”) (Tabone et al., 2021). Pedestrians are therefore not instructed directly by allocentric eHMIs. In crossing situations, they have to first interpret the AV’s intent and subsequently decide whether to cross, based on their own judgement. Although the interpretation of allocentric messages is more demanding, there seems to be a consensus in literature that communicating the state, awareness and/or intent is the preferable approach compared to instructing the pedestrian via an egocentric eHMI. This is because egocentric eHMIs address one specific pedestrian. When several possible addressees are present, this might lead to confusion (Weber et al., 2019). Further, there might be legal implications in case of an accident that happened while a pedestrian followed the instruction to cross (Tabone et al., 2021).

The ability to switch to another road user's perspective and the understanding of eHMIs in general might be mitigated if the perceiver is cognitively loaded. Being a pedestrian is cognitively demanding as it requires visual and cognitive attention in a dynamic environment (Stavrinos et al., 2018). Possible effects of cognitive load on the comprehension of different messages and perspectives should therefore be investigated more precisely. Hence, the research question of this study was the following: Which message and perspective of eHMIs is understood best (in terms of accurate comprehension and speed) – even if the perceiver is cognitively loaded? To answer this question, an online study was conducted. The participants indicated whether they could safely cross, based on pictures of an AV that showed eHMIs communicating different messages (yielding/non-yielding) from different perspectives (egocentric/allocentric/am-biguous).

In line with the above reasoning, an earlier study by Eisma and colleagues explored the effects of message and perspective on the understandability of *text-based* eHMIs while taking cognitive load into account (Eisma et al., 2021). In this study, the applied egocentric messages were understood best, and their cognitive memory task had no significant effect. The authors argued that these findings might only be representative for text-based eHMIs and might therefore not generalize to the various other eHMI designs that have been proposed. In line with other authors, they called for further research as there is no consensus on the content, perspective, and modality of eHMIs yet (Dey et al., 2020). Regarding modality, visual eHMIs can communicate information with the highest density and in more detail compared to haptic, acoustic or body language eHMIs (Bengler et al., 2020). So, visual eHMIs should be pursued. Apart from visual text-based eHMIs, the application of icons seems promising: Research on icon-based traffic signs and digital icons indicates that they are effective after some training because their meaning gets more familiar over time (Goonetilleke et al., 2001). They are legible from far distance (Rettenmaier et al., 2020) and do not exclude persons who cannot read a certain language. Research on whether the findings by Eisma and colleagues generalize to eHMI icons therefore seems warranted. As their cognitive load task did not induce visuospatial load, which is prominent in traffic (Dommes, 2019), we conducted a conceptual replication employing a visuospatial cognitive load task and nonverbal eHMI icons.

METHODS

In the following the participants, apparatus and stimuli, procedure as well as the measures and analysis are described.

PARTICIPANTS

Participants were recruited via e-mail, social networks (Facebook, LinkedIn etc.), and via the participant pool of Technische Universität Dresden. In total, 85 participants completed the experiment. 50 were female (59 %) and 35 were male (41 %). Their age ranged from 18 to 78 years (M = 36.4, SD = 17.0). Knowledge of the German language was required to understand the instructions. Informed consent was obtained from all participants. Undergraduate psychology students received course credit for the participation.

APPARATUS AND STIMULI

An experiment was programmed and conducted online using LabVanced (Finger et al., 2017). A dual task paradigm was applied. The participants' primary task was to indicate whether it was safe to cross in reaction to still images of an AV that was equipped with one of six eHMI icons. Three different icons which represented different perspectives (cf. Figure 1, left) were designed: An egocentric perspective, which signalled from the pedestrian's point of view (i.e., "You, the pedestrian, can cross/not cross"), an allocentric perspective, which signalled from the point of view of the vehicle (i.e., "I, the vehicle intend to yield/do not intend to yield"), and an ambiguous perspective that could be interpreted from both perspectives. A crossed-out version of each icon was designed to ensure that one version of every icon messaged that the AV was yielding and another one that it was not yielding. This added up to six icons in total. In three pretests, the six eHMIs were iteratively adapted to achieve similar levels of comprehension accuracy and speed irrespective of their perspective and message. The final icons were attached to a vehicle, which drove on the driving lane of a parking lot (cf. Figure 1, right). The camera angle of the image was from the pedestrian's point of view and depicted an AV that seemed to drive towards the pedestrian. The eHMI icon was placed on the grill of the AV as this area is in the pedestrians' visual focus (Dey et al., 2019). A driver was not included because fully automated vehicles do no longer need/allow human intervention (SAE, 2021). The participants were instructed to imagine they want to cross in front of the AV and to decide as quickly as possible whether they can cross or not. At the start of the primary crossing the question "Can I go?" and a short reminder that clicking F = *yes* and J = *no* appeared on the screen. Then, the image depicting the AV with an eHMI was shown and participants indicated whether they can go or not. Consequently, the image disappeared, and the screen was blanked. Altogether, this sequence lasted 8000 ms independently of the time until a key was pressed.

The secondary task was to perform a visuospatial delayed match-to-sample working memory task. It consisted of a memory and a recognition task. For the memory task, a white dot was displayed in one of 20 possible positions on a 4 x 5 grid for 1000 ms. Participants had to remember either zero (low cognitive load), three (medium load), or six positions (high load) and maintain them for the following recognition task. The levels of cognitive load were chosen based on earlier research to ensure that the task was cognitively demanding but not too difficult (Höller-Wallscheid et al., 2017). The recognition task started with the question "Which position is novel?". White dots with numbers written on them were simultaneously presented in trials. These dots appeared on the 4 x 5 grid in all of the previously presented positions apart from one dot, which was shifted by one box. Participants had to identify this dot. They responded by pressing the respective number on their keyboard. There was no time limit for this recognition task. The time span of maintaining the spatial cues of the memory task was 15.2 s.

PROCEDURE

After the introduction, informed consent, instruction, and practice trial, participants performed the 18 main experimental trials (6 eHMIs x 3 levels of cognitive load). The order of the trials was completely randomized. Each trial had a fixed combination of the two factors which were the primary interpretation task and the secondary visuospatial memory task. Hence, the trials per se were the same for every participant. The participants' primary interpretation task was embedded in the secondary visuospatial memory task. Each trial under medium and high cognitive load started with the memory task: The respective white dots were shown. The primary interpretation task in reaction to the eHMI followed. Then, participants performed the recognition task of the secondary visuospatial memory task. At the end of each trial, participants were asked about their subjective cognitive load on a scale from 1 = *very low effort* to 10 = *very high effort*. This question marked the end of one trial and participants could take a short break. After all 18 main experimental trials, the participants rated the clarity of each of the eHMI icons subjectively without time pressure (1 = *very low* to 10 = *very high)*.

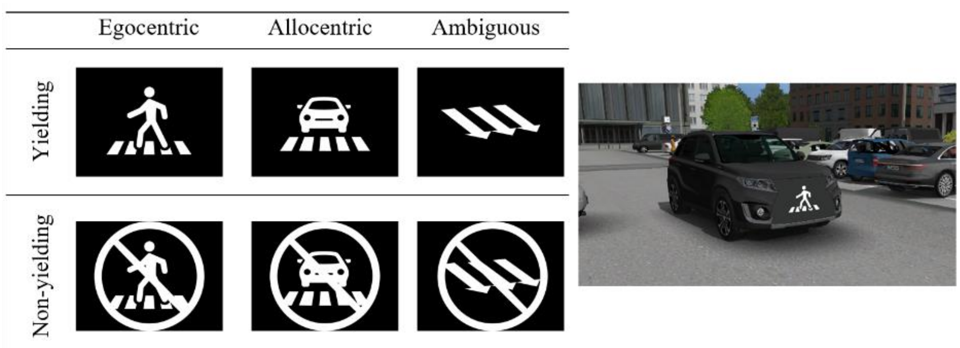

Figure 1. Left: The six eHMI icons arranged by perspective and message. Right: A sample screenshot of the base-image used in the experiment showing one of the icons.

MEASURES AND ANALYSIS

The participants' interpretation of the eHMIs, the response time, subjective clarity ratings, and subjective cognitive load were measured. The subjective cognitive load was measured to confirm that the manipulation of cognitive load worked as intended. As proposed by Eisma and colleagues, an objective clarity score was calculated to make the clarity of each eHMI icon comparable with each other. The objective clarity score depended on the participants' crossings: Objective clarity score (in %) = | 2 x (share of correct answers in % - 50 %) |. As there is no correct answer for ambiguous icons, the share of the predominant answer was used instead of the correct one (Eisma et al., 2021). A score of 0 % indicates that the message is very unclear (meaning that half of participants understood that they can cross, and the other half understood the opposite). As data did not meet the assumption of normal distribution, nonparametric tests were applied. Pairwise comparisons were corrected by Bonferroni corrections for multiple comparisons.

RESULTS

The objective clarity scores of the six eHMI icons are depicted in Table 1. As can be seen, objective clarity was highest for the egocentric, lower for the allocentric, and lowest for the ambiguous eHMIs. The objective clarity scores were higher for the yielding compared to the non-yielding messages for all three perspectives.

Table 1: Objective clarity scores for each eHMI.

	eHMI					
	Egocentric		Allocentric		Ambiguous	
	Yielding	Non-yielding	Yielding	Non-yielding	Yielding	Non-yielding
S	95 %	93 %	51 %	40 %	22 %	17 %

Note. The score could range from 0 % (= objectively very unclear, random decisions) to 100 % (= objectively very clear, exclusively correct decisions).

Regarding the subjective clarity, scores were descriptively highest for the egocentric eHMIs. See Table 2 for an overview of the subjective clarity scores. A nonparametric Friedman test indicated a significant difference in the subjective clarity ratings for the six different eHMIs, $\chi^2 (5) = 199.07$, $p < .001$. Pairwise comparisons revealed that the subjective clarity scores for the egocentric eHMIs were significantly higher than the score for the allocentric and ambiguous perspective, all $p < .001$. There was no significant difference between the ambiguous and allocentric eHMIs. Pairwise comparisons between the subjective clarity scores for different eHMI messages, but within one perspective, were only significant for the comparison between the ambiguous yielding and non-yielding eHMI, with a higher clarity score for the yielding eHMI, $p < .05$.

Table 2: Subjective clarity scores for each eHMI.

	eHMI					
	Egocentric		Allocentric		Ambiguous	
	Yielding	Non-yielding	Yielding	Non-yielding	Yielding	Non-yielding
M	9.26	8.55	5.15	3.95	5.34	3.87
SD	1.98	2.29	3.15	2.65	3.17	2.61

Note. Subjective clarity scores could range from 1 = *very low* to 10 = *very high.*

As can be seen in Table 3, response time (RT) was shorter for yielding than non-yielding messages. According to a nonparametric Friedman test, there were significant differences in RT between the six eHMIs, χ^2 (5) = 164.60, $p < .001$. This difference was significant for the egocentric ($p < .05$) and the ambiguous ($p < .001$) but not the allocentric eHMI perspective, $p = .116$. Pairwise comparisons revealed that response times for the egocentric yielding eHMI were shorter than the other perspectives ($p < .001$), the egocentric yielding eHMI being the overall shortest. Significant differences for the response times between allocentric and ambiguous eHMIs were only found when comparing yielding with non-yielding eHMIs of these perspectives, $p < .01$.

Table 3: Mean response times for each eHMI perspective and message.

	eHMI					
	Egocentric		Allocentric		Ambiguous	
	Yielding	Non-yielding	Yielding	Non-yielding	Yielding	Non-yielding
M	1703	2009	2593	3057	2345	3028
SD	755	969	1176	1405	1044	1445

Note. Mean response times in ms could range from 0 to 8000 ms.

Regarding cognitive load, there were no significant differences in objective clarity between the three levels. Pairwise comparisons of the response times between the levels of cognitive load were also nonsignificant (all $p > .05$) although an initial test showed that the response times significantly differed between the three levels, χ^2 (2) = 7.08, $p < .05$. Nevertheless, there was a trend that response times were longer under low (M = 2551 ms, SD = 982 ms) than under high (M = 2392 ms, SD = 897 ms; p = .052), and medium cognitive load (M = 2438, SD = 1046 ms; p = .078).

DISCUSSION

The aim of this study was to investigate which messages and perspectives of eHMI icons are understood best. Furthermore, it was tested whether there are differences in the participants' understanding of the eHMIs when they are cognitively loaded. Our results indicate that pedestrians understood those eHMI messages better and faster that tell them to cross the street than those that instructed them not to do so. In terms of perspective, eHMIs that directly referred to the behavior that is expected from the pedestrian were considerably better understood than those that referred to the AV's behavior or were ambiguous. There was no difference between the latter two regarding understandability. This pattern manifested in objective (interpretations, response times, objective clarity) as well as subjective measures (subjective clarity). These results are in line with previous research (Bazilinskyy et al., 2019; Eisma et al., 2021). Furthermore, the subjective and objective clarity ratings we identified for eHMI icons that refer to the behaviour that is expected from the pedestrian, were very similar to the ones Eisma and colleagues report for their text-based eHMIs (Eisma et al., 2021). This is probably due to the instructive character of these eHMIs (Bazilinskyy et al., 2019). Moreover, cognitive load did not influence the understanding. Interestingly, we replicated the previously observed trend that eHMIs were understood faster under high cognitive load. We initially figured that this finding might be due to the nature of the original task which did not induce *visuospatial* load (Eisma et al., 2021). Consequently, cognitive load should not be considered a central concern in eHMI research and design. However, we cannot preclude possible effects of other forms of distraction, like concurrent visual multitasking. Additionally, one should note that our findings are limited by the fact that we used still images. They lacked movement, which is an important factor in crossing decisions of course.

Overall, we conclude that eHMI icons might be understood as correct and quick as text-based eHMIs. We advise caution regarding eHMIs that communicate that the AV is not yielding / that the pedestrian cannot cross. This study corroborates previous evidence that eHMIs that refer to the behavior that is expected from the pedestrian (rather than the AV's own behavior) are understood best. Further, eHMIs are understood equally correct and quick even when the observer is cognitively loaded.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research as part of the project CADJapanGermany: HF [funding number 16ES1035].

REFERENCES

Bazilinskyy, P., Dodou, D., & de Winter, J. (2019). Survey on eHMI concepts: The effect of text, color, and perspective. *Transportation Research Part F: Traffic Psychology and Behaviour*, *67*, 175–194. https://doi.org/10.1016/j.trf.2019.10.013

Bengler, K., Rettenmaier, M., Fritz, N., & Feierle, A. (2020). From HMI to HMIs: Towards an HMI framework for automated driving. *Information*, *11*(2), 61. https://doi.org/10.3390/info11020061

Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Riener, A., Martens, M., & Terken, J. (2020). Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. *Transportation Research Interdisciplinary Perspectives*, *7*(100174), 150–174. https://doi.org/10.1016/j.trip.2020.100174

Dey, D., Walker, F., Martens, M., & Terken, J. (2019). Gaze patterns in pedestrian interaction with vehicles: Towards effective design of external human-machine interfaces for automated vehicles. *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 369–378. https://doi.org/10.1145/3342197.3344523

Dommes, A. (2019). Street-crossing workload in young and older pedestrians. *Accident Analysis & Prevention*, *128*, 175–184. https://doi.org/10.1016/j.aap.2019.04.018

Eisma, Y. B., Reiff, A., Kooijman, L., Dodou, D., & de Winter, J. C. F. (2021). External human-machine interfaces: Effects of message perspective. *Transportation Research Part F: Traffic Psychology and Behaviour*, *78*, 30–41. https://doi.org/10.1016/j.trf.2021.01.013

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). *LabVanced: A unified JavaScript framework for online studies*. International Conference on Computational Social Science, Cologne, Germany.

Goonetilleke, R. S., Martins Shih, H., Kai On, H., & Fritsch, J. (2001). Effects of training and representational characteristics in icon design. *International Journal of Human-Computer Studies*, *55*(5), 741–760. https://doi.org/10.1006/ijhc.2001.0501

Höller-Wallscheid, M. S., Thier, P., Pomper, J. K., & Lindner, A. (2017). Bilateral recruitment of prefrontal cortex in working memory is associated with task demand but not with age. *Proceedings of the National Academy of Sciences*, *114*(5), 830–839. https://doi.org/10.1073/pnas.1601983114

Rettenmaier, M., Schulze, J., & Bengler, K. (2020). How much space is required? Effect of distance, content, and color on external Human–Machine Interface size. *Information*, *11*(7), 346. https://doi.org/10.3390/info11070346

SAE. (2021). *J3016 Standard: Taxonomy and definitions for terms related to on-Road motor vehicle automated driving systems*. SAE. https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic

Stavrinos, D., Pope, C. N., Shen, J., & Schwebel, D. C. (2018). Distracted walking, bicycling, and driving: Systematic review and meta-analysis of mobile technology and youth crash risk. *Child Development*, *89*(1), 118–128. https://doi.org/10.1111/cdev.12827

Tabone, W., de Winter, J., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D., Sheridan, T. B., & Stanton, N. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation Research Interdisciplinary Perspectives*, *9*, 100293. https://doi.org/10.1016/j.trip.2020.100293

Weber, F., Chadowitz, R., Schmidt, K., Messerschmidt, J., & Fuest, T. (2019). *Crossing the street across the globe: A study on the effects of eHMI on pedestrians in the US, Germany and China*. 515–530.