Intelligent Human Systems Integration (IHSI 2024), Vol. 119, 2024, 70–78
https://doi.org/10.54941/ahfe1004470

Participants' Speed-Accuracy Trade-Off Behavior in High-Stress Situations in Simulator Studies

Marcel Usai[1], Philip Schäfer[2], Nicolas Herzberger[1], and Frank Flemisch[1]

[1]Fraunhofer FKIE, 53343 Wachtberg & RWTH Aachen University, 52056 Aachen, Germany
[2]RWTH Aachen University, 52056 Aachen, Germany

ABSTRACT

This paper describes the phenomenon of user speed-accuracy trade-off in decision making in time and safety critical domains. We have observed such a behavior in two simulator studies, using the example of a takeover request (TOR) in a driving simulator and a search and collect task in an Industry 4.0 virtual reality (VR) environment. In both studies, some participants were observed trading decision accuracy for reaction speed, ignoring obvious visual cues, resulting in a failure of the task, which could have been prevented. The paper describes the phenomena and raises but does not yet solve the question, how to react to this behavior, either by adapting the process of Human Systems Integration, or by adapting the design, e.g., with additional escalation patterns of a co-system.

Keywords: Human-machine cooperation, User studies, User experience, Speed-accuracy trade-off, Human behavior, Human systems integration

INTRODUCTION

User studies in virtual reality (VR) or driving simulators are often used to include end users or other stakeholders in the design process of human-machine systems. The simulation environment enables tests of system limit or system failure scenarios and allows to collect participant feedback in these situations. This possibility enables test cases in which, for example the technical co-system does not function properly anymore, or the human reaches her or his stress or workload limit.

A general observation in user studies is, that there are intraindividual differences in reaction time and quality. Depending on the subjective perceived urgency and danger, the actual trade-off made by participants might be completely different from what is required by the use case. Underlaying paradigms include regulatory models of behavior (e.g. Voß, 2020, based on Summala, 1988), situational awareness (Endsley, 1995), the task regulation model (Hockey, 1997) and speed-accuracy trade-offs (e.g. Wickens et al., 2013, Duckworth et al., 2018) which influence decisions and actions under uncertainty to solve uncertainty dilemmas (e.g., Flemisch and Baltzer et al. (in press), Flemisch et al. (submitted)).

Based on two exemplary studies, the authors want to highlight challenges in the design of human-AI cooperation, which may lead to a mismatch between required and actual trade-offs in reaction time and quality, especially under high time pressure and in dangerous situations. This raises the questions, a) how human-machine systems can be designed and humans integrated so that reaction time and quality are well balanced and b) if it is possible for an AI co-system to detect when a human is making incorrect speed-accuracy trade-offs, depending on the use case design, and which options exist to intervene.

This paper cannot provide an extensive answer to these challenges. It does, however, provide a description of the problem, enriched with examples and ideas on how to address these issues in the design phase of a human-machine system with the overall goal to foster an intelligent Human Systems Integration.

First Example: Driving Simulator Study on Takeover Behavior in Highly Automated Driving

The first exemplary study is a driving simulator study, described in full detail by Usai et al. (in press). N = 24 Participants followed a driving course including two different use cases (see Figure 1), in which the automation requested a takeover by the participants, resulting in 48 recorded reactions to a TOR.

Figure 1: Snapshots of both use cases in all three different HMI designs (Usai et al., in press).

While driving, whenever a dedicated lane for automated driving appeared, they entered highly automated driving mode and proceeded to play a Tetris game on a center stack display. The first use case consisted of an unregulated four way crossing that had to be crossed straight and which three other vehicles entered at the same time, of which the right one would always take the participants' right of way by starting to move when participant would start to move.[1] The second use case included the appearance of a breakdown vehicle standing in the center lane of a three-lane highway with dense traffic on the left lane, leaving only two choices to proceed without a fatal crash: Either changing to the right lane and overtaking the breakdown vehicle or stopping in front of it. Participants were divided into three groups to test three different interaction designs of the takeover request (TOR).

All TOR were triggered using an auditory warning, followed by a takeover design, which was different for each group. The first design would simply hand over control right after the TOR, regardless of human reaction. The second design would transfer control as soon as the human takes over or enter a minimum risk maneuver (MRM), if the human would not take over. Finally, the third design is based on the second one and adds a second warning layer, which is attention-sensitive and only triggered if the driver did not react to the first warning at all (based on gaze and control input data), as well as visual cues to it to guide the driver's attention. This includes a symbol, prompting the driver to take over, and, in the first use case (crossing), red marking on the ground in front, and for the second use case a red marking to the left (see Figure 1, bottom right), a green arrow on the ground in front pointing to the right lane and a green marking of the lower right edge of the windshield. All these cues are meant to guide the driver to take over the driving task and help decide on which action to take next. As soon as participants took over, all visual cues were turned off.

Figure 2 displays one participant of group three at the time of takeover in the second scenario. The front screen shows an extensive implementation of a head up display (HUD), featuring especially a red wall towards the left lane to signal the dense traffic and a green arrow to and a green corner on the right lane, signaling that it is safe to change to the right lane, because there is no traffic.

Figure 2: Driving simulator study setup. One of three groups of participants received extensive HUD feedback (design 3; red "wall" and green arrow as displayed in the figure).

[1]To give more details on the background of this second use case: It was simulated that the ego vehicles and two others were interconnected, and their automations cooperatively decided on a priority list, which allowed the participant to drive first. The one to the right of the participants' vehicle, however, did not have such capabilities, which made the automation issue a takeover request to the participants to fall back to a human-human cooperation between the two vehicles' drivers.

To determine the participants' reactions after the TOR, video recordings of the situations were evaluated. All participants reacted by reaching for the control inputs and changing their gaze behavior. Whenever participants performed an action other than a light braking[2] right away (less than one second after the first reaction), it is assumed that they traded for speed as a takeover strategy. Whenever they wait or look at different areas of interest other than the front view, i.e., look in mirrors or the instrument cluster, it is assumed that they traded for accuracy.

Outcomes of the situation are classified into successes and failures, based on whether they experienced a crash within the simulation or not. A failure can be a collision with the vehicle upfront, with other traffic or with the guard rail on the highway. Results, split by reaction, are presented in Table 1. For $N = 24$ participants and two use cases, there are in total 48 encounters; $n = 9$ participants experienced design 1, $n = 7$ design 2, and $n = 8$ design 3. Reactions are classified as "no reaction" (participant did not give input on lateral or longitudinal controls other than light braking), "lane change (LC) right" (participant did change to right lane and might have used the pedals), "lane change (LC) left" (participant did change to left lane and might have used the pedals; this always coincided with a collision with left lane traffic), and "brake only" (participant did brake, but gave no lateral input).

Table 1. Classification of successes and failures of the outcomes of all 48 situations.

successes	no reaction	LC right	LC left	brake only	total
Design 1	0	7	0	2	9
Design 2	3	2	0	2	7
Design 3	1	3	0	7	11
failures	no reaction	LC right	LC left	brake only	
Design 1	8	0	0	1	9
Design 2	1	0	4	2	7
Design 3	0	0	3	2	5

The results of interest regarding the participants' speed-accuracy trade-off are failures of participants in Design 3 when performing a lane change to the left (only in use case 2: breakdown vehicle). Whilst in general, Design 3 turns out to lead to fewer failures, still three participants collided with traffic while changing lanes, despite the design itself warning them to not change to the left lane by placing a big, red wall on the left side, and encouraging changing to right lane with other cues. Two of those three participants were observed to decide for a lane change right away, while the third one first observed the situation before acting on the decision. This leads to the alternative hypothesis that those participants did either not recognize the visual cues, purposefully ignored them or did not understand their meaning, all within the timeframe

[2] As in Design 1, control is transferred right after the TOR, vehicle velocity drops right away. To be able to compare Design 1 with the other two, similar behavior of the human i.e., a light braking right after the TOR, is ignored.

starting with the reaction to the TOR and ending with the decision execution of the lane change.

Second Example: "Impact of Visual and Auditory Warning Signals on Path Compliance in Virtual Reality Under Time Pressure"

In the second study, $N = 8$ participants had to complete a search and collect task in a virtual reality (VR) warehouse environment (see Figure 3) under time pressure. Visual and auditive warning signals were activated to alert users when they have strayed from a predefined path and should trigger their return to the intended walking route. A within-subject design method was used to investigate whether the cues achieve the desired effect to prompt users to return to the predefined path.

Figure 3: The VR warehouse scenario in which the test subjects had to navigate to and select the highlighted object on one of the four pillars. Leaving the grid-marked path triggered the warning effects.

To immerse the users into the VR warehouse setting the HTC Vive Pro Eye head-mounted-display connected wirelessly to an Intel i7 workstation with an Nvidia GeForce 2080ti graphics card. The simulation was always rendered with over 75 frames per second. The physical laboratory space exceeded the virtual warehouse in size and was free of obstacles, allowing users to navigate the simulation freely without physical restrictions.

The intent behind the visual (a) and auditive (b) cues is to signal the user that something is amiss, leveraging alertness known from gaming environments (visual effect) or parking situation (auditive signal) to encourage a change in behavior. The signals serve as a form of punishment and aim to both alert the user to a problem (in our case, straying from the path) and to encourage them to restore the previous state (returning to the predefined path). While the warning signals are activated, the user should feel a sense of discomfort.

a) The visual discomfort is inspired by the known damage effect in shooter games where the display turns red every time a bullet is hitting the player. To replicate this in a non-gaming context, a pulsing transparent red layer was overlaid on the user's view. Over time, the frequency of the pulsing increased, possibly intensifying the urgency for the user to correct their course.

b) The auditive discomfort is inspired by the known warning sound from a parking situation if an obstacle is close. The frequency is synced with the pulsing of the visual effect and was set to a higher volume than the common background noises of the environment.

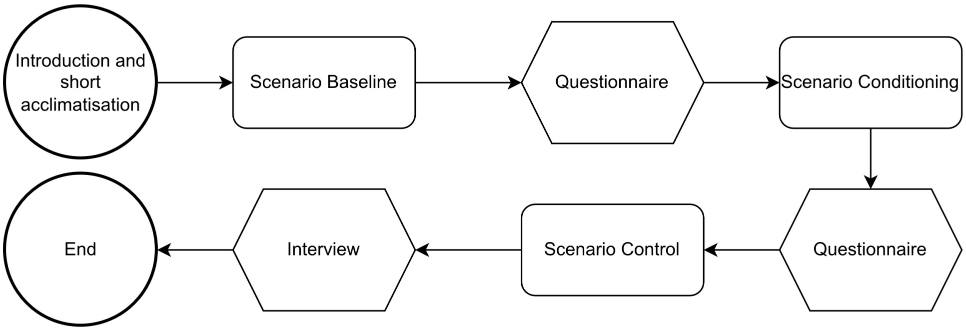

Figure 4: Flowchart of the study. After the introduction, the test subjects first went through a scenario without trigger effects, followed by a questionnaire, the conditioning and control scenario and a final interview.

After a short introduction on how to interact in virtual reality, the users spawned at the starting point in the warehouse as shown in Figure 4. They were given around 30 seconds to acclimatize to the new setting before they get teleported back to the starting point and the baseline scenario began. Their task was to navigate by natural walking to the highlighted object located on one of the four pillars and select them with the controller. There was only one object highlighted at the same time. The participants were told that their time needed to finish the scenario was measured. In the *Scenario Baseline*, no warning signals were activated if the participants went off the metal grids. In the following *Scenario Conditioning* they had to repeat the tasks as in the *Scenario Baseline* but while off the metal grids, visual and auditive signals were triggered as described before. In the *Scenario Control* the signals were deactivated again. At the beginning of each scenario, the participants were placed back to the starting point. The participants were not informed about the warning signals before the experiment and had to find out for themselves what triggers them and eventually, how to deactivate them again.

To answer whether the warning signals have an impact on the adherence to stick with the predefined path paired t-tests were employed comparing the time spent off the path between the following scenarios:

Baseline vs. Conditioning: The paired t-test comparing the Baseline and Conditioning scenarios yielded a T-statistic of 0.273 and a P-value of 0.793.

This indicates that there is no statistically significant difference in 'Time off path' between these two scenarios.

Baseline vs. Control: In the comparison between Baseline and Control scenarios, the analysis resulted in a T-statistic of 1.799 and a P-value of 0.115. This suggests that, while there was a trend towards a difference in 'Time off path' between Baseline and Control, this difference was not statistically significant.

Conditioning vs. Control: The test comparing Conditioning and Control scenarios showed a T-statistic of 2.397 and a P-value of 0.048. This result indicates a statistically significant difference in 'Time off path' between these scenarios.

Table 2. The time participants spent abroad the predefined path in seconds while performing the search and collect task in the different scenarios.

Time off path [s]			
n = 8	**Baseline**	**Conditioning**	**Control**
Mean	19.902	18.610	11.756
Std. Deviation	17.078	15.133	10.506
Minimum	1.040	4.670	0.280
Maximum	48.710	53.430	35.850

Interviews revealed varied perceptions of the warning signals among the participants. Of the eight, only three (P1, P2, and P6) recognized the signals as warnings and consciously chose to stay on the path thereafter. P7 acknowledged the peeping sound as an off-path indicator, yet opted to prioritize speed over comfort by disregarding it. Conversely, P4 identified the warnings as indicative of an error but was unable to discern their specific cause.

Design Choices for Human-Machine Systems

In both previous examples, which displayed warnings in simulation environments using transparent red layers, it happened that users were not stopped by them. It should be noted here that these are qualitative individual observations. Nevertheless, such observations accumulate across studies, use cases and situations, which was the motivating factor for this publication.

It is interesting to note that in these cases, participants often tend to prioritize reaction or execution time over accuracy (e.g. Schaefer et al., 2022), whether consciously, as in example 2, or unconsciously, as in example 1.

In the introduction, the questions were raised as to whether a) human-machine systems can be designed in such a way that reaction time and quality are well balanced and b) it is possible for an AI co-system to detect when a human is making incorrect speed-accuracy trade-offs, and to react appropriately.

A possible answer to question a) could be an interaction design which focusses on the actual integration of the human in the overall system. And as trivial as this may sound, countless socio-technical systems today are more likely to have been incrementally adapted to user experience (UX) and interaction in late stages of the development or even after the release of the system, rather than actually having focused on usability in the early stages of development. Especially in safety-relevant use cases, it is relevant to evaluate and minimize single cases of failure early enough in the development cycle.

The early integration of humans into the system can be implemented using tailored development methods (e.g., bHSI innovation turbine, Flemisch et al., 2022), allowing to identify unexpected reactions of future users and their causes at an early stage and to develop suitable warning methods together with future users (Human Systems Exploration, Flemisch et al., 2013, Preutenborbeck et al., 2024 (in press)) which can be adapted in real time (Exploration sandbox, e.g., Bielecki et al., 2020).

Question b) could be addressed by learning from user reactions and subsequent results observed in the past and using them to evaluate behavior and reaction patterns observed in real time. This approach has been used, for example, to predict the takeover quality of drivers by analyzing their orientation response to a TOR and comparing this orientation response with previously recorded ones (Herzberger, in press, Flemisch & Herzberger et al., in press). Similarly, a database could be used to compare gaze paths, for example, which would be constantly fed with further gaze sequences if such an ignoring of a cue is recorded. In this way, patterns could be identified that could initially be responded to with a further warning level and, in a second step, an interaction design adapted to these patterns could be developed.

CONCLUSION AND OUTLOOK

In two different user studies, a similar behavior of participants ignoring parts of a warning was observed. While most of the participants understood, that the co-system was pressuring for human action, hints on which action to take were still ignored by the same participants.

This observation was only made on a fraction of participants, however, depending on urgency and criticality of the warning, the design of human-machine system needs to work for all users and balance reaction speed and accuracy accordingly. One expedient solution would be, to integrate human users into the system design as early as possible. Another solution could be to discover these speed-accuracy trade-offs in real time, and allow a co-system to react on it, e.g., by giving hints or training after the fact, or by reacting in real time by using stronger escalation patterns. The phenomenon also opens the question of how much freedom we build into our human-machine systems, and how much unsafe speed-accuracy trade-offs we tolerate as a society and as individuals to balance safety and freedom in a self-determined way.

ACKNOWLEDGMENT

This publication was partly funded within the Priority Programme 1835 "Cooperative Interacting Automobiles (CoInCar)" of the German Science Foundation (DFG).

REFERENCES

Bielecki, K., López Hernández, D., Bloch, M., Baltzer, M., Schmidt, R. & Wasser, J. et al. (2020) Enhancing Reality: Adaptation Strategies for AR in the Field. In: Schmorrow, D. & Fidopiastis, C. M. (Eds.) *Augmented cognition: 11th International Conference, AC 2017, held as part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, proceedings, International Conference on Human-Computer Interaction*. Springer: Cham, pp. 304–315.

Duckworth, R. A., Potticary, A. L. & Badyaev, A. V. (2018) Chapter One - On the Origins of Adaptive Behavioral Complexity: Developmental Channeling of Structural Trade-offs. In: Naguib, M., Barrett, L., Healy, S. D., Podos, J., Simmons, L. W. & Zuk, M. (Eds.) *Advances in the Study of Behavior*. Academic Press, pp. 1–36.

Endsley, M. R. (1995) Measurement of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1), 65–84. Available from: https://doi.org/10.1518/001872095779049499.

Flemisch, F., Baltzer, M., Abbink, D., Siebert, L. C., van Diggelen, J. & Herzberger, N. D. et al. (in press) Towards a Dynamic Balance between Humans and AI-based Systems. In: *Multidisciplinary Reserach Handbook on Meaningful Human Control over Artificial Intelligence Systems*. Edward Elgar Publishing.

Flemisch, F., Herzberger, N. D., Usai, M., Baltzer, M. C. A., Schwalm, M. & Voß, G. et al. (in press) Cooperative Hub for cooperative research on Cooperatively Interacting Vehicles: Use-Cases, Design and Interaction patterns. In: Stiller, C. (Ed.) *Cooperative Interacting Vehicles*. Springer.

Flemisch, F., Preutenborbeck, M., Burlage, L. & Ripkens, A. (submitted) *Dilemma Model of Cooperative Guidance and Control to solve the speed-certainty-dilemma in Human-Autonomy Teaming: First Sketch*, 4th IEEE International Conference on Human-Machine Systems, Toronto.

Herzberger, N. D. (in press) *Erfassung der Übernahmefähigkeit von Fahrpersonen im Kontext des automatisierten Fahrens*. Shaker.

Hockey, G. R. (1997) Compensatory control in the regulation of human performance under stress and high workload; a cognitive-energetical framework. *Biological Psychology*, 45(1-3), 73–93. Available from: https://doi.org/10.1016/s0301-0511(96)05223-4.

Preutenborbeck, M., Wolf-Monheim, F., Carey, E., Canu, P., Herzberger, N. & Usai, M. et al. (2024 (in press)) Stakeholder-driven Design and Exploration of Sociotechnical Systems in Virtual Reality. In: *Proceedings of the 34th CIRP Design Conference*. Springer.

Schaefer, P., Vaskevic, A. & Meixner, G. (2022) Guided by the hint: How shader effects can influence object selection in Virtual Reality. In: *Human Factors in Virtual Environments and Game Design, 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022),* July 24-28, 2022. AHFE International.

Summala, H. (1988) Risk control is not risk adjustment: the zero-risk theory of driver behaviour and its implications. *Ergonomics*, 31(4), 491–506. Available from: ht tps://doi.org/10.1080/00140138808966694.

Usai, M., Herzberger, N. D., Yu, Y. & Flemisch, F. (in press) Confidence Horizons: Dynamic Balance of Human and Automation Control Ability in Cooperative Automated Driving. In: Stiller, C. (Ed.) *Cooperative Interacting Vehicles*. Springer.

Voß, G. M. I. (2020) *Regulatives Fahrerverhalten und dessen Relevanz für das automatisierte Fahren*.

Wickens, C. D., Hollands, J. G., Banbury, S. & Parasuraman, R. (2013) *Engineering Psychology and Human Performance*. Routledge: London [England], New York.