



OPEN Detecting unacceptable behavior of an autonomous vehicle using electroencephalography

Maren A. K. Bertheau^{1,2} & Christoph S. Herrmann^{1,3,4}✉

In the present study, we investigated event related potentials (ERPs) in the context of autonomous driving - specifically in left-turn situations through oncoming traffic. We recorded electroencephalography while participants ($n = 33$) observed a simulated autonomous vehicle executing a left turn maneuver through oncoming traffic. In the ERP, we observed an increased N2 (251 to 431 ms) when the AV behaved incongruently to the participant's assessment of the turn situation compared to when it behaved congruently. There were no significant effects in N1 (142 to 182 ms), P2 (177 to 237 ms), nor P3 (439 to 689 ms). This suggests that, in human-AV interaction, ERP-based devices might, in the future, be able to identify critical situations. However, further research is needed to bring current findings from fundamental research closer to application.

Keywords Autonomous driving, ERP, N1, N2, P2, P3

Highly autonomous vehicles (AVs) (SAE level 4) are becoming increasingly prevalent. Highly and fully AVs (SAE level 4 & 5) bear a great potential for increased road safety, energy efficiency, and social participation for people without driver's licenses, for instance, underage, elderly, or individuals with disabilities¹. However, even when federally approved for real-world traffic, an AV's non-human nature may lead to behavior that is formally lawful, but not necessarily socially or humanly acceptable. For instance, unacceptable behavior by the AV could arise in novel circumstances that the AV has not encountered before² or when its behavior would not fully align with the users' individual attitudes, personality, risk aversion, driving experience, etc.³. Additionally, AVs can take information from sensors beyond human sensory abilities into account (e.g., radar, infrared, etc.)^{4,5}. With these advantages, they would be able to perform certain traffic maneuvers more safely and swiftly than a human driver^{4,6}. Consequently, the behavior of an AV, even with federal approval, could at times be unexpected or unpredictable and thus likely unacceptable to the human user.

In line with the 'Autonomous Vehicle Acceptance Model'⁷ such unexpected or unacceptable behavior could induce stress and anxiety⁸ and lead to distrust in AVs, and thus to a refrain from using AVs⁷. A system could attempt to remedy this by adequately and individually adjusting its behavior by, e.g., giving an explanation⁹. However, for these adjustments to be successful, timing is crucial¹⁰. Therefore, critical or unacceptable situations need to be identified, especially in time-critical maneuvers.

To translate this complex need into simplified experimental terms, we define unacceptable situations as those in which an AVs exhibits behavior incongruent with the user's expectations. From an experimental design perspective, this is advantageous because it allows the creation of equal numbers of trials between experimental (incongruent) and control (congruent) conditions, independent of individual preferences.

In time-critical driving maneuvers, it is impractical to measure user perceptions (or ratings of acceptability) of AV behavior via questionnaires¹¹. Physiological measures may provide an unobtrusive way to assess user perceptions of AV behavior during dynamic and complex driving situations¹¹. In particular, electroencephalography (EEG) has cost-efficient mobile solutions, e.g.,¹², and a high temporal resolution that is well suited to investigate human - AV interaction in time-critical traffic situations. A frequently used measure in the EEG is the event related potential (ERP), a brain response time-locked to an event such as the presentation of a stimulus. These signals occur within (ms) and precede any voluntary motor reaction¹³.

¹Department of Psychology, Experimental Psychology Lab, Carl-von-Ossietzky University, Ammerländer Heerstr. 114-118, 26111 Oldenburg, Germany. ²Department of Informatics, Karlsruhe Institute of Technology (KIT), Am Fasanengarten 5, 76131 Karlsruhe, Germany. ³European Medical School, Cluster for Excellence "Hearing for All", Research Center Neurosensory Science, Carl-von-Ossietzky University, Carl-von-Ossietzky-Straße 9-11, 26129 Oldenburg, Germany. ⁴Research Center Neurosensory Science, Carl von Ossietzky University, 26129 Oldenburg, Germany. ✉email: christoph.herrmann@uni-oldenburg.de

When reviewing previous studies on the acceptability of decisions by AVs in traffic situations, to the best of our knowledge, EEG has never been thoroughly utilized before. Yet, there are several ERPs components with systematic amplitude modulations in tasks requiring similar cognitive processing as those required for assessing the incongruence of an AV's behavior. In foundational ERP research, there is a wide variety of frequently used tasks to manipulate congruence and incongruence. In the visual domain, commonly used tasks are the Eriksen Flanker task¹⁴, the Stroop task¹⁵, as well as Decision Making or Estimation tasks, e.g.,^{16–19}. Incongruent Eriksen Flanker (e.g., HSHH click S) as well as incongruent Stroop trials (e.g., “red” printed in blue) evoke an increased fronto-central N2²⁰ and modulate the parietal P3 response in the ERP^{20,21}. These ERPs might reflect the increased cognitive control²⁰, specifically inhibitory control or response inhibition^{20,22}. In loss trials from Decision Making or Estimation Tasks,^{16–19} (where the outcome was incongruent to a participant's expectation), the ERP shows a feedback related negativity (FRN), a fronto-central negative deflection peaking around 200–300 ms after negative feedback^{23–26}, which is then followed by a frontal P3a at around 300–400 ms and a parietal P3b at around 400–500 ms²⁶. The FRN is thought to play a central role in reinforcement learning processes^{26–28} and is functionally interpreted as tracking the difference between actual and expected outcomes, referred to as reward prediction errors²⁸.

Admittedly, all of these tasks are extremely simplified in comparison to traffic situations. Nevertheless, the underlying mechanisms of cognitive control, such as conflict resolution and retrieval, are highly relevant to successfully assess complex traffic maneuvers such as turning left through oncoming traffic. When choosing whether to turn left through a specific gap, conflict resolution is highly relevant. Every new gap needs to be assessed so that users can expect how the specific gap will be handled by the AV. Subsequently, users would need to retrieve their expectations to assess whether it was congruent or incongruent to the AV's behavior.

Folstein & van Petten²⁰ summarized that both N2 and P3 are broadly speaking associated with the significance of events (e.g., relevant stimuli). The N2 is generated in three distinct brain regions that are believed to serve different cognitive functions. The anterior N2 is associated with novelty and mismatch in attended stimuli. Further, the fronto-central N2 is associated with response inhibition (Go-No Go task) and the posterior N2 is associated with visual attention²⁰. However, while functionally distinct from the view of basic research, all these paradigms and mechanisms have in common that the N2 marks significant stimuli that are either new, rare, attended/awaited, or errors. The P3 is commonly separated into a more frontally generated P3a and a more parietally generated P3b. The P3a is usually associated with orienting attention to unexpected or significant events, whereas the P3b is associated with updating of working memory²⁰. Since we are interested in the unexpected behavior of AVs, and based on our previous research²⁹, we would expect an increased P3a in response to incongruent trials. Hence, we suspect that the ERP components N2 and P3 might play a role when an AV behaves incongruently to users' expectations in traffic. Further, earlier components could also be involved. The N1 is thought to play a crucial role in visual attention and serves as an index of discrimination processes³⁰. The prefrontal P2 is thought to play a role in stimulus-response mapping, in other words, classifying stimuli and matching them to the correct response³¹.

For the present study, we aimed to investigate the feasibility of using the ERP technique to capture participants' acceptance of AV behavior in dynamic traffic situations simulated in a laboratory environment. Our results might open new avenues to unobtrusively evaluate reciprocal communication between human users and AVs. Thus, this work is at the intersection of basic psychophysiological research and applied human-machine interaction development.

Hence, we developed a left turn task using driving simulations in a laboratory environment to effectively investigate ERPs during autonomous driving maneuvers. Driving simulations are advantageous as they provide a high control over the whole experimental situation as compared to e.g., Wizard of Oz approaches, where participants believe to interact with an actual AV in the field, but which is actually (sometimes remotely) operated by an unseen human confederate: see e.g.,³². Further, the electrically shielded laboratory environment provides a higher EEG data quality as compared to field setups³³.

We investigated left turn situations through oncoming traffic because they are (1) a frequently occurring traffic maneuver, (2) a high risk for accidents, (3) a known difficulty for AVs, and (4) allow for personal opinions or style. (1) The frequency of this maneuver is relevant, because for reliable ERP results it is necessary to plausibly repeat a certain task/situation numerous times³³. (2) To exemplarily showcase the proneness for accidents of left turn maneuvers, note that in 2023 there were 97.481 turning, turning-in, and crossing accidents with personal injury on German roads, which accounts for 34% of total accidents that year³⁴. (3) Furthermore, Wiegand, Eiband, Haubelt & Hussmann¹⁰ identified turning situations to cause existing AVs to behave unexpectedly using thematic analysis of real-world experience reports. (4) What makes turning left especially interesting to us, is that even if the maneuver is carried out without an accident, there is still room for personal preferences: Some vehicle users are more risk averse than others and therefore deem different gap sizes in the oncoming traffic as save enough, hence their assessment might frequently differ from the AV's.

In our left turn task, participants indicated whether they would take a left turn when there were gaps of variable length in oncoming traffic in a left turn situation. Next, the participants observe the AV's behavior as it either executes the turn or waits. This behavior can align (“congruent”) or conflict (“incongruent”) with the participant's decision.

We hypothesized that participants would rate congruent trials as more acceptable than incongruent ones (*Acceptability*).

Regarding the physiological measure, EEG, we hypothesized that the amplitudes of the N2, and P3 components of the ERP would be greater after the presentation of the AV's behavior in incongruent trials (see “*Methods*”) as compared to congruent trials. Additionally, since we use dynamic rather than static stimulus material, we wanted to ensure that we were replicating the temporal sequence of ERP components that are typically observed

in response to viewing static images. Hence, we also included earlier components in our analysis and expected increased N1 and P2 amplitudes in incongruent trials compared to congruent trials.

Results

Behavioral results

After artifact exclusion, 98.70% ($\bar{x} = 197.39$ (*standard deviation* (*SD*) = 2.62) per participant) trials went into analysis. In 69.95% ($\bar{x} = 137.91$ (*SD* = 46.90)) of trials, participants indicated that they wanted to turn and vice versa in 30.05% ($\bar{x} = 59.55$ (*SD* = 47.75)) that they wanted to wait (see supplementary Table 1).

Overall, participants rated the *Acceptability* of the AV's behavior higher when it was congruent with their own decision ($t(32) = 12.02$, $p < .001$, $d = 2.09$) (see Fig. 1).

Regarding the question after debriefing, if they knew they were not interacting with a real artificial intelligence (AI), 23.5% indicated “yes”, 44.1% “not sure”, and 32.4% “no”.

EEG results

The task was designed such that out of a total of 200 trials, the AV behaved congruently with the participant's decision in 50% and incongruently in 50% of the trials. Additionally, in 50% of the trials, the AV turned through the proposed gap in traffic while in the other 50%, it waited for another gap. This resulted in different numbers of trials in each combination in the final analysis: After artifact correction, in 34.86% ($\bar{x} = 68.73$ (*SD* = 23.13)) the AV congruently turned, 14.92% ($\bar{x} = 29.58$ (*SD* = 24.30)) the AV incongruently turned, 15.13% ($\bar{x} = 29.97$ (*SD* = 23.56)) the AV congruently waited, and 35.09% ($\bar{x} = 69.18$ (*SD* = 23.88)) the AV incongruently waited (see supplementary Table 1). After artifact rejection, there were on average $\bar{x} = 98.70$ (*SD* = 1.31) congruent trials, and $\bar{x} = 98.76$ (*SD* = 1.75) incongruent trials (see supplementary Table 1) per participant for the final analysis.

The following section presents the results of the two analysis of variance (ANOVAs) (*Condition* $\times$ *Electrodes*) conducted to analyze the N1, P2, N2, and P3 components of the ERP. To control the false discovery rate only $p < 0.0167$ is considered significant³⁵.

N1 results

For the N1 (142 to 182 ms), we found no significant main effect of *Condition* on the amplitude (μV) ($F_{(1,32)} = 9.90$, $p = .27$) (see supplementary Table 2). The standard errors (SEs) of the incongruent and congruent N1 amplitude overlap (see Fig. 2).

There was a significant main effect on *Electrodes* ($F_{(2,96)} = 48.24$, $p(\text{Greenhouse} - \text{Geisser}(GG)) < .001$, $\eta_p^2 = .60$) (see supplementary Table 2 & 3). This effect describes the topography over the midline electrodes. As typical for a N1 the amplitude was the lowest at parietal and occipital sites. (see Fig. 3)

Furthermore, there was no two-way interaction *Condition* $\times$ *Electrodes* ($F_{(2,96)} = .92$, $p_{(GG)} < .43$) in the N1 time window (see supplementary Table 2).

P2 Results

For the P2 (177 to 237 ms), we found no significant main effect of *Condition* on the amplitude (μV) ($F_{(1,32)} = .47$, $p = .50$) (see supplementary Table 4). The SEs of the incongruent and congruent P2 amplitude overlap (see Fig. 3).

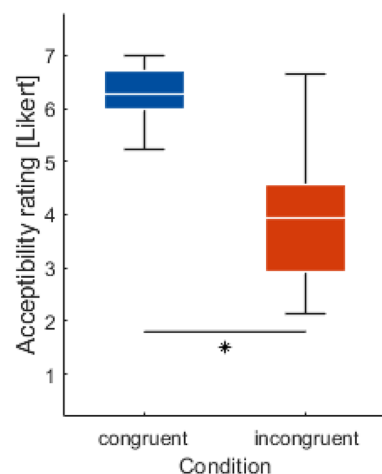


Fig. 1. Acceptability ratings after each trial (see “Methods”). The Likert scale ranged from ‘completely acceptable’ (7) to ‘completely unacceptable’ (1). The *Condition* congruent is always shown in blue and incongruent in red. On each box, the central marks indicate the median, and the bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. Statistical significance is indicated with *.

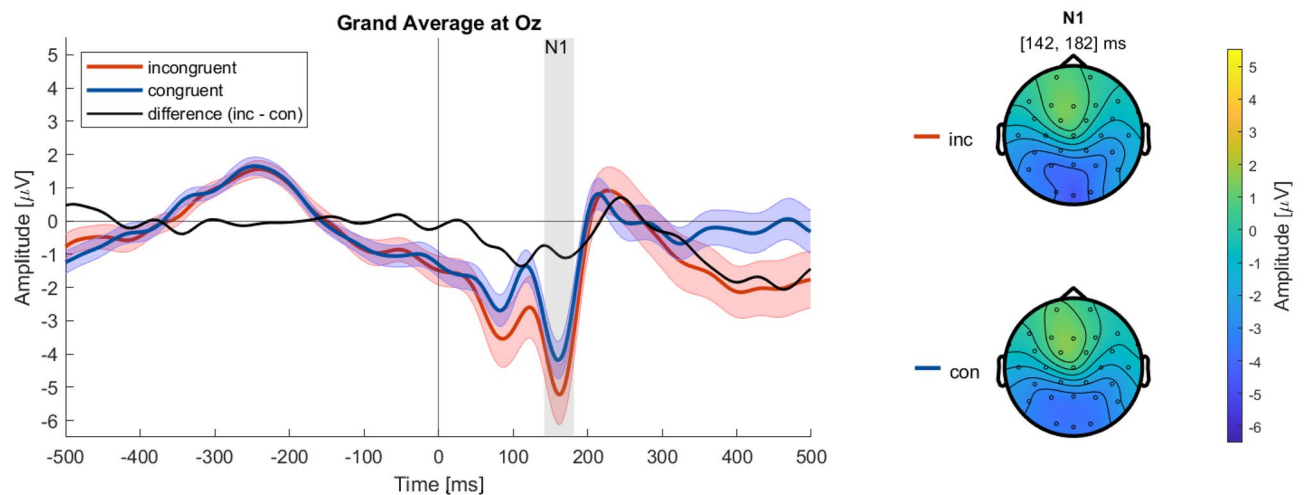


Fig. 2. The grand average ERPs at electrode Oz time-locked to the presentation of the AV's decision (see “Methods”) for both conditions, incongruent (red), congruent (blue), as well as their difference (black), were filtered from 0.1 to 20 Hertz (Hz) and with a baseline from – 500 to 0 ms. The shaded error bars indicate the SE. The gray shading indicates the identified time interval of the N1 component corresponding to the topographies. The topographies show the N1 in the incongruent, and congruent condition, sharing one color bar.

There was a significant main effect on *Electrodes* ($F_{(2,64)} = 26.74, p_{(GG)} < .001, \eta_p^2 = .50$) (see supplementary Table 4 & 5), which again illustrates the topography of this component consisting of an increased amplitude over fronto-central electrodes (see Fig. 3).

Furthermore, there was no two-way interaction *Condition x Electrodes* ($F_{(2,64)} = .50, p_{(GG)} < .52$) in the P2 time window (see supplementary Table 4).

N2 results

For the N2 (251 to 431 ms) we found a significant main effect of *Condition* on the amplitude (μV) ($F_{(1,32)} = 15.14, p < .001, \eta_p^2 = .32$) (see supplementary Table 6). The amplitude of the N2 was greater in incongruent vs. congruent trials (see Fig. 3).

There was also a significant main effect on *Electrodes* ($F_{(2,64)} = 34.68, p_{(Greenhouse - Geisser(GG))} < .001, \eta_p^2 = .52$) (see supplementary Table 6 & 7). This effect describes the topography in both conditions, showing the highest amplitude at Fz, and then in descending order at Cz and at Pz.

Furthermore, there was no two-way interaction *Condition x Electrodes* ($F_{(2,64)} = .54, p_{(GG)} < .50$) in the N2 time window (see supplementary Table 6).

Exploratorily, we conducted a correlation analysis to examine whether a difference in subjective acceptability between congruent and incongruent conditions is significantly associated with a difference in N2 amplitude between those two conditions. Frontally, we find a significant Pearson correlation ($r = .47, p = .006$), however not centrally ($r = .25, p = .16$), nor parietally ($r = -.04, p = .83$).

P3 results

Similarly to the N1 and P2, for the P3 (439 to 689 ms) we found no significant main effect of *Condition* on the amplitude (μV) ($F_{(1,32)} = .77, p = .39$) (see supplementary Table 8). During the P3, the SEs of the incongruent and congruent amplitude overlap (see Fig. 3).

There was a significant main effect on *Electrodes* ($F_{(2,64)} = 35.32, p_{(Greenhouse - Geisser(GG))} < .001, \eta_p^2 = .53$) (see supplementary Table 8 & 9). In the topography, this is visible as the amplitude of the P3 descriptively being higher at Fz and Cz than at Pz (see Fig. 3).

Furthermore, there was no two-way interaction *Condition x Electrodes* ($F_{(2,64)} = 3.39, p_{(GG)} < .06$) in the P3 time window (see supplementary Table 8).

Discussion

In the present study, we investigated modulations of the N1, P2, N2, and P3 components of the ERPs, when an AV maneuvers left turn situations through oncoming traffic.

Regarding the behavioral results, as we expected, participants rated trials where the AV behaved congruent to their assessment as more acceptable than trials with incongruent behavior (see Fig. 1). Note, that the distribution of acceptability rating in the incongruent condition was quite wide. Possibly since AV never produced an accident, some participants rated also incongruent trials as fairly acceptable. This pattern was also present when dividing trials with respect to whether the AV turned or waited: participants rated turning congruently as more

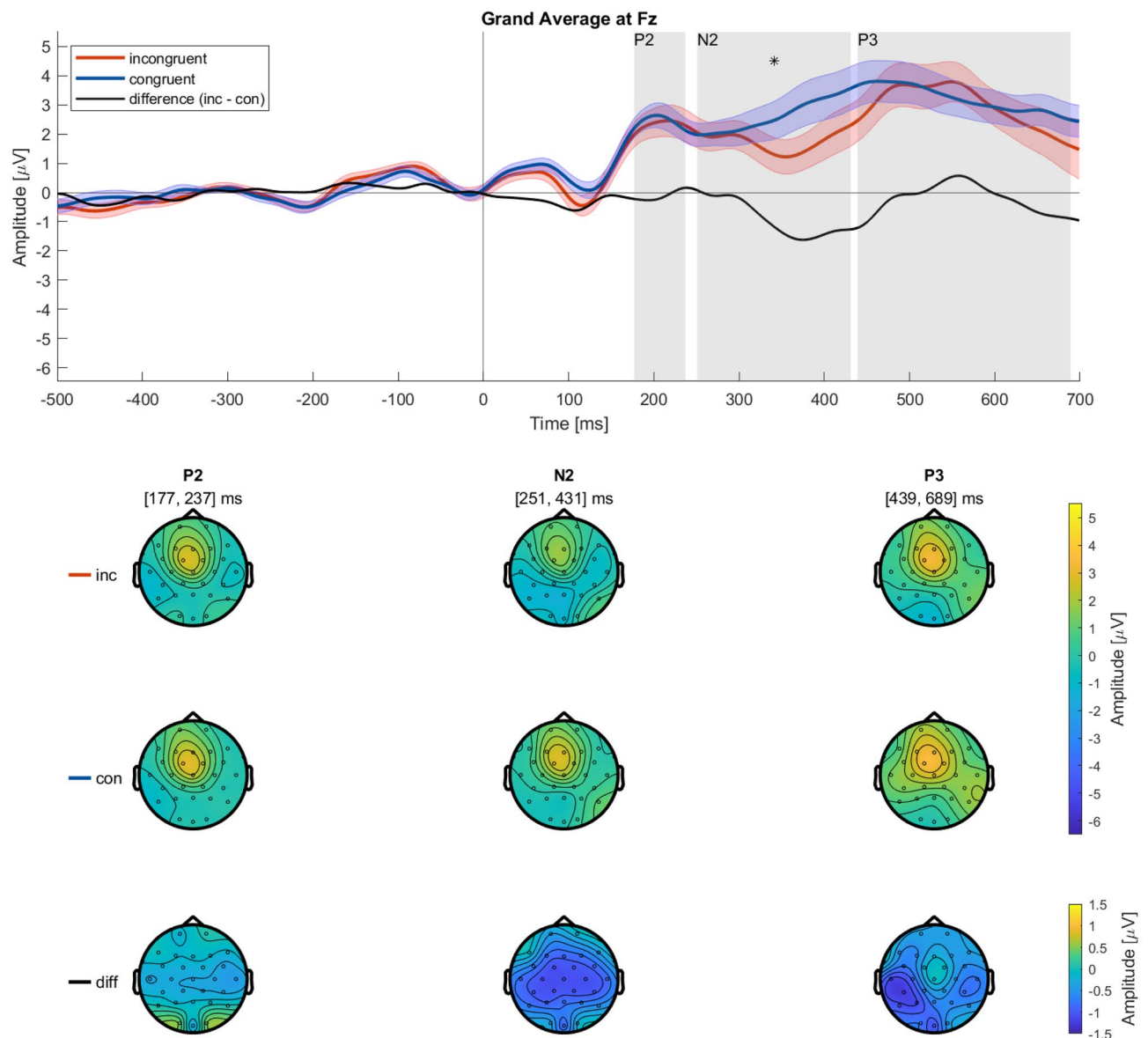


Fig. 3. The grand average ERP at electrode Fz time-locked to the presentation of the AV's decision (see "Methods") for both conditions, incongruent (red), congruent (blue), as well as their difference (black), were filtered from 0.1 to 20 Hz and with a baseline from -500 to 0 ms. The shaded error bars indicate the SE. The gray shading indicates the time intervals of the components P2, N2, and P3 corresponding to the topographies below. * indicate significant time intervals. The topographies show the ERP components P2, N2, and P3 column-wise in the given time intervals. The first row shows topographies from the incongruent, and the second row from the congruent condition. The third shows their difference. Note, that the two conditions share one color bar, whereas we used a smaller color bar for the difference between conditions.

acceptable than turning incongruently, and they rated waiting congruently as more acceptable than waiting incongruently (see supplementary Fig. 1).

In the debriefing, less than one in four participants indicated that they at some point knew that they were not interacting with an AI but a random system. However, these participants could still agree or disagree with its behavior. Therefore, we kept these participants for the analysis to avoid suffering from a lower experimental power.

Overall, participants demonstrated expected behavior, which leads us to infer that participants engaged thoroughly with the task, thus serving as a foundation for the following EEG interpretation.

In the ERP time-locked to the AV's decision (see Fig. 4), there was a significant main effect of *Condition* in the N2 component, but not in N1, P2, and P3. In the ERP components N2 descriptively, there was an increased negativity in the incongruent compared to the congruent *Condition* (see Fig. 3).

In contrast to the majority of ERP studies, we used dynamic instead of static stimulus material, specifically videos generated in a driving simulation environment showing left turn maneuvers by an AV through oncoming

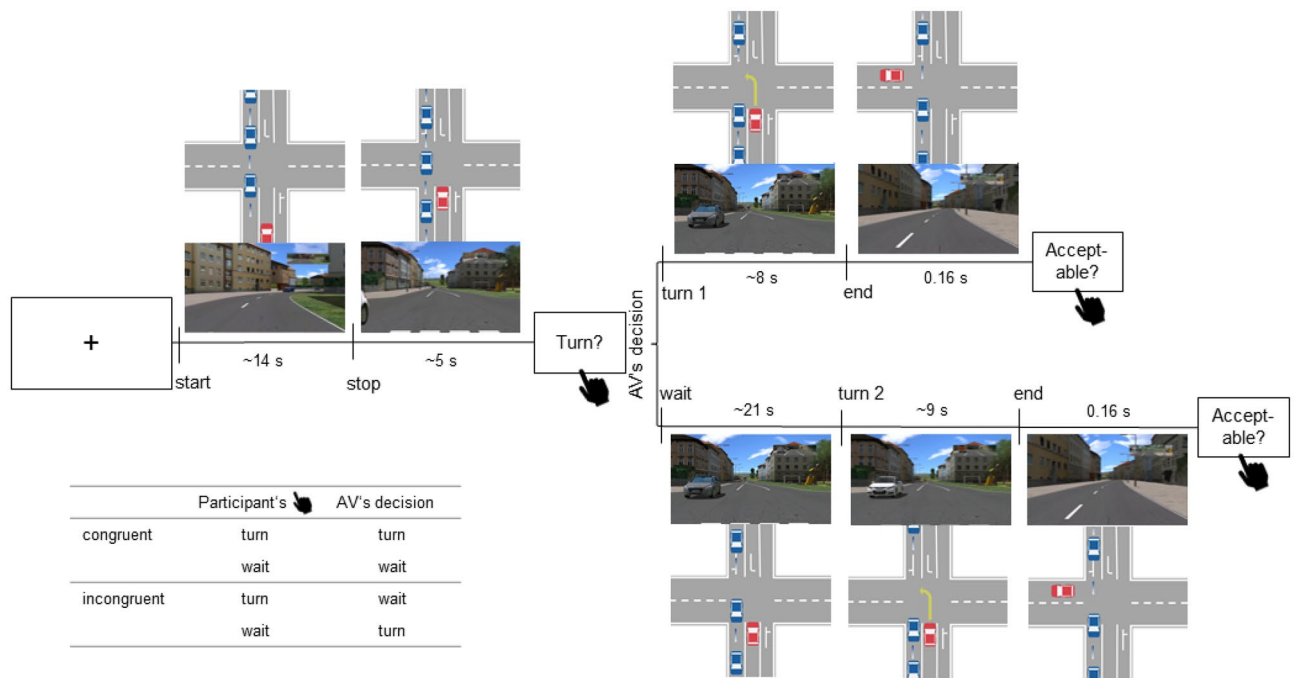


Fig. 4. Schematic overview of an exemplary trial in the LT task. The flow chart illustrates the trial's time course, with all relevant events marked by vertically labeled ticks. For each relevant event, a screenshot from a corresponding video and a schematic drawing of the traffic situation at that moment are provided. The videos show the driver's perspective of the red vehicle in the schematic drawings. Each trial begins with a cross at the center of the screen, then the AV approaches the crossing (start) and stops (stop). At the first possible gap between oncoming vehicles (blue vehicles in the schematic drawings), participants were asked to indicate via button press whether they wanted to turn left through the gap ('Turn?'). As illustrated in the table, in congruent trials the AV's decision complies, in incongruent trials it diverges from the participants' decision. The AV either turns (turn 1) or waits (wait) for the next possible gap, where it eventually turns (turn 2). The videos ended after a successful turn (end). At the end of each trial, participants were asked to rate the acceptability of the AV's behavior (Acceptable?). See [supplementary Methods](#) for the exact wording of the used items. On each section of the timeline, the mean duration of each phase is indicated in s.

traffic. Therefore, we analyzed early components to confirm if we can obtain stable ERPs with our material. Even though there was no difference between conditions in the N1, the topographies (see Fig. 2) reveal the typical pattern of an occipital negativity³⁶. We interpret this as a strong indicator of the credibility of our ERP data. The N1 is thought to play a crucial role in visual attention and as an index of discrimination processes³⁰. Further, it is associated with the processing of spatial visual information³⁷, which is relevant to successfully assessing traffic situations. However, as the N1 is an early ERP component, in this case, it likely reflects differences in the physicality of the stimulus material (size, position, contrast, etc.)^{38,39}, and therefore, here needs to be interpreted with caution.

Regarding the subsequent components, P2, N2, and P3, their topographies (see Fig. 3) all showed increased amplitudes in fronto-central sites, whereas the amplitudes in parietal and occipital sites appeared lower. Visibly, these topographies resemble the typical topographies of a P3a⁴⁰ and this dominant topography already builds up during the earlier components. Hence, when further interpreting the N2 we concentrated on the difference topography contrasting incongruent - congruent trials (see Fig. 3) because in the ERP plot (see Fig. 3) the N2 negativity is only prominently visible in the incongruent condition (red line) but not in the congruent (blue line). Therefore, the difference topography incongruent - congruent of the N2 likely reflects the actual N2 topography, isolated from the ramping up towards the P3a. The observed N2 shows a central distribution spreading widely over frontal, parietal, as well as lateralized sites (see Fig. 3). Additionally, we showed a positive correlation between N2 difference and Acceptability rating difference. Participants who rate incongruent trials as less acceptable display a more negative N2 amplitude. The N2 ERP component was shown to have different generators and functions with distinct topographies depending on specific task modalities²⁰. Roughly, the N2 is frequently subdivided into (1) the anterior or fronto-central N2 in response to deviating non-targets, and response inhibition, (2) the posterior N2, in response to deviant targets, and (3) the posterior contra-lateral N2 (N2pc), a marker of spatial visual attention^{20,33}. The wide distribution in the present topography, however, suggests that not one isolated function was necessary to solve our task, but rather a combination of anterior and posterior N2 was at play. As the N2 is especially sensitive to incongruent stimuli and task modalities^{20,41–47}, and significantly correlated with the Acceptability ratings we would argue that the N2 might be a helpful indicator when identifying critical situations, even in complex stimulus material such as traffic using ERPs.

Contrary to our hypothesis, there were no differences between the *Conditions* regarding the amplitudes of both ERP components P2 and P3. In both conditions, the P2 and P3 peaks are located fronto-centrally (see Fig. 3). The prefrontal P2 is thought to play a role in stimulus-response mapping, in other words, classifying stimuli and matching them to the correct response³¹. The lack of differences in P2 amplitudes in our conditions suggests that this functionality was equally important during congruent as well as during incongruent trials.

For the P3a, we found a fronto-central topography that did not differ between conditions. The P3a is usually associated with orienting attention to significant events, whereas the P3b is associated with updating of working memory²⁰. This might suggest that an AV executing a left turn through oncoming traffic by itself represents a significant event.

To address some limitations, in the present study, we focused on plausible traffic situations for an AV with approval for road use. Hence, our simulation did not entail truly failed traffic maneuvers (accidents), since for those it is unambiguously clear that they are unacceptable. Further, accidents are difficult to plausibly simulate, and they should already be minimized by manufacturers, especially since there is evidence showing that for future AV users adoption is tied to substantially increased safety (4–5 times safer) in comparison to human drivers⁴⁸. The simulated AV in this study always turns successfully and never has an accident. Over time, participants more frequently indicated that they wanted to turn even through the smaller gaps between vehicles. This suggests that over system usage time, participants showed increased trust towards the AV, which could be interpreted as a positive sign for future acceptance of AV. Additionally, the fact that the AV always performed a successful turn maneuver resulted in a wide distribution of acceptability ratings in the incongruent condition. Hence, there might be different levels of acceptability in such human-AV interaction situations. Thus, it might be worthwhile to differentiate those in a future experimental design. Further, this study was designed in such a fashion that the AV behaved congruently to the participant's assessment in 50% of trials and incongruently in 50%. Consequently, in the congruent condition, there was an increased number of trials where the AV turned, whereas, in the incongruent condition, there was an increased number of trials where the AV waited for another gap. Therefore, the observed increase of the N2 response is particularly noteworthy: It occurred when the AV did not show the expected behavior (turning), which would have resulted in immense changes in the visual field, but waited - so participants' visual perspective remained the same in most incongruent trials. However, the impact of the changes in the visual field on our effect could be a subject for further study. Taken together, the significant event was the AV not showing an action but remaining inactive. We find this a promising indication that in the future successful ERP-based brain computer interface (BCI) applications in the context of AVs.

To give some future outlook, in the present study, we set out to investigate the usefulness of ERPs in the context of identifying critical situations with AVs. For this purpose, we decided to start in a rather controlled laboratory environment. We could show that it is possible to find high-quality ERPs even in complex moving stimulus material. In the next steps, the reliability of our findings would need to be tested in setups closer to field application: e.g., in a Wizard of Oz environment in an actual car and using smaller, mobile EEG devices. Further, to develop an actual ERP based device to identify unacceptable traffic situations, it would be necessary to employ single-trial classifiers (e.g.,^{25,49–53}). In all those studies, the classifiers have been fed with training data from situations providing clear wrong and right outcomes. Since traffic is often more ambiguous, it might be advisable to include data from more frequently occurring stimulus material, such as ours, in these lines of research. However, this was beyond the scope of the current study and is a wide field for further research.

In conclusion, we investigated the usefulness of ERPs in the context of AVs. In our controlled laboratory environment driving simulation study, we could show that it is possible to find high quality ERPs even in moving stimulus material. When observing an AV executing a traffic maneuver in a fashion that contradicts participants' assessment, we find increased amplitudes in the ERP components N1 and N2, but no significant difference in P2 and P3. This suggests that in human-AV interactions, it might be possible to use ERP based devices to identify critical situations. However, further research is needed to bring current findings from fundamental research closer to application.

Methods

Note, that this is the same set of participants, procedure, and apparatus than in our previous Moral Machine (MM) study²⁹ (under revision), because both studies were recorded simultaneously.

Participants

Thirty-six volunteers were recruited to participate in this study to achieve sufficient statistical power in our design. Participants were included in the study if they were over 18 years old, right-handed, native German speakers, had normal or corrected to normal vision, and had no psychiatric or neurological diseases. Additionally, participants were required to have a valid German driver's license and be legally fit to drive under German regulations (§31 StVZO, and §316 StGB) at the time of the experiment. All participants gave written informed consent and received 10 € per hour for participation. The research protocols were approved by the Commission for Research Impact Assessment and Ethics ("Kommission für Forschungsfolgenabschätzung und Ethik") at the Carl von Ossietzky University of Oldenburg and complied with all relevant ethical regulations. Two participants were excluded because they participated in only one of the two recording sessions. Another participant was excluded because they clicked 'turn' in all trials. Hence, there were no 'congruent wait' and 'incongruent turn' trials for that participant. Therefore, the data of 33 participants (mean age = 24.455, SD = 3.042, age = 3.042, age range = 19–34 years, male = 18) were included in the analysis.

Procedure

The experiment consisted of two sessions, conducted on different days (days between sessions: mean=6.618, SD=8.707). Each session took ~ 150 – 200 min.

	Session I				Session II			
Task	LT training	LT	MM training	MM	MM training	MM	LT training	LT
Conditions	con	con & inc	con	con & inc	con	con & inc	con	con & inc
Trials	3	100 (in 4 blocks)	6	100 (in 2 blocks)	6	100 (in 2 blocks)	3	100 (in 4 blocks)
Time [min]	~3	~80 (in 4 blocks)	~1	~20 (in 2 blocks)	~1	~20 (in 2 blocks)	~3	~80 (in 4 blocks)

Table 1. Overview of procedure.

During the first experimental session, participants gave written informed consent and filled in the demographic questionnaires. In both sessions, the EEG and the peripheral physiological sensors were positioned. Participants were then seated in an electrically shielded, dimly lit chamber at ~ 120 cm distance in front of a display. First, participants were familiarized with the two tasks, Left Turn (LT), and MM, in a short training. To achieve somewhat similar training times, for the LT task there were three congruent trials, and for the MM task there were six congruent trials.

The main experiment consisted of both tasks. The order of the tasks was counterbalanced both between sessions and participants. The current paper discusses the data from the LT task. For results from the MM task see²⁹ (under revision). The LT task of the main experiment was divided into four (~ 20 min) blocks. Between blocks, participants had the opportunity to take short, self-paced breaks. After the second session, they were debriefed that they did not interact with an AI but a deterministic abstract machine. Then they were asked if they knew they were not interacting with a real AI, with the answering options “yes”, “not sure”, and “no”. Table 1 summarizes the overall procedure.

Left turn task

For both tasks, there was a cover story that the AV would make elaborate, data-driven decisions. In the debriefing at the end of the last experimental session, it was clarified that all of the AV’s supposed decisions were predetermined by a random number generator.

Each trial began with a cross that was presented at the center of the screen for 1000 ms. Then, participants saw short video sequences from the driver’s perspective of an AV approaching an intersection with oncoming traffic. At the stop line, the simulation was interrupted and participants were asked to assess whether they would turn left in the next gap between oncoming vehicles or whether they would prefer to wait. Participants were instructed to ‘answer from their gut’ when they were ready and to indicate their opinion using the left and right arrow buttons on a standard keyboard. After pressing an arrow, participants viewed a video in which the AV behaved either congruently or incongruently with their decision. After each trial, participants rated the acceptability of the AV’s decision on a Likert scale ranging from 1, completely unacceptable, to 7, completely acceptable. This range was chosen to include a neutral middle point (4) while ensuring that differences remain distinguishable in the human working memory⁵⁴. See Fig. 4 for an overview of the task.

We created 30 different vignettes. Each participant performed 200 trials in total in random order. Half of the trials were presented in the congruent setting, and half in the incongruent setting. Thus, each scenario was presented at least 3x in a congruent and 3x in an incongruent setting (see Table 1).

Apparatus

Hardware

The stimulus material was presented on a Samsung SyncMaster display: P247GH, 1920 x 1080 pixels, 60 Hz refresh rate.

The EEG data were recorded using a 32-channel actiCAP snap electrode system with the standard 10-10 layout (BrainProducts GmbH, Gilching, Germany). The reference electrode was placed on the tip of the nose and a vertical electro-oculogram (EOG) electrode was placed under the right eye. The ground electrode was placed at electrode location FPz. Thus, the final EEG signal was acquired from 30 active Ag-AgCl electrodes. Impedance was measured before recording and kept below 20 kΩ. EEG signals were digitized at a rate of 1000 Hz by a 16-bit ActiChamp EEG amplifier.

Software

The task was programmed using the Psychtoolbox^{55–57} in MATLAB 2021a (The MathWorks, Inc., Natick, MA, USA).

The video sequences (see Fig. 4 for screenshots) were created in the commercial driving simulation software SILAB Version 6 (WIVW GmbH, Würzburg, Germany).

Physiological signals were recorded using BrainVision Pycorder software (BrainProducts GmbH, Gilching, Germany).

Data preprocessing and plotting were performed using the FieldTrip toolbox (⁵⁸; <http://fieldtriptoolbox.org>) in MATLAB 2022b (The MathWorks, Inc., Natick, MA, USA).

All statistical analyses were computed using SPSS 29 (IBM, Armonk, NY, USA).

Data analysis

EEG preprocessing

The EEG data were then demeaned, detrended, and filtered between 1 and 35 Hz using a Butterworth filter (ft_preprocessing)⁵⁸. The data were then segmented into 1000 ms epochs without overlap. Epochs containing

data that exceeded a threshold of 1000 μV were rejected. Then, an independent component analysis (ICA) was performed, and components containing eye and muscle movement artifacts were visually identified and rejected. The remaining components were projected back onto the raw EEG data.

The raw data were then filtered from 0.1 to 35 Hz, and cut around the trigger at “AV’s decision” from –2500 to 2500 ms (see Fig. 4). This trigger marked the first frame in the video where the motion of the vehicle differed in either the car remaining still or starting to move. Using accumulated z-score thresholding ($\text{ft_artifact_zvalue}$)⁵⁸, trials with a $z \geq 20$ were rejected. The trials from the two experimental sessions were merged, and the data were downsampled to 250 Hz. Since our stimulus material was dynamic throughout, which might have caused the baseline to be unstable, we accounted for this by applying a wide baseline correction window from –500 to 0 ms. For the ERPs, we averaged the epochs from each AV’s decision (see Fig. 4) for each participant. Then we calculated a grand average over participants for each condition separately (congruent, and incongruent).

To identify the ‘N1’, ‘P2’, ‘N2’, and ‘P3’ components of the ERP, we applied a peak picking algorithm (`findpeaks()`) on the grand average collapsed over all conditions. For each component, we defined a time window and target electrodes for identifying the peak or trough amplitudes (μV) respectively: N1 (100–200 ms at Oz)³³, P2 (150–250 ms at Fz)⁵⁹, N2 (200–350 ms at Fz)²⁰, and P3 (250–500 ms at Cz)⁴⁰. For each component, we defined a relevant time window around the identified peaks or troughs: N1 (162 ± 20 ms), P2 (207 ± 30 ms), N2 (341 ± 90 ms), and P3 ($489 - 50 + 200$ ms). Note, that the windows’ widths were chosen to ensure that they covered the whole component. Subsequently, the P3 time window was shifted to the right to ensure that there was no overlap with the N2 window and to better capture the skewed shape of the P3 (see gray shadings in Fig. 3). These four component windows were then applied to the previously calculated grand averages for the conditions (congruent and incongruent). All samples in one time window were collapsed for the three relevant electrodes (Fz, Cz, Pz) and exported for inference statistical analyses.

Statistical analysis

To examine the differences in acceptability ratings between conditions, we conducted a repeated measures t-test.

To examine the differences in the ERP components between conditions, we conducted four two-way repeated measures ANOVAs with the factors *Condition* (congruent, incongruent) $\times$ *Electrodes* (Fz, Cz, Pz), one for each relevant component N1, P2, N2, and P3. Note that for N1, we also included Oz in the analysis. To account for the sphericity problem GG corrections were used to correct the df. Within each of these four ANOVAs, we controlled the false discovery rate³⁵. Accordingly, we applied adjusted α levels to the three ANOVA effects to determine significance: .0500 for the largest p value, .0333 for the middle p value, and .0167 for the smallest p value³⁵.

Each ANOVA archives a strong statistical power of $1 - \beta \geq .94$ with 33 complete datasets, and a conservatively assumed correlation between repeated measures of $r = .2$, and a conventionally medium-sized effect ($f = .25$)^{60–62} (see [supplementary Methods](#) for details).

Furthermore, as a post-hoc procedure to further disentangle significant three-level main effects and interaction effects, we employed Bonferroni corrected t-tests for C relevant comparisons on the simple main effects.

Exploratorily, we performed a Pearson correlation analysis between N2 amplitude differences ($N2_{\text{incongruent}} - N2_{\text{congruent}}$) and difference acceptability ratings ($\text{incongruent} - \text{congruent}$).

Data availability

The data are available from the corresponding author upon reasonable request.

Received: 29 January 2025; Accepted: 1 September 2025

Published online: 12 September 2025

References

- Li, D., Huang, Y. & Qian, L. Potential adoption of robotaxi service: the roles of perceived benefits to multiple stakeholders and environmental awareness. *Transp. Policy* **126**, 120–135 (2022).
- Rocklage, E. Teaching self-driving cars to dream: a deeply integrated, innovative approach for solving the autonomous vehicle validation problem in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC) 1–7 (IEEE, 2017).
- Šucha, M. & Aernochová, D. Driver personality as a valid predictor of risky driving. *Transport. Res. Procedia* **14**, 4286–4295 (2016).
- Schoettle, B. Sensor fusion: a comparison of sensing capabilities of human drivers and highly automated vehicles <https://public.ebooks.umich.edu/E&oe/umtriswt/pdf/swt-2017-12.pdf> (2017).
- Li, D., Liu, A., Pan, H. & Chen, W. Safe, efficient and socially-compatible decision of automated vehicles: a case study of unsignalized intersection driving. *Automotive Innov.* **6**, 281–296 (2023).
- Rydzewski, A. & Czarnul, P. Human awareness versus Autonomous Vehicles view: comparison of reaction times during emergencies. In *2021 IEEE Intelligent Vehicles Symposium (IV)* 732–739 (IEEE, 2021).
- Hewitt, C., Politis, I., Amanatidis, T. & Sarkar, A. Assessing public perception of self-driving cars. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (eds. Fu, W.-T. et al.) 518–527 (ACM, 2019).
- Schick, B., Seidler, C., Aydogdu, S. & Kuo, Y.-J. Driving experience versus mental stress at assisted lateral guidance. *ATZ Worldwide* **121**, 68–73 (2019).
- Ha, T., Kim, S., Seo, D. & Lee, S. Effects of explanation types and perceived risk on trust in autonomous vehicles. *Transport. Res. F: Traffic Psychol. Behav.* **73**, 271–280 (2020).
- Wiegand, G., Eiband, M., Haubelt, M. & Hussmann, H. “I’d like an explanation for That!” Exploring reactions to unexpected autonomous driving. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* 1–11 (ACM, 2020).
- Lohani, M., Payne, B. R. & Strayer, D. L. A review of psychophysiological measures to assess cognitive states in real-world driving. *Front. Hum. Neurosci.* **13**, 57 (2019).
- Bleichner, M. G. & Debener, S. Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG. *Front. Hum. Neurosci.* **11**, 163 (2017).

13. MacKay, D. M. Do we “control” our brains?. *Behav. Brain Sci.* **8**, 546 (1985).
14. Eriksen, B. A. & Eriksen, C. W. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* **16**, 143–149 (1974).
15. Stroop, J. R. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **18**, 643–662 (1935).
16. Miltner, W. H., Braun, C. H. & Coles, M. G. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a “generic” neural system for error detection. *J. Cogn. Neurosci.* **9**, 788–798 (1997).
17. Nieuwenhuis, S., Yeung, N., Holroyd, C. B., Schurger, A. & Cohen, J. D. Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cerebral Cortex* **14**, 741–747 (2004).
18. Holroyd, C. B., Nieuwenhuis, S., Yeung, N. & Cohen, J. D. Errors in reward prediction are reflected in the event-related brain potential. *NeuroReport* **14**, 2481–2484 (2003).
19. Hu, X., Xu, Z., Li, Y. & Mai, X. The impact of trust decision-making on outcome processing: evidence from brain potentials and neural oscillations. *Neuropsychologia* **119**, 136–144 (2018).
20. Folstein, J. R. & van Petten, C. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* **45**, 152–170 (2008).
21. Zurrón, M., Pouso, M., Lindín, M., Galdo, S. & Díaz, F. Event-related potentials with the Stroop colour-word task: timing of semantic conflict. *Int. J. Psychophysiol.* **72**, 246–252 (2009).
22. Leon-Carrion, J., García-Orza, J. & Pérez-Santamaría, F. J. Development of the inhibitory component of the executive functions in children and adolescents. *Int. J. Neurosci.* **114**, 1291–1311 (2004).
23. Cohen, M. X., Elger, C. E. & Ranganath, C. Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage* **35**, 968–978 (2007).
24. Frank, M. J., Woroch, B. S. & Curran, T. Error-related negativity predicts reinforcement learning and conflict biases. *Neuron* **47**, 495–501 (2005).
25. Chavarriaga, R., Sobolewski, A. & Del Millán, J. R. Errare machinale est: the use of error-related potentials in brain-machine interfaces. *Front. Neurosci.* **8**, 208 (2014).
26. Ullsperger, M., Fischer, A. G., Nigbur, R. & Endrass, T. Neural mechanisms and temporal dynamics of performance monitoring. *Trends Cogn. Sci.* **18**, 259–267 (2014).
27. Faßzligbender, L., Krause, D. & Weigelt, M. Feedback processing in cognitive and motor tasks: a meta-analysis on the feedback-related negativity. *Psychophysiology* **60**(12), e14439 (2023).
28. Walsh, M. M. & Anderson, J. R. Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neurosci. Biobehav. Rev.* **36**, 1870–1884 (2012).
29. Bertheau, M. A. K., Boetzel, C. & Herrmann, C. S. Event-related potentials reveal incongruent behavior of autonomous vehicles in the moral machine dilemma. *Sci. Rep.* **15**, 16048 (2025).
30. Vogel, E. K. & Luck, S. J. The visual N1 component as an index of a discrimination process. *Psychophysiology* **37**, 190–203 (2000).
31. Aydin, M., Lucia, S., Casella, A., Bello, B. & Russo, F. Bayesian interpretation of the prefrontal P2 ERP component based on stimulus/response mapping uncertainty. *Int. J. Psychophysiol.* **199**, 112337 (2024).
32. Detjen, H., Pfleging, B. & Schneegass, S. A Wizard of Oz field study to understand non-driving-related activities, trust, and acceptance of automated vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* 19–29 (ACM, 2020).
33. Luck, S. J. An introduction to the event-related potential technique 2nd ed. <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=798344> (MIT Press, 2014).
34. Statistisches Bundesamt. Unfälle mit Personenschaden, Verunglückte: Deutschland, Jahre, Unfalltyp/Unfallart, Ortslage, Schwere der Verletzung, Tabelle 46241-0005 (ed Statistisches Bundesamt) 2024. <https://www-genesis.destatis.de/genesis/online?sequenz=statistikTabellen&selectionname=46241#abreadcrumb>.
35. Hidden multiplicity in exploratory multiway ANOVA. Prevalence and remedies. *Psychon. Bull. Rev.* **23**, 640–647 (2016).
36. Herrmann, C. S. & Knight, R. T. Mechanisms of human attention: event-related potentials and oscillations. *Neurosci. Biobehav. Rev.* **25**, 465–476. <http://www.sciencedirect.com/science/article/pii/S0149763401000276> (2001).
37. Wascher, E., Hoffmann, S., Sängler, J. & Grosjean, M. Visuo-spatial processing and the N1 component of the ERP. *Psychophysiology* **46**, 1270–1277 (2009).
38. Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K. & Herrmann, C. S. Size matters: Effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clin. Neurophysiol.* **115**, 1810–1820 (2004).
39. Fründ, I., Busch, N. A., Körner, U., Schadow, J. & Herrmann, C. S. EEG oscillations in the gamma and alpha range respond differently to spatial frequency. *Vision. Res.* **47**, 2086–2098 (2007).
40. Polich, J. Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* **118**, 2128–2148 (2007).
41. Li, Y., Wang, Y., Zhang, B., Wang, Y. & Zhou, X. Electrophysiological responses to expectancy violations in semantic and gambling tasks: a comparison of different EEG reference approaches. *Front. Neurosci.* **12**, 169 (2018).
42. van de Weijer, J., Paradis, C., Willners, C. & Lindgren, M. Antonym canonicity: temporal and contextual manipulations. *Brain Lang.* **128**, 1–8 (2014).
43. White, K. R., Crites, S. L., Taylor, J. H. & Corral, G. Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Social Cogn. Affect. Neurosci.* **4**, 191–198 (2009).
44. Wang, P., Yang, Y.-P., Tan, C.-H., Chen, Q.-W. & Cantfort, T. Gender stereotype activation versus lexical semantic activation: an ERP study. *J. Gen. Psychol.* **144**, 283–308 (2017).
45. Debruille, J. B. The N400 potential could index a semantic inhibition. *Brain Res. Rev.* **56**, 472–477 (2007).
46. Delogu, F., Brouwer, H. & Crocker, M. W. Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain Cogn.* **135**, 103569 (2019).
47. Soto, M. & França, A. I. The functional analysis of the N400 component: lexical access, integration or can we have it both ways?. *Revista Lingüística* **16**, 521–562 (2020).
48. Liu, P., Yang, R. & Xu, Z. How safe is safe enough for self-driving vehicles?. *Risk Anal.* **39**, 315–325 (2019).
49. Wang, X., Chen, H.-T., Wang, Y.-K. & Lin, C.-T. Implicit robot control using error-related potential-based brain-computer interface. *IEEE Trans. Cogn. Dev. Syst.* **15**, 198–209 (2023).
50. Nakamura, K. & Natsume, K. Detection of error-related potentials during the robot navigation task by humans. In *2020 International Conference on Computational Intelligence (ICCI)* 153–158 (IEEE, 2020).
51. Ehrlich, S. K. & Cheng, G. A feasibility study for validating robot actions using EEG-based error-related potentials. *Int. J. Soc. Robot.* **11**, 271–283 (2019).
52. Lopes-Dias, C., Sburlea, A. I. & Müller-Putz, G. R. Online asynchronous decoding of error-related potentials during the continuous control of a robot. *Sci. Rep.* **9**, 17596 (2019).
53. Vi, C. T., Jamil, I., Coyle, D. & Subramanian, S. Error related negativity in observing interactive tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (eds. Jones, M. et al.) 3787–3796 (ACM, 2014).
54. Colman, A. M., Norris, C. E. & Preston, C. C. Comparing rating scales of different lengths: equivalence of scores from 5-Point and 7-Point scales. *Psychol. Rep.* **80**, 355–362 (1997).
55. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
56. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).

57. Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R. & Broussard, C. What's new in Psychtoolbox-3. *Perception* **2007**, 1–16 (2007).
58. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
59. Crowley, K. E. & Colrain, I. M. A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin. Neurophysiol.* **115**, 732–744 (2004).
60. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
61. Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. Tests for correlation and regression analyses. Statistical power analyses using G*Power 3.1. *Behav. Res. Methods* **41**, 1149–1160 (2009).
62. Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. Quantitative Methoden. Band 2 5th ed. Supplementary material retrieved from https://www.lehrbuch-psychologie.springernature.com/sites/default/files/2022-01/Rasch_A5_978-3-662-63283-3_Kapitel_3_GPower_R_Ergaenzung.pdf (Springer, 2021).

Acknowledgements

This research was supported by the Research Training Group “Social Embeddedness of Autonomous Cyber Physical Systems” (SEAS) by the University of Oldenburg, as well as by the Innovation Campus for Future Mobility (www.icm-bw.de) and by the Helmholtz Association within the Core Informatics project in Karlsruhe. Here, I would like to especially thank Prof. Dr. Maike Schwammberger (KIT, Karlsruhe, Germany) for her support. I would like to thank Dr. Martin Fränze (OFFIS, Oldenburg, Germany) for providing access to the driving simulator and Dr. Lars Weber (DLR-SE, Oldenburg, Germany) for programming the driving simulation in SILAB, specifically creating the traffic scenarios and the driving automation systems. Additionally, I want to thank current and past members of our team, Dr. Sreekari Vogeti, Dr. Cindy Boetzel, Dr. Heiko I. Stecher, and Dr. Florian Kasten for their support, as well as our students Maria Stollmann and Pablo Alexis Olguín Baxman for their help with piloting and data recordings.

Author contributions

M.A.K.B., and C.S.H. conceptualized the study. M.A.K.B. conducted the experiment, analysed the results, and wrote the manuscript. C.S.H. contributed to the data analysis and conceptualization of the manuscript. All authors reviewed and approved the submitted manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-18305-2>.

Correspondence and requests for materials should be addressed to C.S.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025