

# Effect of Partially Automated Driving on Mental Workload, Visual Behavior and Engagement in Nondriving-Related Tasks: A Meta-Analysis

Human Factors  
2025, Vol. 0(0) 1–30  
© 2025 Human Factors  
and Ergonomics Society



Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/00187208251323132  
[journals.sagepub.com/home/hfs](https://journals.sagepub.com/home/hfs)



Nicola Vasta<sup>1</sup> and Francesco Biondi<sup>1,2</sup> 

## Abstract

**Objective:** The goal of this meta-analysis is to investigate the effect of partial automation on mental workload, visual behavior, and engagement in nondriving-related tasks.

**Background:** The literature on the human factors of operating partially automated driving offers mixed findings. While some studies show partial driving automation to result in suboptimal mental workload, others found it to impose similar levels of workload to the ones observed during manual driving. Likewise, while some studies evidence a marked increase in off-road glances when the automated system was engaged, other work has failed to replicate this pattern.

**Method:** 41 studies involving 1482 participants were analyzed using the PRISMA approach.

**Results:** No significant differences in mental workload were found between manual and partially automated driving, indicating no changes in mental workload between the two driving modes. A higher likelihood of glancing away from the forward roadway and engaging in nondriving-related tasks was found when the partially automated system was engaged.

**Conclusion:** Although the adoption of partial driving automation comes with some intended safety benefits, its use is also associated with an increased engagement in nondriving-related activities.

**Application:** These findings add to our understanding of the safety of partial automation and provide valuable information to Human Factors practitioners and regulators about the use and potential safety risks of using these systems in the real-world.

## Keywords

driving automation, cognitive workload, visual behavior, distraction, detection response tasks, meta-analysis, safety, automated driving, workload, partial automation

## Introduction

Approximately 1.19 million people die each year as a result of road crashes worldwide, with activities like speeding and using cellphones being among the most dangerous contributors to road

---

<sup>1</sup>University of Windsor, Canada

<sup>2</sup>University of Utah, USA

### Corresponding Author:

Francesco Biondi, Human Systems Lab, University of Windsor,  
2555 College Ave, Windsor, Ontario N9B 3P4, Canada;  
e-mail: [francesco.biondi@uwindsor.ca](mailto:francesco.biondi@uwindsor.ca)

fatalities (WHO, 2023). In response to this, a growing number of automobile manufacturers have introduced driving automation that can control the vehicle's behavior to limit the detrimental effect that these activities have on safety. The Society of Automotive Engineers (SAE, 2021) identifies six levels of driving automation, ranging from 0 (manual driving) to 5 (fully automated driving). Currently, SAE level 2 (L2) systems, also known as partially automated driving, are becoming more commonplace on our roads, with a projected market share of 60% by 2025 in the United States (Statista, 2023). These systems can assist the driver by controlling both the vehicle's steering and acceleration in selected conditions, provided the driver remains vigilant and ready to take control whenever necessary (SAE, 2021). Yet, preliminary crash data show that using these systems may lead to drivers disengaging from the driving task more often compared to manual driving, thus posing a risk to safety (e.g., NHTSA, 2022; NTSB, 2020b).

Operating partially automated systems is expected to switch the role of the human driver from *vehicle operator* to *system supervisor* (Biondi et al., 2019; Cabrall et al., 2019). This drastic shift in driver responsibilities is hypothesized to reduce the overall demand of driving, thus impacting the driver's mental workload (cf. Solís-Marcos et al., 2017). Mental workload refers to the amount of mental resources required to process information, make decisions, and perform actions during a task (cf. Young et al., 2015). Suboptimal levels of mental workload (whether too high or too low) can negatively impact driving performance and vigilance (Biondi, 2024; McWilliams & Ward, 2021). For instance, a demanding task, such as manually operating a vehicle while talking on a cellphone, may increase mental workload, thereby reducing the resources available for driving. Conversely, an arguably easier task, such as supervising the functioning of an automated system, may lower mental workload, thus freeing up more mental resources (cf. Biondi, 2024, Engström, Tech, Kingdom, & Victor, 2013). However, while reducing mental workload could potentially allow drivers to allocate more mental resources to identifying hazards, recent research suggests that such reduction in workload may instead be accompanied by a general disengagement from

driving-related activities (e.g., Biondi et al., 2023; Solís-Marcos et al., 2018) and an increased engagement in nondriving-related tasks (NDRT), such as looking away from the road or using a cellphone (Hungund & Kumar Pradhan, 2023; Zhang et al., 2021).

Research on the effect of automated driving on mental workload has produced mixed findings. Using self-reported metrics, Stapel et al. (2019) found a reduction in workload when the L2 system was engaged. A similar pattern was observed by Radhakrishnan et al. (2022) who also found a reduction in physiological activation during partially automated driving, a pattern that the authors interpreted as a reduction in drivers' mental workload. Likewise, the decline in detection task performance observed by Biondi et al. (2018) during L2 driving was also interpreted as lower mental workload. These patterns are in conflict with the work by Lohani and McDonnell (Lohani et al., 2021; McDonnell et al., 2021; McDonnell et al., 2023) who, instead, found partial automation not to produce any changes in mental workload when compared to manual driving. Additional data by Kim et al. (2023) revealed an opposite pattern, showing that operating an L2 system increased drivers' workload. The authors attributed this increase to the additional demands resulting from supervising the functioning of the L2 system over an extended period of time.

Conflicting findings can also be observed in the literature investigating behavioral changes resulting from partially automated driving. Works by Solís-Marcos et al. (2018) and Biondi and Jajo (2024) have evidenced a reduction in forward glances accompanied by an increase in glances directed toward the vehicle's touchscreen when the L2 system was engaged. However, conflicting results were found by Gaspar and Carney (2019) who, despite observing a slight increase in glance duration toward the vehicle's touchscreen when the L2 system was engaged, failed to see an increase in the total time spent looking away from the road in this condition. Similarly, Goncalves et al. (2020) found that gaze concentration on the forward roadway did not differ between manual and L2 driving.

This seemingly fractured literature, combined with a greater presence of partially automated vehicles on our roads, justifies the need for a

deeper investigation of the differences in mental workload, visual behavior and NDRT engagement between manual and partially automated driving. With this said, the current study has three main objectives.

- (1) *Explore differences in mental workload between manual and partially automated driving.* There is conflicting evidence on the effect that operating partially automated systems have on drivers' mental workload. This pattern could possibly be the result of studies adopting a combination of self-reported, physiological, and behavioral approaches for mental workload assessment, which are often found to offer incongruent results (Matthews et al., 2020; Tao et al., 2019; Longo et al., 2022). Here, we investigate this topic with the goal of dissecting the diverse methodologies adopted in the literature.
- (2) *Explore differences in visual behavior between manual and partially automated driving.* With partially automated driving reducing the overall demands of driving, here we investigate the effect that the drastic transition in the driver's responsibilities has on their ability to maintain their eyes on the road when the L2 system is operational. Further analyses are conducted to investigate possible changes associated with operating simulated versus on-road partially automated systems.
- (3) *Explore differences in NDRT engagement between manual and partially automated driving.* Building on the work conducted for objective 2, this objective aims to further investigate the effect that operating a partially automated system has on NDRT engagement. Additional analyses are conducted to investigate potential age and sex-related differences.

Similar meta-analyses have investigated the human factors of automated driving. For example, earlier work by de Winter et al. (2014) explored changes in mental workload resulting from operating driving automation. However, in doing so the authors largely examined drivers' self-reported mental workload resulting from operating either driver assistance (or SAE level 1) or highly automated systems (or SAE level 4 and 5), with limited attention being paid to partially automated

driving. Weaver and DeLucia (2018) investigated the human factors of adopting driving automation requiring shared vehicle control, but their study largely centered on the transition of control between the human driver and the automated system during conditionally automated driving or SAE level-3 automation; a topic that was also investigated by the meta-analytical work by Zhang et al. (2019). Similar work was conducted by Shahini and Zahabi (2022) who, in addition to focusing on transitions of control between manual and automated driving, also explored the mental workload resulting from operating partially and highly automated systems. In our study we investigate changes in mental workload between manual and partially automated driving while taking into consideration potential differences resulting from adopting diverse self-reported, physiological, and behavioral metrics. Additionally, in all our analysis, we also examine the moderating effects of age and sex. Altogether, although previous meta-analytical works have addressed related topics, we believe the distinct focus of the current research makes a unique contribution to the Human Factors literature on driving automation.

## Method

In conducting this meta-analysis, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework was adopted (Page et al., 2021). The entire procedure is detailed in Figure 1. All the studies selected for review achieved a sufficient level of quality by reporting detailed descriptions of experimental designs and protocols. Therefore, each study was considered of equal importance and was not coded for quality (cf. Zhang et al., 2019). The ROB 2 tool (Sterne et al., 2019) was employed to assess the risk of bias in the selected papers. A detailed description of this assessment is provided in the Supplementary Material (see Supplemental Figures 1 and 2). The protocol was not preregistered. Data and scripts for the analyses are available upon request to the authors.

## Search Strategy

An in-depth search was conducted in five popular databases: Scopus, Web of Science (WoS), PubMed, PsycINFO, and PsycArticles. The search was

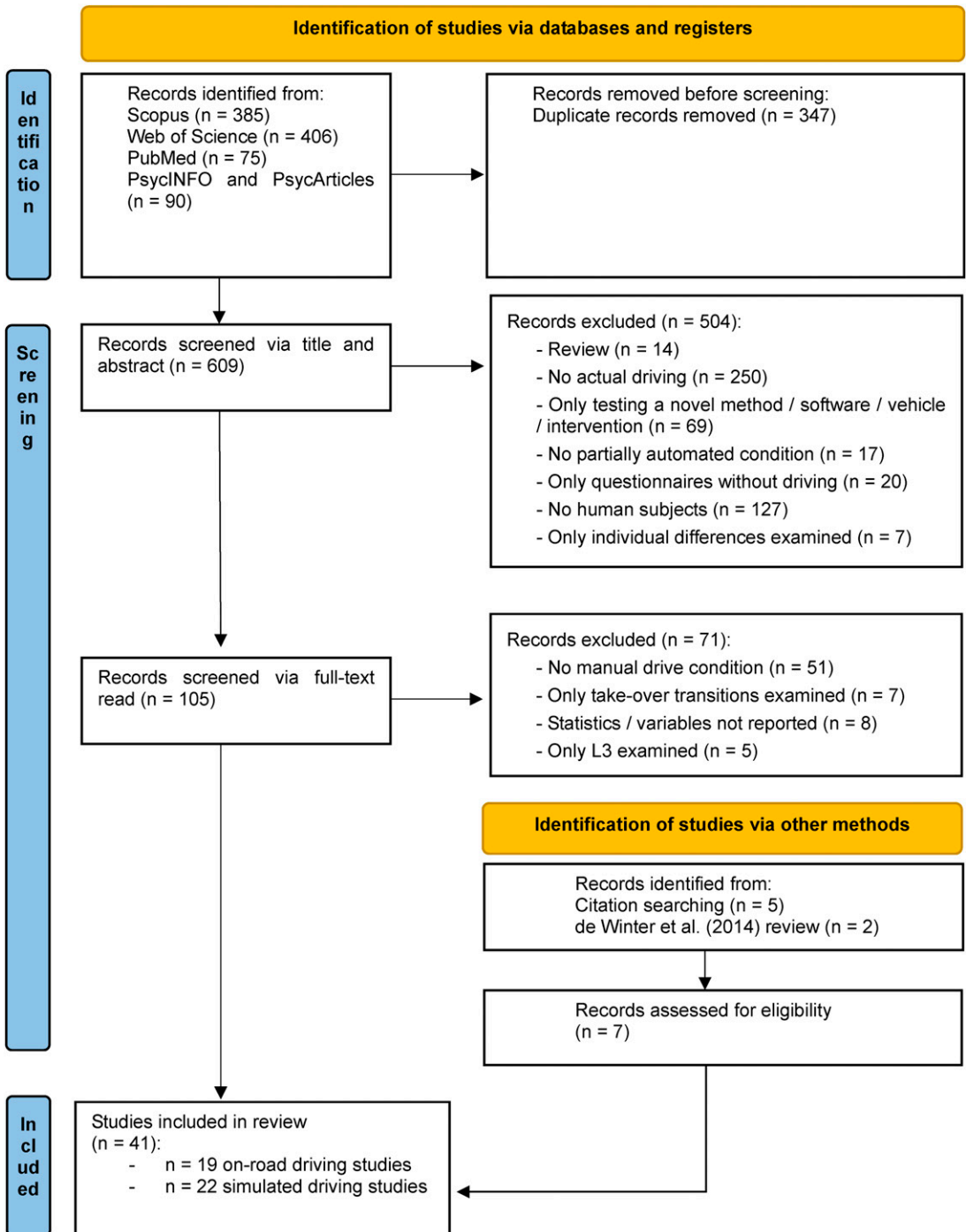


Figure 1. PRISMA flowchart, showing in detail identification, screening, and inclusion steps (Page et al., 2021).

conducted on May 26th of 2024 on peer-reviewed articles published between January 1st, 2010, and May 26th, 2024. January 1st, 2010, was selected as the start date as it predates the commercialization of partially automated systems as well as the first version of the SAE driving automation taxonomy published in 2016. Following Paré et al. (2015), we preferred to limit the study selection by using a carefully selected set of keywords (see Table 1), rather than adopting a more extensive approach. This was done for purposes of rigor and clarity and to ensure the complete reproducibility of the selection process. Once duplicates were removed,  $n = 609$  papers were included in the review for screening (see Figure 1 for a visual display of the selection process).

**Inclusion and Exclusion Criteria**

Studies were selected according to our inclusion and exclusion criteria (see Figure 1). Only empirical studies published in peer-reviewed journal or conferences and involving human participants were included in our review: review papers, methodological papers, studies based on simulations of data without actual participants, studies that did not involve human participants or a driving task, and studies that did not undergo a peer-review process were excluded. Studies must include both a manual driving condition and a partially automated driving condition as per the SAE taxonomy, and they must compare these conditions by testing differences in either mental

workload, visual behavior, or NDRT engagement. Studies investigating only partially automated driving, only manual driving, only take-over transitions or testing individual differences between participants were excluded. Moreover, we chose to exclude studies involving only SAE level 3 automation from the review and meta-analysis. This was done considering the meaningful differences between L2 and L3 systems in terms of operational design domains (e.g., L2 requires the driver to supervise the functioning of the automated at all times, whereas this is not required when the L3 system is engaged and operational), and also given the extremely limited literature on the comparison between manual and L3 driving in on-road experiments (only one study was identified). Studies testing novel software, vehicles, or interventions (e.g. training programs) were excluded as they can largely modify drivers’ behavior, therefore confounding potential differences between manual and partially automated driving (Casner & Hutchins, 2019; Ebnali et al., 2019). Finally, studies with asymmetries between driving mode conditions (i.e., nondriving-related task presented only during partially automated driving) or reporting nonquantitative data or incomplete data were excluded.

**Study Selection**

The screening was performed independently by the two authors, who reached agreement on the

**Table 1.** Keywords Included in the Review. Notes: The Asterisk (\*) Allow for Alternative Suffixes of These Keywords (e.g., Distract\* Stands for Both Distracted and Distraction).

Distraction and Mental Workload (OR)	(AND)	SAE Level of Automation (OR)	(AND)	Autonomous Driving (OR)
distract*, “risky driv*”, “risky behav*”, “inattent*”, “attention*”, “mental workload”, “cognitive workload”, “mental effort”, “cognitive load”, “situation* awareness”		“partial*”, “SAE level 3”, “SAE level-3”, “SAE level 2”, “SAE level-2”		“ADS”, “automated systems” “autonomous systems” “automated vehicles” “autonomous vehicles” “automated driving” “autonomous driving” “autonomous driving systems” “automated driver systems” “automation”

final selection. This was done to decrease the risk of bias and the risk of error in the selection process (Page et al., 2021). First, we reviewed the titles and abstracts to ensure that the papers met our inclusion criteria. Once this first screening was completed, selected papers were read in their entirety by both authors and then screened once again (see Figure 1). Next, a citation search was performed on the selected papers to identify articles that had not emerged from the keyword analysis (as suggested by Page et al., 2021). Furthermore, the review of de Winter et al. (2014) was screened to find studies that could match our inclusion criteria. Five additional articles were identified through the citation search, while two more were identified from the review of de Winter et al. (2014). When the same data were used by two different studies, the most complete study was retained, while the less complete study was included only if it presented relevant additional findings. After the study selection process was completed,  $n = 41$  studies were included in the present work, with 19 studies being on-road driving experiments and 22 simulated driving experiments.

### *Dependent Variables Extracted*

The variables relevant to our objectives were extracted and grouped according to the constructs they were intended to measure (i.e., mental workload, visual behavior or NDRT engagement). See below for a complete overview of the variables extracted and the rationale behind their inclusion.

**Mental Workload.** Mental workload assessments can vary depending on the metrics being used (Longo et al., 2022; De Waard, 1993; Butmee et al., 2019). Furthermore, there is not a single elective measure that can reliably discriminate between high and low mental workload (Charles & Nixon, 2019). Therefore, to obtain a comprehensive overview of the differences in mental workload between manual and partially automated driving, physiological, subjective, and behavioral metrics were included in the analyses. Descriptions of these metrics are presented below.

**Physiological Measures.** Spectral electroencephalogram (EEG) is sensitive to changes in mental workload. In particular, increased alpha and theta

power are usually associated with greater workload (McDonnell et al., 2023) and fatigue (Zhang et al., 2021). Similarly, differences in event-related potentials (ERPs) amplitude during an oddball paradigm (i.e., a secondary task involving the presentation of a series of identical sounds randomly interspersed with rare deviant stimuli) are thought to reflect variations in mental workload, with larger ERPs' amplitudes measured after each oddball sound corresponding to more mental resources available due to lower levels of mental workload (Figalová et al., 2024; Luck, 2014). Variations in blood flow as recorded by functional near-infrared spectroscopy (fNIRS) can indicate variations in the amount of mental workload experienced by the driver, with increased blood flow indicating higher workload (Saikia et al., 2021; Sibi et al., 2017). Pupil diameter and blink rate reflect variations in mental workload, with higher blink frequencies and increased pupil size indicating higher mental workload (Radhakrishnan et al., 2023; Tsai et al., 2007). Skin conductance level (SCL) and skin conductance response (SCR) have been shown to positively correlate with levels of mental workload during driving. Higher levels of drivers' mental workload are usually associated with both an increase in heart rate and a decrease in heart rate variability, as measured by the root mean square of successive differences (RMSSD) (Lohani et al., 2021; Radhakrishnan et al., 2022). It is important to note, however, that while we attribute these physiological changes to differences in mental workload, many of these measures are also sensitive to variations due to other mental or physical states or to measurement artifacts (cf. Luck, 2014).

**Behavioral Measures.** The ISO Detection Response Task (DRT) is a standardized metric of mental workload. It requires responding to the presentation of a visual, vibrotactile, or auditory stimulus presented every 3–5 s. Differences in DRT RT and accuracy are expected to reflect different levels of mental workload (ISO, 2015), with slower RT and lower accuracy indicating increased mental workload (for a review, see Biondi, 2024).

**Subjective Measures.** Self-reported metrics of mental workload can offer useful information

regarding the subjective mental workload experienced by drivers. The NASA Task Load Index (NASA-TLX) is the most common instrument requiring participants to self-rate their level of workload often on a 21-point Likert scale (Hart & Staveland, 1988). Most studies in this meta-analysis that used self-reported measures employed NASA-TLX scores or single scales of this questionnaire (see Figure 2). However, it is important to note that recent studies have questioned the validity of NASA-TLX, suggesting that subjective measures often diverge from other indicators, such as physiological and behavioral measures (e.g., de Winter, 2014; Rubio et al., 2004; Matthews et al., 2020).

**Visual Behavior.** Visual behavior can provide useful insights into drivers' visual attention (cf. Hungund & Kumar Pradhan, 2023). Therefore, we investigated visual behavior considering gaze and glances toward nondriving relevant areas (i.e., off-road glances, vertical and horizontal gaze dispersion, and gazes toward instruments panels or billboards) and away from driving relevant areas (i.e., gazes toward front road, hazards, or vehicle's mirrors).

**NDRT Engagement.** Similarly to visual behavior, engaging in NDRT (e.g., texting, holding the phone, talking to a passenger, using the radio, navigating with the GPS, and performing an easy ad-hoc experimental task) can provide information about attention allocation, with greater engagement in NDRT suggesting decreased attention toward driving relevant areas (cf. Dogan et al., 2019; Hungund & Kumar Pradhan, 2023).

### Meta-Analytic Approach

Mental workload, visual behavior and NDRT engagement were analyzed via three separate meta-analyses, conducted using the R package "metafor" (Viechtbauer, 2010). For the studies eligible for inclusion in the meta-analysis, relevant statistics were extracted: Cohen's  $d$  was calculated from means and standard deviations. When reported, partial eta squared ( $\eta_p^2$ ) was transformed to Cohen's  $d$  via the following formula (Lakens, 2013):

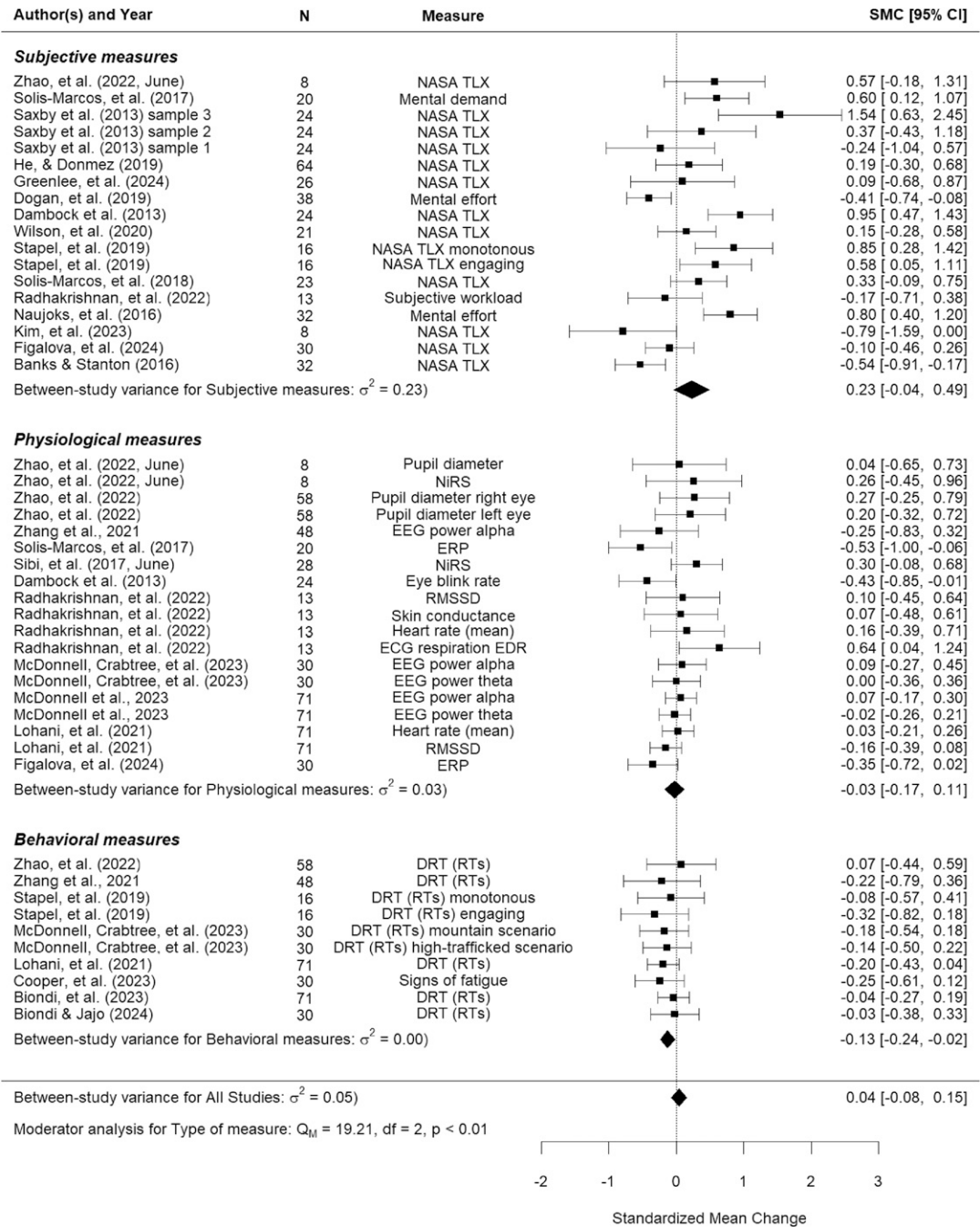
$$d = ((N - 1)/N) \times (\sqrt{\eta_p^2 / (1 - \eta_p^2)}).$$

Unlike raw mean differences, Cohen's  $d$  quantifies the mean difference on a standardized scale,

allowing to compare effect sizes calculated from different measures (Borenstein et al., 2021; Rosenthal & Dimatteo, 2001). When means and standard deviations were not provided, a graphic software (i.e., "PlotDigitizer") was used to extract the missing numerical values from figures. When these metrics were not available,  $t$ -values and  $p$ -values were used instead. When none of these were available for a particular result, that result was excluded from the meta-analysis (see Tables 2 and 3 for an overview of all the findings included in the meta-analysis). Since the majority of the studies included in this review used a within-subject design (i.e., 29 out of 40) we chose the Standardized Mean Changes (SMCs) as test statistic as it is appropriate when quantifying the difference between two within-subject conditions (Gibbons, 1993; Morris et al., 2002). Cohen's  $d$ ,  $t$ -values, and  $p$ -values were converted to SMCs using the "escal" function of the "metafor" package (Viechtbauer, 2010). Their respective sampling variance was calculated using the same function. The sampling variance quantifies the extent to which the effect size is expected to vary from study to study, with higher sampling variances indicating that the reported effect size may not be reliably replicable (Morris et al., 2002).

Next, we addressed potential issues arising from using different types of designs in the same meta-analysis following Morris et al. (2002)'s guidelines. First, we transformed effect sizes from different designs into a single metric, that is, the SMC. Second, we included the study design (between- vs. within-subjects) as a moderator variable in all meta-analyses. Additionally, since sampling variance depends on both sample size and study design (Morris et al., 2002), we conducted two different sampling variance estimations: within-subjects studies' sampling variance was estimated using Gibbons's formula, which accounts for repeated measures designs (Gibbons, 1993); between-subjects studies' sampling variance was estimated using Hedges' formula (Hedges, 1983, 1982), which accounts for independent groups designs (for reference, see equations A1 and A3 in Morris et al., 2002).

Finally, since multiple effect size estimates were included for each study, a multivariate Random Effect Model was preferred over a univariate model to account for the dependency among effect sizes



**Figure 2.** Forest plot of the Standardized Mean Changes (SMC) for mental workload. Negative SMCs indicate higher workload estimates during partially automated driving, while positive SMCs indicate higher workload during manual driving. Black squares represent the SMCs for each study, while the diamonds represent the aggregated SMCs estimated with random-effects models.  $Q_M$ -test: test of moderator, with significant results indicating that the variable is a significant moderator.

**Table 2.** Selected Studies Investigating Mental Workload, Visual Behavior and NDRT Engagement During On-Road Driving.

Authors	N	Sample Characteristics	Design	Drive Duration	Dependent Measures	Construct	Key Findings
Banks & Stanton (2016)	32	Age M(SD) 18–65 38 (10.8) Sex NA	Within	40 min	NASA Task Load Index	Workload	Higher overall workload during L2 compared to MAD. In detail, higher scores in mental demand, temporal demand, effort and frustration during automated driving
Biondi & Jejo (2024)	30	Age M(SD) 22 (4.36) Sex 13 F, 17 M	Within	80 min	Detection response task RTs and accuracy Total eyes off the road time (TEORT); Total glance time by area of interest (AOI); average and maximum glance duration by AOI	Workload Visual behavior	No differences between L2 and MAD in both RTs and accuracy Increased TEORT during L2 compared to MAD Increased total, average and maximum glance duration toward a touchscreen during L2 Mixed results for total, average and maximum glance duration toward rearview and side mirrors No differences between L2 and MAD overall However, RTs increased at a greater rate during L2 relative to MAD over time
Biondi et al. (2023)	71	Age M(SD) 40.8 (6.11) Sex 25 F, 46 M	Within	160 min	Detection response task RTs	Workload	No differences between L2 and MAD Higher workload during L2 compared to naturalistic driving (i.e., participants could choose whether to drive in L2 or manual mode) No differences between naturalistic driving and MAD.
Cooper et al. (2023)	30	Age M(SD) 18–55 35.73 (9.34) Sex 12 F, 18 M	Within	At least 40 min per day for 6–8 weeks	Signs of fatigue (coded from video tape) Manual NDRT engagement	Workload NDRT engagement	No differences between L2 and MAD Greater NDRT engagement during L2 compared to naturalistic driving Greater NDRT engagement during MAD compared to naturalistic driving
Dunn et al. (2021)	150 1. 30 2. 120	Age M(SD) 1. 46.5 (12.2) 2. 25–54 Sex 1. 8 F, 22 M 2. 60 F, 60 M	Within	Data collected for at least one year	Visual-manual NDRT engagement	NDRT engagement	1. More likely to engage in visual or manual NDRT during L2 compared to MAD. No differences in engagement in cognitive NDRT 2. No differences in engagement in visual or manual NDRT. More likely to engage in cognitive NDRT during MAD compared to L2 1. Greater %EORT during L2 compared to MAD 2. No differences in %EORT.
Figalová et al. (2024)	30	Age M(SD) 22–64 40.36 (13.73) Sex 16 F, 14 M	Within	120–150 min	Percentage of eyes off-road (EORT) NASA Task Load Index ERPs amplitude (auditory oddball)	Visual behavior Workload	No differences between L2 and MAD. Increased ERPs amplitude during MAD compared to L2, indicating that more attentional resources were available to process environmental sounds during MAD.
Gaspar & Carney (2019)	10	Age 25–62 Sex 4 F, 6 M	Within	13 min for 5 days	Mean and maximum glance duration toward touchscreen and instrument cluster	Visual behavior	Increased mean and maximum glance duration during L2 compared to MAD.
Kim et al. (2023)	8	Age M(SD) 44.13 (8.41) Sex 1 F, 7 M	Within	95 min	Total-eyes-off-road time (TEORT) NASA Task Load Index	Visual behavior Workload	No differences in mean TEORT Longest TEORT observed during L2 Higher perceived workload during L2 compared to MAD.
Kraff et al. (2018)	33	Age M(SD) 22–77 52.82 (15.93) Sex 3 F, 30 M	Between	120 min	Total glance duration (TGD) Perceived strain	Visual behavior Workload	Increased TGD toward a display and decreased TGD toward the road during L2 compared to MAD. No differences between L2 and MAD.

(continued)

Table 2. (continued)

Authors	N	Sample Characteristics	Design	Drive Duration	Dependent Measures	Construct	Key Findings
Lohani et al. (2021)	71	Age M(SD) 1. 39 2. 32 2. 41–64 52.72 (6.33) Sex 1. 13 F, 26 M 2. 12 F, 20 M	Within	80 min	Heart Rate Root mean square of successive differences in normal heartbeats Detection response task RTs	Workload Workload Workload	No differences between L2 and MAD. No differences between L2 and MAD. No differences between L2 and MAD.
McDonnell et al. (2023)	71	Age M(SD) 1. 39 2. 32 2. 41–64 52.2 (6.37) Sex 1. 13 F, 26 M 2. 12 F, 20 M	Within	80 min	EEG (frequency band power)	Workload	No differences on frontal theta power and on parietal alpha power
McDonnell et al., 2023	30	Age M(SD) 35.73 (9.34) Sex 12 F, 18 M	Within	72 min	Detection response task RTs EEG (frequency band power) Eye glance patterns	Workload Workload Visual behavior	Slower RTs during L2 compared to MAD RTs decreased over time only when driving under L2, and not during MAD. No differences in frontal theta power and parietal alpha power No differences between the group driving in L2 and the group driving in MAD.
Mueller et al. (2022)	20	Age M(SD) 1. 42 (8) 2. 47 (14) Sex 1. 5 F, 5 M 2. 5 F, 5 M	Between	60 min	Manual NDRT engagement Perceived mental effort	NDRT engagement Workload	No differences between L2 and MAD overall However, experienced drivers with L2 engaged more in NDRT during L2, compared to MAD. Higher perceived mental effort during MAD, compared to L2
Naajoks et al. (2016)	32	Age M(SD) 20–70 47.19 (16.08) Sex 8 F, 24 M	Within	60 min	Manual NDRT engagement Perceived mental effort	NDRT engagement Workload	No differences between L2 and MAD overall However, experienced drivers with L2 engaged more in NDRT during L2, compared to MAD. Higher perceived mental effort during MAD, compared to L2
Noble et al. (2021)	19	Age M(SD) 47.8 (9.83) Sex 6 F, 13 M	Within	Data collected for up to one year	Total eyes off road time (TEORT); Mean glance off roadway (MGOR); Single longest glance off-road (SLG); Number of off-road glances (NORG) Visual NDRT engagement	Visual behavior NDRT engagement	Greater TEORT, SLG, and NORG when automated driving was active, compared to when it was not active No differences in MGOR. No differences overall in the likelihood of engaging in NDRT However, more likely to browse on cell phone during L2
Reagan et al. (2021)	10	Age M(SE) 46.7 (5) Sex 4 F, 6 M	Within	Data collected for 4 weeks	Visual-manual NDRT engagement	NDRT engagement	No difference between L2 and MAD in overall NDRT engagement However, increased cell phone manipulations in the second part of the drive, but not in the first, during L2

(continued)

**Table 2.** (continued)

Authors	N	Sample Characteristics	Design	Drive Duration	Dependent Measures	Construct	Key Findings
Solis-Marcos et al. (2018)	23	Age M(SD) 43.5 (8.5) Sex 2 F, 21 M	Within	NA	NASA Task Load Index  Eye glances Visual NDRT engagement	Workload  Visual behavior NDRT engagement	No difference between MAD and L2 in overall workload No difference between MAD and L2 in groups with different levels of experience with L2 Higher frequency of glances toward the front and left mirror during MAD compared to L2 Longer and more glances toward the dashboard and to an additional visiomotor task during L2 compared to MAD Drivers experienced with L2 spent more time looking at an additional visiomotor task and the tablet during L2
Sapfel et al. (2019)	16	Age M(SD) 21–69 ~42 (14) Sex 1 F, 17 M	Within	180 min	Detection response task RTs NASA Task Load Index Heart rate	Workload Workload Workload	Slower RTs during L2 compared to MAD. Higher perceived workload during MAD compared to L2 However, no differences between MAD and L2 for inexperienced drivers No difference between L2 and MAD
Wilson et al. (2020)	21	Age M(SD) 22–68 41 (12) Sex 8 F, 13 M	Within	40 min	NASA Task Load Index	Workload	No differences between MAD and L2

Notes: MAD = manual driving; L2 = SAE level 2 driving. Key findings in italics are only discussed with a narrative approach due to insufficient information or to the originality of the finding. When a study includes more than one sample, the sample size and demographics are shown for each sample.

**Table 3.** Selected Studies Investigating Mental Workload, Visual Behavior and NDRT Engagement During Simulated Driving.

Authors	N	Sample Characteristics	Design	Drive Duration	Dependent Measures	Construct	Key Findings
Carsten et al. (2012) Two samples	49 / 25 2 24	Age M(SD) 47.8 (11) Sex 5 F, 44 M	Within	135 min	Eye gaze toward the road Visual NDRT engagement	Visual behavior NDRT engagement	Higher time spent watching the road during MAD, compared to L2 Less likely to engage in NDRT during manual driving, compared to L2
Damböck et al. (2013)	24	Age M(SD) 23–57 30.5 Sex 4 F, 20 M	Within	NA	NASA Task Load Index Eye blink rate Horizontal gaze dispersion	Workload Workload Visual behavior	Higher workload scores during MAD compared to L2 Increased blinks during L2 compared to MAD. Increased horizontal gaze dispersion during L2 compared to MAD.
Dogan et al., 2019	38	Age M(SD) ~37 (9) Sex	Within	12 min	Mental workload	Workload	Higher workload scores during L2 compared to MAD.
Eddine et al. (2024)	27	Age M(SD) 26 (3.62) Sex 13 F, 25 M	Within	20 min	Eye fixation on billboards and driving-related areas	Visual behavior	More and longer eye fixations during L2 compared to MAD Greater time fixating driving-related areas during MAD compared L2
Feldhütter et al. (2019)	45	Age M(SD) 39.04 (5.98) Sex 10 F, 35 M	Within	50–55 min	Visual behavior towards driving relevant areas in the different driving phases (attention ratio)	Visual behavior	More likely to watch driving relevant areas during MAD compared to L2
Goncalves et al. (2020)	29	Age M(SD) 21–60 34.21 (8.94) Sex 14 F, 15 M	Within	Under 120 min	Percentage of fixations on the road center; horizontal gaze dispersion	Visual behavior	No differences in percentage of fixations on the road center No differences in horizontal gaze dispersion
Greenlee, Delucia, & Newton, 2022	26	Age M(SD) 19.19 (1.32) Sex 15 F, 11 M	Between	40 min	NASA Task Load Index	Workload	No difference in the overall score between groups However, higher scores in the physical demand subscale during MAD compared to L2
Hatfield et al. (2019)	24	Age M(SD) 18–21 18.88 (1.4) Sex 14 F, 10 M	Between	15 min	Eye movements (scanning of the environment)	Visual behavior	No differences in vertical and horizontal eye movements between groups

(continued)

**Table 3.** (continued)

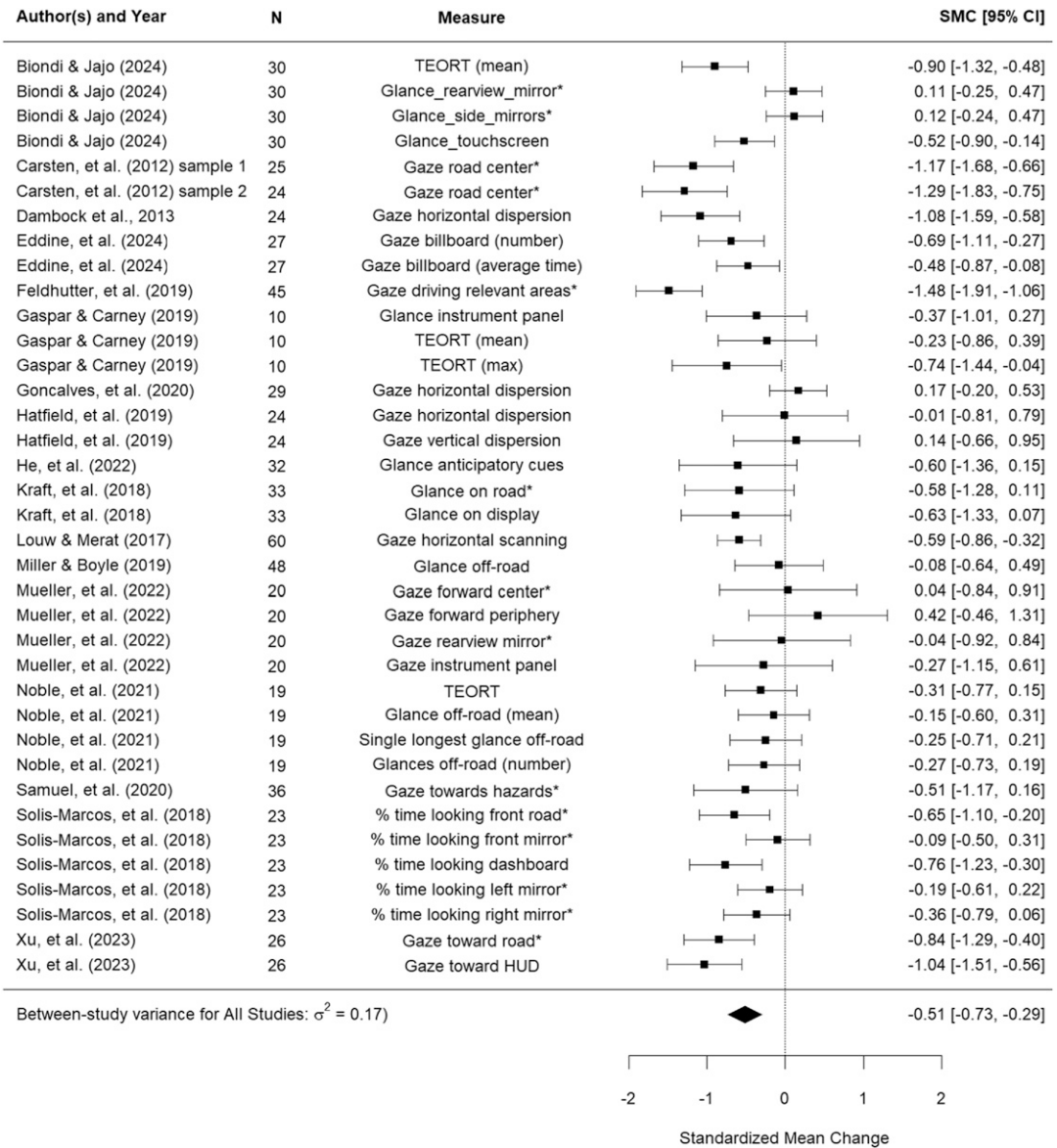
Authors	N	Sample Characteristics	Design	Drive Duration	Dependent Measures	Construct	Key Findings
He & Dommez (2019)	64	Age M(SD) 18–58 28.8 (5.1) Sex 32 F, 32 M	Between	20 min	NASA Task Load Index Visual-manual NDRT engagement	Workload NDRT engagement	No differences in workload between groups More manual interactions with NDRT and longer glances toward NDRT in the group driving with L2 compared to the group driving with MAD No differences in rate of glances toward NDRT between MAD and L2
May; He et al., 2022 Two studies with the same dataset	32	Age M(SD) 28.8 (6.8) Sex 18 F, 18 M	Between	20 min	Eye glances toward anticipatory cues, roadway	Visual behavior	More time spent looking at anticipatory cues in the group driving in L2 than the group driving in MAD No differences between the two groups when considering glances to roadway
Louw & Merat (2017)	60	Age M(SD) 36.16 (12.38) 28 F, 32 M	Within	40 min	Visual NDRT engagement Eye movements (horizontal scanning and gaze pitch)	NDRT engagement Visual behavior	More time spent looking at the NDRT in the group driving in L2 than the group driving in MAD. Higher horizontal scanning during L2 compared to MAD.
Miller & Boyle (2019)	48	Age M(SD) 25–54 38.5 (9.4) Sex 24 F, 24 M	Between	40 min for 3 days	Eyes-off-road glances	Visual behavior	No differences between groups in mean glance duration overall However, longer mean glance duration in the last part of the drive compared to the first part during L2, but not during MAD.
Radhakrishnan et al. (2022)	13	Age M(SD) 42 (17) Sex 4 F, 9 M	Within	18 min	ECG-derived respiration rate Root mean square of successive differences in normal heartbeats Mean heart rate Number of skin conductance responses per minute	Workload Workload Workload Workload	Higher respiration rates during MAD compared to L2 No differences between MAD and L2 No differences between MAD and L2 No differences between MAD and L2
Samuel et al. (2020)	36	Age M(SD) 18–25 22.25 (1.82) Sex 14 F, 22 M	Between	12 min	Self-reported workload ratings Eye movements anticipation towards hazards	Workload Visual behavior	No differences between MAD and L2 No differences between the group driving in L2 and the group driving in MAD in eye movements anticipation toward hazards
Saxby et al. (2013) Three samples	72 2. 24 3. 24	Age M(SD) 1. 24 18–40 19.92 (2.65) Sex NA	Between	10, 30 and 50 min	NASA Task Load Index	Workload	No differences in the first two samples driving for 10 and 30 min, respectively Higher workload for the MAD group compared to the L2 group in the third sample, that drove for 50 min

(continued)

Table 3. (continued)

Authors	N	Sample Characteristics	Design	Drive Duration	Dependent Measures	Construct	Key Findings
Sibi et al. (2017)	28	Age M(SD) 17–71 31.11 (12.76) Sex 10 F, 18 M	Within	22 min	fNIRS	Workload	No differences between MAD and L2
Solis-Marcos et al. (2017)	20	Age M(SD) 22–34 27.1 (3.8) Sex 9 F, 11 M	Within	5 min	Mental demand sub-scale in the NASA Task Load Index ERP amplitude (auditory oddball)	Workload Workload	Higher mental demand during MAD compared to L2 Increased ERPs amplitude during MAD compared to L2
Xu et al. (2023)	26	Age 23–50 Sex 6 F, 20 M	Within	20 min	Eye gaze toward road and HUD	Visual behavior	More likely to gaze on the HUD and less likely to gaze on the road during L2 compared to MAD.
Zhang et al. (2021)	48	Age M(SD) 20–35 24.83 (2.81) Sex 24 F, 24 M	Between	60 min	NASA Task Load Index Detection Response Task RTs and accuracy	Workload Workload	No differences between MAD and L2 No differences between MAD and L2 in both RTs and accuracy overall
Zhao, Liu, et al., 2022	58	Age M(SD) 20–30 22.03 (2.08) Sex 29 F, 29 M	Between	60 min	EEG (frequency band power) Detection Response Task RTs	Workload Workload	However, slower RTs during L2 after 40 min of drive No differences in alpha power overall. However, increased alpha power during L2 after 40 min of drive No differences between groups in overall RTs However, faster RTs in the first 10 min of driving for the group driving in L2 compared to the group driving in MAD.
Zhao, Liu, et al., 2022, June	8	Age M(SD) 25–42 29.5 (5.39) Sex 3 F, 5 M	Within	NA	Pupil diameter NASA Task Load Index fNIRS Pupil diameter	Workload Workload Workload	Left eye: No differences in the overall pupil diameter However, smaller pupil diameter in the first 10 min of driving during L2 Right eye: Smaller pupil diameter during L2 compared to MAD Moreover, smaller pupil diameter in the first 20 min of driving during L2 Higher workload scores during MAD compared to L2 No differences in oxygenated blood No differences in pupil diameter

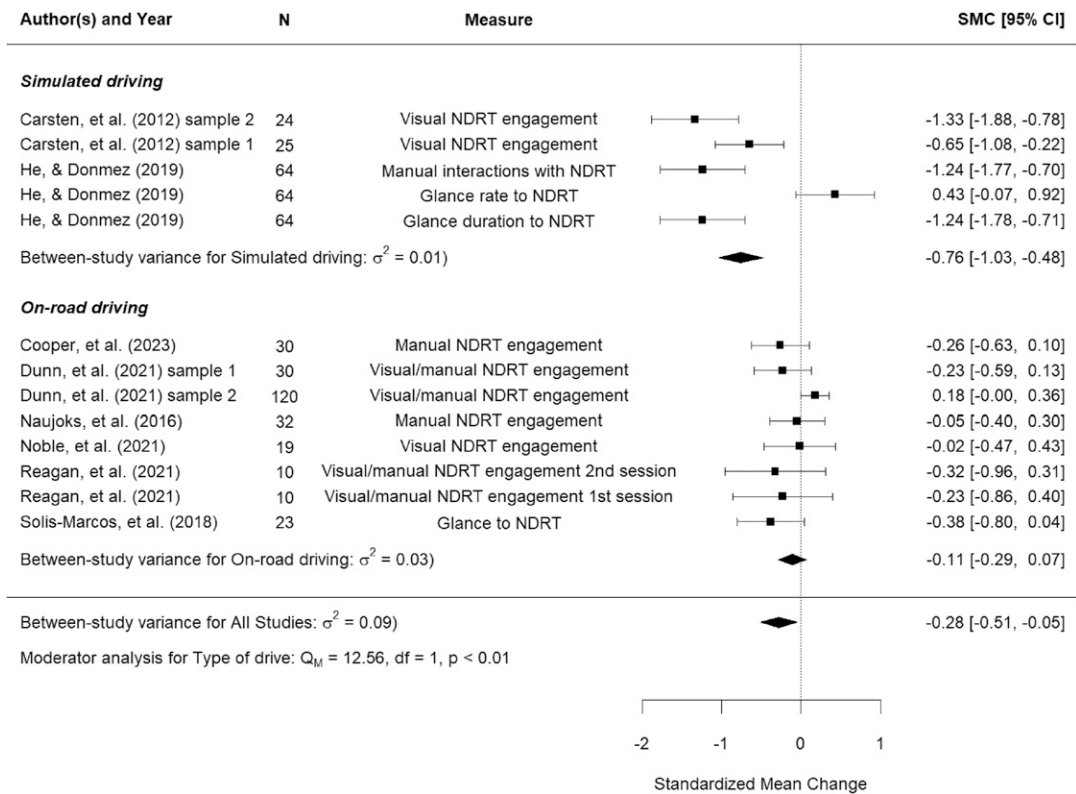
Notes: MAD = manual driving; L2 = SAE level 2 driving. Key findings in *italics* are only discussed with a narrative approach due to insufficient information or to the originality of the finding. When a study includes more than one sample, the sample size and demographics are shown for each sample.



**Figure 3.** Forest plot of the Standardized Mean Changes (SMC) for visual behavior. Negative SMC indicate reduced gazes or glances toward driving relevant areas and increased gaze or glances toward nondriving relevant areas during partially automated driving. Black squares represent the SMCs for each study, while the diamond represents the aggregated SMC estimated with a random-effects model. Measures marked with \* are reverse-coded.

originating from the same study (Berkey et al., 1996; Konstantopoulos, 2011; Olkin & Gleser, 2009). The estimates for the Random Effect Models were computed using the restricted maximum likelihood estimator (Viechtbauer, 2005; Raudenbush, 2009). Moderator analysis was conducted to explore the influence of the Type of drive

(on-road driving vs. simulated driving) on the differences between partially automated and manual driving in each meta-analysis, as drivers might behave differently in real road conditions compared to simulations. Moderator analyses were also conducted to examine how participants' sex (i.e., percentage of females) and the average age of



**Figure 4.** Forest plot of the Standardized Mean Changes (SMC) for NDRT engagement. Negative SMC indicate increased NDRT engagement during partially automated driving. Black squares represent the SMCs for each study, while the diamonds represent the aggregated SMCs estimated with random-effects models.  $Q_M$ -test: test of moderator, with significant results indicating that the variable is a significant moderator.

drivers in each study influenced differences in mental workload, visual behavior, and NDRT engagement. This was done given that there is evidence that age and sex might affect some of these variables (cf. Cantin et al., 2009). Additionally, a moderator analysis was performed to investigate the influence of the Type of measure used (physiological vs. behavioral vs. subjective) when assessing mental workload.

### Publication Bias

Publication bias refers to the phenomenon by which nonsignificant results are less likely to be published in peer review journals and conferences than significant results. To assess publication bias, we performed rank correlation tests for funnel plot asymmetry using the Kendall's tau statistics included in the R package "metafor" (Begg & Mazumdar, 1994; Viechtbauer, 2010). This test

examines the correlation between the absolute values of effect sizes and their corresponding sampling variances. A significant correlation indicates that larger effect sizes come from studies having high sampling variance (i.e., studies with small sample sizes and/or between-subject designs), which suggests the presence of publication bias. A nonsignificant correlation indicates that the effect sizes are not dependent on sample size or design, thus suggesting the absence of publication bias.

### Results

Characteristics of on-road and simulated driving studies are described in Tables 2 and 3, respectively. Additional study characteristics can be found in Supplemental Tables 1 and 2 in the Supplementary Material. In this section, we present the aggregated characteristics of the studies and the results of the meta-analytic method.

## Study Characteristics

$N = 41$  articles, encompassing 47 experiments, were included in the meta-analysis. Studies ranged from 2012 and 2024, and reported data from 1482 participants, of which 537 were female and 841 males (when reported). Participants were mostly recruited through advertisements posted online or in universities, while only one study (i.e., Feldhütter et al., 2019) recruited participants from among automaker employees.

## Meta-Analysis Results

The results of each Random Effect Model are presented in detail below: the estimates of the Standardized Mean Change (SMC) are presented within text and in Figures 2, 3, and 4, while the between-study variance component ( $\sigma^2$ ) is presented in Figures 2, 3, and 4. The assessment of heterogeneity across studies (Q) was also conducted and reported within the text. Significant Q-values suggest that the true effect sizes are heterogeneous, while nonsignificant Q-values indicate that the variability in the observed effect sizes is smaller than would be expected based on sampling variability alone, and that the true effect sizes are relatively consistent across studies (Cochran, 1954). The size of the SMC estimates can be interpreted in the same way as Cohen's  $d$ : a small mean change is 0.20, a medium mean change is 0.50, and a large mean change is 0.80 (Hedges & Olkin, 2014; Cohen, 2013). The design of the experiment (within- vs. between-subjects) was not found to be a significant moderator in any meta-analyses (mental workload:  $Q_M(1) = 0.985, p = .321$ ; visual behavior:  $Q_M(1) = 2.196, p = .138$ ; NDRT engagement:  $Q_M(1) = 1.210, p = .271$ ). Hence, following Morris et al. (2002)'s guidelines, it was not included in the analyzes.

**Mental Workload.** For the purpose of the present meta-analysis, negative SMCs indicate higher workload during partially automated mode, while positive SMCs indicate higher workload during manual mode. In total, 47 different effects were collected from 26 independent samples analyzing mental workload, with 13 employing on-road driving experiments and 13 employing simulated driving experiments. A two-level multivariate

Random Effect Model was conducted, including effects sizes and sampling variances in the first level of analysis and the studies in the second level as random intercepts. This analysis did not result in a significant difference (SMC = 0.039,  $p = .503$ , 95% C.I. = [-0.076, 0.154]), suggesting that, overall, there are no differences in the levels of mental workload between manual and partially automated driving. Moderator analyses indicated that the Type of drive (on-road vs. simulated), the average age of drivers, and the percentage of females in the sample were not significant moderators ( $Q_M(1) = 1.180, p = .277$ ;  $Q_M(1) = 0.404, p = .525$ ; and  $Q_M(1) = 2.151, p = .142$ , respectively). In contrast, the Type of measure (physiological vs. subjective vs. behavioral) emerged as a significant moderator ( $Q_M(2) = 19.21, p < .001$ ). In detail, the Random Effect Model including behavioral measures indicated a small negative effect size (SMC = -0.132,  $p = .019$ , 95% C.I. = [-0.242, -0.022]), suggesting that mental workload was slightly higher during partially automated driving. In contrast, the Random Effect Model including physiological and subjective measures did not indicate a significant difference (physiological: SMC = -0.028,  $p = .698$ , 95% C.I. = [-0.171, 0.115]; subjective: SMC = 0.226,  $p = .093$ , 95% C.I. = [-0.038, 0.489]).

Finally, the heterogeneity assessment revealed that subjective workload was the only group with significantly heterogeneous effect sizes ( $Q(16) = 76.93, p < .001$ ), suggesting that variations in subjective measures of workload might be influenced by other unknown factors. Effect sizes estimated from physiological ( $Q(18) = 24.99, p = .125$ ) and behavioral measures ( $Q(9) = 2.90, p = .968$ ) did not show significant heterogeneity. To explore this further, we reasoned on the possibility that subjective mental workload could be influenced by the presence (or absence) of an NDRT. Five studies included an NDRT in both driving modes, while twelve studies did not include any NDRT. Moderator analysis showed that the presence of an NDRT was a significant moderator ( $Q_M(1) = 4.29, p = .038$ ). Specifically, a positive effect size was found when an NDRT was included in both driving modes (SMC = 0.578,  $p < .001$ , 95% C.I. = [0.308, 0.847]), indicating that subjective mental workload was higher during manual driving compared to partially automated driving.

In contrast, when no NDRT was present, the analysis did not reveal a significant difference (SMC = 0.059,  $p = .716$ , 95% C.I. = [-0.262, 0.381]).

**Visual Behavior.** In this meta-analysis, we attempted to quantify the visual behavior toward both driving and nondriving-related areas. Therefore, variables measuring visual behavior toward drive-relevant areas were reverse coded, so that positive SMCs indicate reduced gaze or glances toward driving relevant areas and increased gaze or glances toward nondriving relevant areas during manual driving, while negative SMCs indicate reduced gaze or glances toward driving relevant areas and increased gaze or glances toward nondriving relevant areas during partially automated driving (in Figure 3, measures marked with an asterisk “\*” are reverse coded).

In total, 37 different effects were collected from 17 independent samples, with 6 involving on-road driving experiments and 11 involving simulated driving experiments. A two-level multivariate Random Effect Model was conducted, including effect sizes and sampling variances at the first level of analysis and the studies as random intercepts at the second level. This analysis resulted in a significant negative mean change (SMC = -0.513,  $p < .001$ , 95% C.I. = [-0.733, -0.294]), suggesting that, overall, drivers’ visual behavior increased toward nondriving relevant areas and decreased toward driving relevant areas during partially automated driving. Moderator analysis indicated that this tendency was slightly stronger in samples with lower percentages of females ( $Q_M(1) = 6.58$ ,  $p = .010$ ,  $\beta = 0.014$ ). In contrast, the Type of drive and the average age of drivers were not significant moderators ( $Q_M(1) = 2.30$ ,  $p = .129$ , and  $Q_M(1) = 0.46$ ,  $p = .498$ , respectively).

Notably, the heterogeneity assessment indicated that the effect sizes were significantly heterogenous ( $Q(36) = 114.12$ ,  $p < .001$ ; see Figure 3), suggesting that visual behavior could be influenced by other unknown factors. To investigate this, we tested whether glances to different areas produced distinct results. Specifically, we hypothesized that excluding glances toward the instrument panel might reveal distinct visual behaviors when compared to glances directed toward other off-road areas (i.e., glances outside the window, vertical and horizontal

dispersion) or on-road areas (i.e., front road and mirrors). Therefore, we included the Type of area (three levels: on-road, off-road, and instrument panel) as a moderator, finding a significant effect ( $Q_M(2) = 9.18$ ,  $p = .010$ ). In detail, all areas showed significant negative mean changes, replicating the general trend outlined in the main analyses, with increased glances toward instrument panels (SMC = -0.655,  $p < .001$ , 95% C.I. = [-0.874, -0.436]), increased off-road glances (SMC = -0.395,  $p < .001$ , 95% C.I. = [-0.626, -0.164]) and decreased on-road glances (SMC = -0.620,  $p = .008$ , 95% C.I. = [-1.078, -0.162]) during partial automation.

**NDRT Engagement.** In this meta-analysis, visual and manual NDRT engagement was assessed: positive SMCs indicate increased NDRT engagement during manual driving, while negative SMCs indicate increased NDRT engagement during partially automated driving.

In total, 13 different effects were collected from 10 independent samples, with 7 involving on-road driving experiments and 3 involving simulated driving experiments. A two-level multivariate Random Effect Model was conducted, incorporating effect sizes and sampling variances at the first level of analysis and the studies as random intercepts at the second level. This analysis revealed a small but significant negative mean change (SMC = -0.281,  $p = .017$ , 95% C.I. = [-0.513, -0.050]), suggesting that NDRT engagement was slightly greater during partially automated driving. Moderator analysis indicated that the Type of drive was a significant moderator ( $Q_M(1) = 12.56$ ,  $p < .001$ ), with simulated driving studies showing a large negative mean change (SMC = -0.756,  $p < .001$ , 95% C.I. = [-1.034, -0.479]) while on-road driving studies did not show a significant mean change (SMC = -0.109,  $p = .237$ , 95% C.I. = [-0.288, 0.071]). Notably, only three samples involved simulated driving experiments, suggesting that the overall SMC was likely driven by few studies employing simulated driving paradigms. In contrast, the average age of drivers and the percentage of females in the sample were not significant moderators ( $Q_M(1) = 0.10$ ,  $p = .746$ ;  $Q_M(1) = 1.45$ ,  $p = .228$ , respectively).

Finally, heterogeneity assessment indicated that only the three simulated driving studies showed significantly heterogeneous effect sizes ( $Q(4) =$

32.70,  $p < .001$ ), whereas the remaining on-road studies did not show significant heterogeneity ( $Q(7) = 11.67, p = 0.112$ ).

### Publication Bias Assessment

The rank correlation tests revealed a significant positive relationship between SMCs and their sampling variances (Kendall's tau = 0.341,  $p < .001$ ), suggesting the presence of publication bias for mental workload. Similar analyses revealed the presence of a publication bias for NDRT engagement (Kendall's tau = 0.564,  $p = .007$ ). The rank correlation tests resulted nonsignificant for visual behavior (Kendall's tau = 0.099,  $p = .398$ ), suggesting the absence of publication bias.

## Discussion

The present work investigated the differences in mental workload, visual behavior, and NDRT engagement between manual driving and partially automated driving. Here, findings are presented by research objective. Each section begins with a discussion of the results of the meta-analysis, followed by a subsection titled "Additional considerations" which includes relevant findings not included in the meta-analysis.

### Explore Differences in Mental Workload Between Manual and Partially Automated Driving

**Meta-Analysis Results for Mental Workload.** Overall, our meta-analysis did not reveal a general difference in mental workload between partially automated and manual driving. Contrary to our hypothesis, we did not find increased workload during manual driving, suggesting that both partially automated and manual driving results in similar mental workload. The publication bias assessment indicated that some nonsignificant results were not published/available. Indeed, even in this review, we were unable to include the nonsignificant results of three studies due to insufficient information (i.e., Kraft et al., 2018; Stapel et al., 2019; Zhang et al., 2021). Given the results of the meta-analysis and the presence of publication bias, it is unlikely that differences in mental workload exist between partially

automated and manual driving. This finding holds particular significance for Human Factors researchers and automobile manufacturers, as it suggests that both manual and L2 mode yield similar levels of mental workload. Notably, such finding is not entirely inconsistent with prior meta-analyses on the same topic. For instance, de Winter et al. (2014) conducted a comprehensive meta-analysis that indicated higher mental workload during manual driving compared to automated driving. However, their analysis included only self-reported measures and encompassed both partially and highly automated systems. In our analysis, when considering subjective measures alone, mental workload appears higher (though on a trending nonsignificant level) during manual driving, consistent with de Winter's findings. Additionally, while de Winter's review focused primarily on simulated studies and included highly automated systems, our study is limited to L2 systems and incorporates a substantial number of on-road studies published after de Winter's review.

Based on our results, the shift in the driver's role from vehicle operator to system supervisor does not consistently lead to a reduction in mental workload. Previous research has already noted that monitoring automated systems can impose a higher workload due to the novelty of the system and the increased number of tasks drivers may perform during partial automation (cf. the EAST framework, Banks & Stanton, 2016, 2019). In this meta-analysis, subjective mental workload was found to be higher during manual driving compared to partially automated driving when drivers were required to perform secondary tasks (SMC = 0.578), but no difference emerged when no secondary tasks were present. This suggests that, in the absence of secondary tasks, drivers perceive similar levels of workload in both driving modes. In other words, supervising the automated system may involve a different set of tasks than manually operating the vehicle, potentially reducing perceived mental workload when secondary tasks are present, but not when they are absent.

Moderator analyses suggested that mental workload also differed depending on the type of measure used. When behavioral measures were employed, a slightly higher mental workload was found during partially automated driving (SMC = -0.132,  $p = .019$ ). Although this might

seem an interesting result, we believe that some considerations should be made. First, it is important to note that the aggregated effect size is small and only two out of eight studies reported a significant main effect (Mcdonnell et al., 2023; Stapel et al., 2019): This, combined with the presence of publication bias, suggests that such an effect is clearly hard to find and might depend on other factors, such as previous experience with automated systems (see below for more details on this topic). Second, as shown in Figure 2, nine out of ten behavioral measures consisted of reaction times during a Detection Response Tasks (DRT). Here, slower reaction times are generally thought to indicate fewer available resources due to the increased mental workload imposed by the primary driving task (cf. ISO, 2015). However, recent research suggests that poorer DRT performance may also be the result of lower mental workload (Biondi, 2024; Biondi et al., 2023). Therefore, to understand the meaning of DRT reaction times, they must be compared with other measures, such as subjective and physiological measures (Biondi, 2024; Charles & Nixon, 2019).

Overall, the meta-analytic findings seem to indicate that partially automated and manual driving likely impose similar levels of mental workload on the driver. However, they also indicate that assessing mental workload with different measures can result in different outcomes. Researchers and automobile manufacturers that are planning to use behavioral measures to assess mental workload are invited not to use only these measures in the future, but compare them with (at least) self-reported or physiological measures (cf. Biondi, 2024).

**Additional Considerations on Mental Workload.** Stapel et al. (2019) examined the effect that the greater experience using L2 systems have on mental workload, finding that only experienced drivers reported lower perceived workload during partially automated compared to manual driving. Similarly, two studies including only participants inexperienced with partially automated systems did not find any significant differences in self-reported workload (Biondi et al., 2023; Biondi & Jajo, 2024), while one study testing drivers inexperienced with automation found slower DRT reaction times during partially automated driving

(Mcdonnell et al., 2023). This pattern seems to align with the work by Dunn et al. (2021) positing that, as drivers become more experienced using vehicle automation, this may lead to greater system complacency and a higher risk of engaging in distracting activities. Within the context of our study, we argue that this pattern could be the result of the lower mental workload experienced by expert L2 users who may seek engagement in NDRT to counter the declining workload.

Some studies tested differences in mental workload over time, reaching mixed results. Three studies found that DRT reaction times increased at a greater rate in L2 mode arguably indicating greater workload over time (Biondi et al., 2023; Zhang et al., 2021; Zhao, Liu, et al., 2022). Similarly, Saxby et al. (2013) found higher NASA-TLX scores during manual driving after a 50-min drive but observed no differences between partially automated and manual driving during both 10- and 30-min drives, suggesting that differences in perceived mental workload may only emerge after a certain amount of time. In contrast, Mcdonnell et al. (2023) and Zhao et al. (2022) found a seemingly opposite pattern with faster DRT RT and a smaller pupil size in the latter section of the L2 automated drive, patterns that would indicate a temporal reduction in mental workload. Overall, while mental workload seems to vary over time, no particular conclusions can be made from these findings. Future studies should consider time when assessing the impact of automated systems on psychological factors.

Among the studies under consideration, only one (Cooper et al., 2023) adopted a naturalistic approach wherein it was up to the driver to decide whether and when engage the L2 system. The authors compared the driver's workload experienced during this naturalistic portion of the study with that recorded during the experimental phase of the research, that is, when drivers were instructed to operate the vehicle in either manual or L2 mode. Results showed that, while a reduction in workload was found in the L2 mode during the experimental phase, no differences between the two modes were found in the naturalistic phase. This pattern is particularly relevant as it suggests that the workload associated with partial automation might stem more from drivers being forced to use automation, rather than being a direct

consequence of automation itself (Cooper et al., 2023). It is also important to note that in all remaining studies, the experimenter was present inside the vehicle (see Supplementary Table 1 for a complete overview), which might skew the generalizability of the meta-analysis results to naturalistic driving conditions (Safi et al., 2014).

### *Explore Differences in Visual Behavior Between Manual and Partially Automated Driving*

*Meta-Analysis on Visual Behavior.* The meta-analytic approach revealed a moderate aggregated effect size, indicating an increase in eye glances and gazes toward nondriving-related areas and a decrease toward driving-related areas during partially automated driving compared to manual driving (SMC =  $-0.513$ ,  $p < .001$ ). Notably, even when removing glances toward the instrument panel from the analyses, results still show an increase in off-road glances (SMC =  $-0.395$ ) and a decrease in on-road glances (SMC =  $-0.620$ ) during partially automated driving. These findings align with the hypothesis outlined in our objectives, suggesting that partial automation allows drivers to visually explore their surroundings by freeing up some of the mental resources devoted to driving (cf. Eddine et al., 2024; Solís-Marcos et al., 2017). Additionally, the absence of publication bias among these effects (Kendall's tau =  $0.099$ ,  $p = .398$ ) indicates that the size of the aggregated effect is not influenced by studies with small sample sizes or between-subject designs.

These findings show that drivers are more inclined to direct their gaze away from the road during partially automated driving. This is consistent with the hypothesis that, during L2 driving, relinquishing control of the vehicle to the automated system may lead some drivers to boredom. In an attempt to counter the impending state of underload, drivers may then start to direct their attention away from driving and toward the surrounding nondriving environment as a way to self-regulate (cf. Biondi, 2024, Engström et al., 2013). The self-regulation hypothesis posits that, as driving demands shift away from desired levels, drivers start to regulate their behavior to shift workload back to optimal levels (cf. Dunn et al.,

2021). While this occurs in conditions of increasing driving demands—for example, a driver that silences the radio or hangs up a call when negotiating a challenging maneuver—it is also frequent in situations of lower workload—for example, a driver that starts fidgeting or picks up their smartphone when waiting at a red light.

These behaviors are particularly perilous as they may detract from the driver's ability to promptly respond to road hazards. It is plausible that, as visual attention is directed away from the road, it becomes more challenging for drivers to properly maintain awareness of the surrounding traffic conditions and react to emerging threats (He et al., 2022; Gaspar & Carney, 2019; see also Merat et al., 2019 for a similar conclusion based on the “Out of the loop” algorithm). With this said, we recognize that characteristics such as the duration and frequency of off-road glances should also be accounted for when informing on their distraction potential. In the *Visual-Manual Driver Distraction Guidelines for In-Vehicle Electronic Devices*, the National Highway Traffic Safety Administration (NHTSA, 2013) spells out what specific visual behaviors yield higher crash risks. For visual-manual tasks to be considered distracting, single glances away from the road should last longer than 2 seconds, with the total eyes-off-the-road-time exceeding 12 seconds. Though useful, yet considered perhaps too stringent by some (Aust et al., 2013; Kidd et al., 2017; Pournami et al., 2015), it is important to note that these guidelines were developed for completing visual-manual tasks during manual driving, a scenario that might not be representative of the human factors of partially automated driving. With our findings showing an increased tendency to look away from the road when the L2 system is engaged, we believe it is important to investigate possible interventions that might counter these potentially perilous behaviors. Recent works have shown that both tactile and auditory warnings can help mitigate the distraction caused by inappropriate visual behavior during partial automation (e.g., Gruden et al., 2022; Mahajan et al., 2021). The effectiveness of driver training should also be explored, as previous research has shown that providing drivers with information regarding the automated system may help them at any level of expertise (cf. Hungund & Kumar Pradhan, 2023). Furthermore, particular

attention should be given to the system design. For instance, the work of Wang et al. (2024) provides recommendations on human-machine interfaces in vehicle automation, suggesting that the use of auditory alerts should be preferable to visual ones. Automobile manufacturers are encouraged to follow the guidelines available in the literature when developing partially automated systems.

Interestingly, our analyses revealed a higher tendency to glance toward nondriving relevant areas in more male-dominated cohorts, that is, in samples with lower percentages of female drivers. This is interesting as this is among the first studies evidencing sex-related differences during partially automated driving. While it is established that males, especially younger ones, are more inclined to risk tasking when driving (Fillmore et al., 2008; Hasanat-E-rabbi et al., 2021), little evidence was available on whether this would translate to operating partially automated driving.

*Additional Considerations on Visual Behavior.* Only one study presented findings that could not be included in the meta-analysis. Miller and Boyle (2019) found that off-road glances tend to increase over a 40-min period during partially automated driving, but not during manual driving. This is an interesting result that suggests that drivers might become distracted more quickly during partial automation. However, no claims can be done based on a single study.

### *Explore Differences in NDRT Engagement Between Manual and Partially Automated Driving*

*Meta-Analysis on NDRT Engagement.* The meta-analytic approach revealed a small effect size, indicating increased NDRT engagement during partially automated driving compared to manual driving ( $SMC = -0.281, p = .017$ ). This outcome aligns with our expectations, suggesting that drivers are more inclined to engage in nondriving tasks (and potentially be distracted by them) during partially automated driving. Moderator analyses revealed that the type of drive was a significant moderator, showing a large effect in simulated driving experiments ( $SMC = -0.756$ ) and no effect in on-road experiments

( $SMC = -0.109, p = .237$ ). Although this meta-analysis only examined results from a limited set of studies ( $n = 8$ ), it is important to note that the aggregated effect sizes consistently indicate greater NDRT engagement during partially automated driving. In addition, of the eight studies under consideration, only two (Cooper et al., 2023; Noble et al., 2021) adopted a naturalistic approach, whereas the remainder was conducted with drivers operating on-road partially automated vehicles while being monitored by researchers or wherein the use of devices like cellphones was outright restricted. Finally, the publication bias assessment revealed the presence of a publication bias, indicating that the aggregated effect size is likely driven by studies with small sample sizes and/or between-subjects designs. In sum, although the evidence on NDRT engagement should be considered preliminary, we believe it holds value for understanding human factors in L2 driving.

*Additional Considerations on NDRT.* Based on the studies reviewed, drivers with more experience with partially automated systems appear to engage NDRT more often during partial automation. Naujoks et al. (2016) and Solis-Marcos et al. (2018) explored the influence of the experience with partially automated systems on secondary tasks engagement, finding that only experienced drivers engaged more in NDRT during partially automated driving compared to manual driving. Although only two studies among those we selected tested this, it is noteworthy that their findings align with Dunn and colleagues' framework (Dunn et al., 2021), suggesting that the increased engagement in nondriving tasks can be due to experienced drivers over-trusting automated systems. This could be concerning, as increased NDRT engagement may hinder the drivers' ability to anticipate hazards (cf. Hungund & Kumar Pradhan, 2023). Finally, only one study (Cooper et al., 2023) compared a naturalistic condition (in which drivers could choose whether to activate or not in partial automation) with conditions where drivers were instructed to operate the vehicle in either manual or L2 mode, finding lower engagement in NDRT during naturalistic driving. While these findings come from a single study, they suggest that more

realistic driving conditions should be considered when investigating partial automation.

## Conclusions

This work reports the findings of meta-analyses evaluating differences in mental workload, visual behavior, and NDRT engagement between partially automated and manual driving. Below, we summarize the main findings and outline some final conclusions along with potential limitations.

Our data show partially automated driving to increase the likelihood of drivers looking away from the forward roadway and engaging in NDRT. In contrast, no significant differences were observed in mental workload between the two driving modes. Combined, these findings align with the literature on self-regulation of driving behavior, which suggests that drivers usually modulate their behavior in an attempt to avoid conditions of either high and low workload (Moore & Brown, 2019; Oviedo-trespalacios et al., 2017; Oviedo-Trespalacios et al., 2018; Strayer et al., 2017). We argue that the tendency to execute more off-road glances and engage in potentially distracting activities during L2 driving may be the result of drivers trying to counter the onset of boredom resulting from supervising the L2 system, a hypothesis that would find alignment in the literature on vigilance decrement (Molloy & Parasuraman, 1996).

We believe that interventions aimed at correcting drivers' visual behavior and reducing NDRT engagement should be prioritized when developing partially automated systems. Following Hungund and Kumar Pradhan (2023), providing drivers with training could be beneficial and could reduce the risk of distractions, as nontrained drivers might not fully understand the limitations of automated systems. Developing guidelines on the importance of maintaining attention to driving relevant elements when using automated systems might also reduce disengagement from driving and, consequently, improve safety. Similarly, we believe consideration should also be given to the design of partially automated systems. In a comprehensive review, Wang et al. (2024) provide guidelines for interface design in driving automation systems, suggesting that auditory alerts are preferable to visual ones. This aligns with the

concerns outlined in our meta-analysis, as visual alerts could divert gaze from the road, potentially encouraging risky visual behavior (cf. NHTSA, 2013). Another interesting finding that emerges from this work is the importance of previous experience with automation. On the one hand, drivers inexperienced with automation tend to report higher workload, likely due to a lack of trust in these systems. On the other hand, drivers experienced with automation are more likely to engage in secondary tasks while driving, possibly due to over-trust in these systems (cf. Dunn et al., 2021). We believe that developing training programs that balance trust in automation might reduce potential risks associated with workload and distractions. Again, future research should leverage these findings to create appropriate training programs for drivers interested in partially automated systems.

Our review also presents some limitations. First, the study selection process only included studies published in peer-reviewed journals or conferences, thus excluding works from non-peer-reviewed outlets. This decision aimed to improve the quality of the analyzed studies, as non-peer-reviewed papers may contain unreliable findings. Second, some findings were omitted due to unclear reporting of results and/or statistics (see Figure 1). This exclusion may have impacted the meta-analysis results, as some data were excluded solely based on reporting issues. To address this, a publication bias assessment was conducted to estimate the likelihood of omitted nonsignificant findings. Third, although a meta-analysis on NDRT engagement was conducted, only a few studies examined this in naturalistic on-road settings, making the conclusions of this analysis less robust compared to those on mental workload and visual behavior. Fourth, only half of the studies included in this work used realistic L2 systems (i.e., on-road studies), while the rest employed driving simulators. Although moderator analysis on mental workload and visual behavior revealed no differences between on-road and simulated studies, it is important to note that our conclusions are also based on studies using systems that may not fully resemble real-world L2 systems. As a result, our findings may not generalize to all types of L2 systems. Moreover, our work does not focus on differences in L2 system designs, as this review is focused specifically on the differences between manual and partially automated

systems. However, when available, information on system design is provided in the Supplementary Material. Finally, only five databases were used during the study selection. However, it is important to note that, in line with the guidelines of Paré et al. (2015), this was done to improve replicability and rigor.

Our findings suggest that while partial automation seems to impose a similar mental workload as manual driving, it may lead to increased visual behaviors toward nondriving relevant elements and slightly greater engagement in secondary tasks. This behavior is concerning as it may impair drivers' ability to react quickly to sudden hazards. Future research on partially automated systems should focus on interventions aimed at improving visual behavior and reducing engagement in secondary tasks.

## Key Points

- No significant differences in mental workload between manual and partially automated driving.
- Increased visual behavior toward non-driving related areas during partial automation.
- Greater engagement in non-driving related tasks during partial automation.
- Previous experience with automation can affect the safety of partially automated systems.

## Acknowledgments

The authors acknowledge the generous contribution from the University of Windsor Research Chair program. They also thank the Natural Science and Engineering Research Council and the Social Science and Humanities Research Council of Canada for their support.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Francesco Biondi  <https://orcid.org/0000-0002-5558-4707>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*(2), 527–535. <https://doi.org/10.3758/s13428-012-0265-2>
- Banks, V. A., & Stanton, N. A. (2016). Keep the driver in control: Automating automobiles of the future. *Applied Ergonomics*, *53*(pt.B), 389–395. <https://doi.org/10.1016/j.apergo.2015.06.020>
- Banks, V. A., & Stanton, N. A. (2019). Analysis of driver roles: Modelling the changing role of the driver in automated driving systems using EAST. *Theoretical Issues in Ergonomics Science*, *20*(3), 284–300. <https://doi.org/10.1080/1463922x.2017.1305465>
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. <https://doi.org/10.2307/2533446>
- Berkey, C. S., Anderson, J. J., & Hoaglin, D. C. (1996). Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine*, *15*(5), 537–557. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960315\)15:5<537::AID-SIM176>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1097-0258(19960315)15:5<537::AID-SIM176>3.0.CO;2-S)
- Biondi, F., Alvarez, I., Jeong, K., Biondi, F., Alvarez, I., & Jeong, K. (2019). Human - system cooperation in automated driving. *International Journal of Human-Computer Interaction*, *35*(11), 917–918. <https://doi.org/10.1080/10447318.2018.1561793>
- Biondi, F. N. (2024). Adopting stimulus detection tasks for cognitive workload assessment: Some considerations. *Human Factors*, *66*(12), 2561–2568. <https://doi.org/10.1177/00187208241228049>
- Biondi, F. N., & Jajo, N. (2024). On the impact of on-road partially-automated driving on drivers' cognitive workload and attention allocation. *Accident Analysis & Prevention*, *200*(March), 107537. <https://doi.org/10.1016/j.aap.2024.107537>
- Biondi, F. N., Lohani, M., Hopman, R., Mills, S., Cooper, J. M., & Strayer, D. L. (2018). 80 MPH and out-of-the-loop : Effects of real-world semi-automated driving on

- driver workload and arousal. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting*, 62(1), 1878–1882. <https://doi.org/10.1177/1541931218621427>
- Biondi, F. N., McDonnell, A. S., Mahmoodzadeh, M., Jajo, N., Balasingam, B., & Strayer, D. L. (2023). Vigilance decrement during on-road partially automated driving across four systems. *Human Factors*, 66(9), 2179–2190. <https://doi.org/10.1177/00187208231189658>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.
- Butmee, T., Lansdown, T., & Walker, G. (2019). Mental workload and performance measurements in driving task: A review literature. *Advances in Intelligent Systems and Computing*. [https://doi.org/10.1007/978-3-319-96074-6\\_31](https://doi.org/10.1007/978-3-319-96074-6_31)
- Cabrall, C. D. D., Eriksson, A., Dreger, F., Happee, R., & Winter, J. De. (2019). Theoretical issues in ergonomics science how to keep drivers engaged while supervising driving automation? A literature survey and categorisation of six solution areas. *Theoretical Issues in Ergonomics Science*, 0(0), 1–38. <https://doi.org/10.1080/1463922X.2018.1528484>
- Cantin, V., Lavallière, M., Simoneau, M., & Teasdale, N. (2009). Mental workload when driving in a simulator: Effects of age and driving complexity. *Accident Analysis & Prevention*, 41(4), 763–771. <https://doi.org/10.1016/j.aap.2009.03.019>
- Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semi-automated driving: Does it matter what aspects are automated? *Human Factors*, 54(5), 747–761. <https://doi.org/10.1177/0018720812460246>
- Casner, S. M., & Hutchins, E. L. (2019). What do we tell the drivers? Toward minimum driver training standards for partially automated cars. *Journal of Cognitive Engineering and Decision Making*, 13(2), 55–66. <https://doi.org/10.1177/1555343419830901>
- Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, 74(2018), 221–232. <https://doi.org/10.1016/j.apergo.2018.08.028>
- Cochran, W. (1954). The Combination of Estimates from Different Experiments. *Biometrics*. <https://www.jstor.org/stable/3001666>
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. New York, USA: Routledge.
- Cooper, J. M., Crabtree, K. W., McDonnell, A. S., May, D., Strayer, S. C., Tsogbaatar, T., Cook, D. R., Alexander, P. A., Sanbonmatsu, D. M., & Strayer, D. L. (2023). Driver behavior while using level 2 vehicle automation: A hybrid naturalistic study. *Cognitive Research: Principles and Implications*, 8(1), 71. <https://doi.org/10.1186/s41235-023-00527-5>
- Damböck, D., Weißgerber, T., Kienle, M., & Bengler, K. (2013). Requirements for cooperative vehicle guidance. In *16th international IEEE conference on intelligent transportation systems (ITSC 2013)*, iisc (pp. 1656–1661). Available at: <https://doi.org/10.1109/ITSC.2013.6728467>
- de Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, 27(1), 196–217. <https://doi.org/10.1016/j.trf.2014.06.016>
- Dogan, E., Honnêt, V., Masfrand, S., & Guillaume, A. (2019). Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 62(1), 494–504. <https://doi.org/10.1016/j.trf.2019.02.010>
- Dunn, N. J., Dingus, T. A., Soccolich, S., & Horrey, W. J. (2021). Investigating the impact of driving automation systems on distracted driving behaviors. *Accident Analysis & Prevention*, 156(December 2020), 106152. <https://doi.org/10.1016/j.aap.2021.106152>
- Ebnali, M., Hulme, K., Ebnali-Heidari, A., & Mazloumi, A. (2019). How does training effect users' attitudes and skills needed for highly automated driving? *Transportation Research Part F: Traffic Psychology and Behaviour*, 66(1), 184–195. <https://doi.org/10.1016/j.trf.2019.09.001>
- Eddine, R. J., Mulatti, C., & Biondi, F. N. (2024). On investigating drivers' attention allocation during partially-automated driving. *Cognitive Research: Principles and Implications*, 9(1), 21. <https://doi.org/10.1186/s41235-024-00549-7>
- Engström, J., Tech, V., Kingdom, U., & Victor, T. (2013). Effects of cognitive load on driving performance: The cognitive control hypothesis. <https://doi.org/10.1177/0018720817690639>

- Feldhütter, A., Härtwig, N., Kurpiers, C., Hernandez, J., & Bengler, K. (2019). *Effect on mode awareness when changing from conditionally to partially automated driving: Volume VI: Transport ergonomics and human factors (TEHF)* (pp. 314–324). Aerospace Human Factors and Ergonomics. Available at: <https://doi.org/10.1007/978-3-319-96074-6>
- Figalová, N., Bieg, H. J., Reiser, J. E., Liu, Y. C., Baumann, M., Chuang, L., & Pollatos, O. (2024). From driver to supervisor: Comparing cognitive load and EEG-based attentional resource allocation across automation levels. *International Journal of Human-Computer Studies*, 182(September 2023), 103169. <https://doi.org/10.1016/j.ijhcs.2023.103169>
- Fillmore, M. T., Blackburn, J. S., & Harrison, E. L. R. (2008). Acute disinhibiting effects of alcohol as a factor in risky driving behavior. *Drug and Alcohol Dependence*, 95(1–2), 97–106. <https://doi.org/10.1016/j.drugalcdep.2007.12.018>
- Gaspar, J., & Carney, C. (2019). The effect of partial automation on driver attention: A naturalistic driving study. *Human Factors*, 61(8), 1261–1276. <https://doi.org/10.1177/0018720819836310>
- Gibbons, J. D. (1993) *Nonparametric statistics: An introduction* (9). Sage.
- Goncalves, R. C., Louw, T. L., Quaresma, M., Madigan, R., & Merat, N. (2020). The effect of motor control requirements on drivers' eye-gaze pattern during automated driving. *Accident Analysis & Prevention*, 148(August), 105788. <https://doi.org/10.1016/j.aap.2020.105788>
- Greenlee, E. T., DeLucia, P. R., & Newton, D. C. (2022). Driver vigilance decrement is more severe during automated driving than manual driving. *Human Factors*, 66(2), 574–588. <https://doi.org/10.1177/00187208221103922>
- Gruden, T., Tomažič, S., Sodnik, J., & Jakus, G. (2022). A user study of directional tactile and auditory user interfaces for take-over requests in conditionally automated vehicles. *Accident Analysis & Prevention*, 174(2021), 106766–106811. <https://doi.org/10.1016/j.aap.2022.106766>
- Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hasanat-E-rabbi, S., Hamim, O. F., Debnath, M., Hoque, M. S., McIlroy, R. C., Plant, K. L., & Stanton, N. A. (2021). Exploring the relationships between demographics, road safety attitudes, and self-reported pedestrian behaviours in Bangladesh. *Sustainability*, 13(19), 10640. <https://doi.org/10.3390/su131910640>
- Hatfield, N., Yamani, Y., Palmer, D. B., Yaloodik, S., Vasquez, V., Horrey, W. J., & Samuel, S. (2019). Analysis of visual scanning patterns comparing drivers of simulated L2 and L0 systems. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(10), 755–761. <https://doi.org/10.1177/0361198119852339>
- He, D., & Donmez, B. (2019). Influence of driving experience on distraction engagement in automated vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(9), 142–151. <https://doi.org/10.1177/0361198119843476>
- He, D., Kanaan, D., & Donmez, B. (2022). Distracted when using driving automation: A quantile regression analysis of driver glances considering the effects of road alignment and driving experience. *Frontiers in Future Transportation*, 3(May), 1–10. <https://doi.org/10.3389/ffutr.2022.772910>
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92(2), 490–499. <https://doi.org/10.1037/0033-2909.92.2.490>
- Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, 93(2), 388–395. <https://doi.org/10.1037/0033-2909.93.2.388>
- Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic Press.
- Hungund, A. P., & Kumar Pradhan, A. (2023). Impact of non-driving related tasks while operating automated driving systems (ADS): A systematic review. *Accident Analysis & Prevention*, 188(April), 107076. <https://doi.org/10.1016/j.aap.2023.107076>
- ISO. (2015). *Detection-response task (DRT) for assessing attentional effects of cognitive load in driving*. ISO/DIS, 17488.
- Kidd, D. G., Dobres, J., Reagan, I., Mehler, B., & Reimer, B. (2017). Considering visual-manual tasks performed during highway driving in the context of two different sets of guidelines for embedded in-vehicle electronic systems. *Transportation Research Part F: Traffic Psychology and Behaviour*, 47(1), 23–33. <https://doi.org/10.1016/j.trf.2017.04.002>
- Kim, J., Revell, K., Langdon, P., Bradley, M., Politis, I., Thompson, S., Skrypchuk, L., O'Donoghue, J., Richardson, J., & Stanton, N. A. (2023). Partially automated driving has higher workload than manual driving: An on-road comparison of three

- contemporary vehicles with SAE Level 2 features. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 33(1), 40–54. <https://doi.org/10.1002/hfm.20969>
- Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three level meta analysis. *Research Synthesis Methods*, 2(1), 61–76. <https://doi.org/10.1002/jrsm.35>
- Kraft, A. K., Naujoks, F., Wörle, J., & Neukum, A. (2018). The impact of an in-vehicle display on glance distribution in partially automated driving in an on-road experiment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 52(1), 40–50. <https://doi.org/10.1016/j.trf.2017.11.012>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science : A practical primer for t -tests and ANOVAs. *Frontiers in Psychology*, 4(November), 863–912. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lohani, M., Cooper, J. M., Erickson, G. G., Simmons, T. G., McDonnell, A. S., Carriero, A. E., Crabtree, K. W., & Strayer, D. L. (2021). No difference in arousal or cognitive demands between manual and partially automated driving: A multi-method on-road study. *Frontiers in Neuroscience*, 15(June), 577418–577512. <https://doi.org/10.3389/fnins.2021.577418>
- Longo, L., Wickens, C. D., Hancock, G., & Hancock, P. A. (2022). Human mental workload: A survey and a novel inclusive definition. *Frontiers in Psychology*, 13(June), 883321–883326. <https://doi.org/10.3389/fpsyg.2022.883321>
- Louw, T., & Merat, N. (2017). Are you in the loop ? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies*, 76(1), 35–50. <https://doi.org/10.1016/j.trc.2017.01.001>
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. The MIT Press.
- Mahajan, K., Large, D. R., Burnett, G., & Velaga, N. R. (2021). Exploring the benefits of conversing with a digital voice assistant during automated driving : A parametric duration model of takeover time. *Transportation Research Part F: Traffic Psychology and Behaviour*, 80(1), 104–126. <https://doi.org/10.1016/j.trf.2021.03.012>
- Matthews, G., De Winter, J., & Hancock, P. A. (2020). What do subjective workload scales really measure? Operational and representational solutions to divergence of workload measures. *Theoretical Issues in Ergonomics Science*, 21(4), 369–396. <https://doi.org/10.1080/1463922x.2018.1547459>
- McDonnell, A. S., Crabtree, K. W., & City, S. L. (2023). *This is your brain on autopilot 2.0: The influence of practice on driver workload and engagement during on-road, partially automated driving amy*. Human Factors. Available at: <https://doi.org/10.1177/00187208231201054>
- McDonnell, A. S., Simmons, T. G., Erickson, G. G., Lohani, M., Cooper, J. M., & Strayer, D. L. (2021). This is your brain on autopilot: Neural indices of driver workload and engagement during partial vehicle automation. *Human Factors*, 65(7), 1435–1450. <https://doi.org/10.1177/00187208211039091>
- McWilliams, T., & Ward, N. (2021). Underload on the road: Measuring vigilance decrements during partially automated driving. *Frontiers in Psychology*, 12(1), 631364. <https://doi.org/10.3389/fpsyg.2021.631364>
- Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Katzaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., Keinath, A., Green, C. A., & Keinath, A. (2019). The “out-of-the-loop” concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology & Work*, 21(1), 87–98. <https://doi.org/10.1007/s10111-018-0525-8>
- Miller, E. E., & Boyle, L. N. (2019). Adaptations in attention allocation: Implications for takeover in an automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, 66(1), 101–110. <https://doi.org/10.1016/j.trf.2019.08.016>
- Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38(2), 311–322. <https://doi.org/10.1177/001872089606380211>
- Moore, M. M., & Brown, P. M. (2019). The association of self-regulation, habit, and mindfulness with texting while driving. *Accident Analysis & Prevention*, 123(May 2018), 20–28. <https://doi.org/10.1016/j.aap.2018.10.013>
- Morris, S. B., DeShon, R. P., Borenstein, M., Hedges, L. V., Higgins, J. P. T., Rothstein, H. R., Winter, J. C. F., Happee, R., Martens, M. H., Stanton, N. A., Gibbons, R. D., Davis, J. M., Hungund, A. P., Pradhan, A. K., Schmalz, X., Altoè, G., & Mulatti, C. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-

- groups designs. *Psychological Methods*, 7(1), 105–125. <https://doi.org/10.1037/1082-989X.7.1.105>
- Mueller, A. S., Cicchino, J. B., Benedick, A., De Leonardis, D., & Huey, R. (2022). Bears in our midst: Familiarity with Level 2 driving automation and attending to surprise on-road events. *Transportation Research Part F: Traffic Psychology and Behaviour*, 90(February), 500–511. <https://doi.org/10.1016/j.trf.2022.09.016>
- National Highway Traffic Safety Administration. (2013). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices*. NHTSA.
- Naujoks, F., Purucker, C., & Neukum, A. (2016). Secondary task engagement and vehicle automation - comparing the effects of different automation levels in an on-road experiment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 38(1), 67–82. <https://doi.org/10.1016/j.trf.2016.01.011>
- NHTSA. (2022). *Summary report: Standing general order on crash reporting for level 2 advanced driver assistance systems*. NHTSA.
- Noble, A. M., Miles, M., Perez, M. A., Guo, F., & Klauer, S. G. (2021). Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. *Accident Analysis & Prevention*, 151(March 2020), 105959. <https://doi.org/10.1016/j.aap.2020.105959>
- NTSB. (2020b). *Highway accident brief: Collision between car operating with partial driving automation and truck-tractor semitrailer*. Delray Beach, Florida. Available at: <https://www.nts.gov/investigations/AccidentReports/Reports/HAB2001.pdf#0Ahttps://trid.trb.org/view/1707891>
- Olkin, I., & Gleser, L. (2009). Stochastically dependent effect sizes. *The handbook of research synthesis and meta-analysis*, 2(1), 357–376.
- Oviedo-trespalacios, O., Haque, M., King, M., Washington, S., & King, M. (2017). Self-regulation of driving speed among distracted drivers: An application of driver behavioral adaptation theory. *Traffic Injury Prevention*, 18(6), 599–605. <https://doi.org/10.1080/15389588.2017.1278628>
- Oviedo-Trespalacios, O., Haque, M. M., King, M., & Demmel, S. (2018). Driving behaviour while self-regulating mobile phone interactions: A human-machine system approach. *Accident Analysis & Prevention*, 118(January), 253–262. <https://doi.org/10.1016/j.aap.2018.03.020>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372(1), n71. <https://doi.org/10.1136/bmj.n71>
- Paré, G., Trudel, M. C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, 52(2), 183–199. <https://doi.org/10.1016/j.im.2014.08.008>
- Pournami, S., Large, D. R., Burnett, G., & Harvey, C. (2015). Comparing the NHTSA and ISO occlusion test protocols: How many participants are sufficient? In *Adjunct proceedings of the 7th international conference on automotive user interfaces and interactive Vehicular Applications* (pp. 110–116). Available at: <https://doi.org/10.1145/2799250.2799255>
- Radhakrishnan, V., Louw, T., Cirino, R., Torrao, G., Lenn, M. G., & Merat, N. (2023). Using pupillometry and gaze-based metrics for understanding drivers' mental workload during automated driving. *Transportation Research Part F: Psychology and Behaviour*, 94(April 2022), 254–267. <https://doi.org/10.1016/j.trf.2023.02.015>
- Radhakrishnan, V., Merat, N., Louw, T., Gonçalves, R. C., Torrao, G., Lyu, W., Puente Guillen, P., & Lenné, M. G. (2022). Physiological indicators of driver workload during car-following scenarios and takeovers in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 87(June 2021), 149–163. <https://doi.org/10.1016/j.trf.2022.04.002>
- Raudenbush, S. (2009). Analyzing Effect Sizes: Random-Effects Models. In *The Handbook Research Synthesis and Meta-Analysis*. New York, USA: Russell Sage Foundation.
- Reagan, I. J., Teoh, E. R., Cicchino, J. B., Gershon, P., Reimer, B., Mehler, B., & Seppelt, B. (2021). Disengagement from driving when using automation during a 4-week field trial. *Transportation Research Part F: Traffic Psychology and Behaviour*, 82(-August 2020), 400–411. <https://doi.org/10.1016/j.trf.2021.09.010>
- Rosenthal, R., & Dimatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for

- literature reviews. *Annual Review of Psychology*, 52(1), 59–82. <https://doi.org/10.1146/annurev.psych.52.1.59>
- Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, 53(1), 61–86.
- SAE (2021). Taxonomy and Definitions for Terms Related to Cooperative Driving Automation for On-Road Motor Vehicles. *SAE International*. [https://www.sae.org/standards/content/j3216\\_202005](https://www.sae.org/standards/content/j3216_202005)
- Safi, S., Grillo, R., De Lauri, A., Billaud, J., & Roszko, E. (2014). REVIEW of social studies special issue methodological choices and challenges. *Review of Social Studies*, 1(1), 6.
- Saikia, M. J., Besio, W. G., & Mankodiya, K. (2021). The validation of a portable functional nirs system for assessing mental workload. *Sensors*, 21(11), 3810. <https://doi.org/10.3390/s21113810>
- Samuel, S., Yamani, Y., & Fisher, D. L. (2020). Understanding drivers' latent hazard anticipation in partially automated vehicle systems. *International Journal of Human Factors and Ergonomics*, 7(3), 282–296. <https://doi.org/10.1504/IJHFE.2020.110093>
- Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied*, 19(4), 287–300. <https://doi.org/10.1037/a0034386>
- Shahini, F., & Zahabi, M. (2022). Effects of levels of automation and non-driving related tasks on driver performance and workload : A review of literature and meta-analysis. *Applied Ergonomics*, 104(July 2021), 103824. <https://doi.org/10.1016/j.apergo.2022.103824>
- Sibi, S., Baiters, S., Mok, B., Steiner, M., & Ju, W. (2017). Assessing driver cortical activity under varying levels of automation with functional near infrared spectroscopy. *2017 IEEE Intelligent Vehicles Symposium (IV)*, 1(1), 1509–1516. <https://doi.org/10.1109/IVS.2017.7995923>
- Solis-Marcos, I., Ahlström, C., & Kircher, K. (2018). Performance of an additional task during level 2 automated driving: An on-road study comparing drivers with and without experience with partial automation. *Human Factors*, 60(6), 778–792. <https://doi.org/10.1177/0018720818773636>
- Solis-Marcos, I., Galvao-Carmona, A., & Kircher, K. (2017). Reduced attention allocation during short periods of partially automated driving: An event-related potentials study. *Frontiers in Human Neuroscience*, 11(November), 537–613. <https://doi.org/10.3389/fnhum.2017.00537>
- Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2019). Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60(1), 590–605. <https://doi.org/10.1016/j.trf.2018.11.006>
- Statista. (2023). Level 2–4 autonomous vehicle sales as a share of total vehicle sales in 2025 and 2030, by automation level. <https://www.statista.com/statistics/1230101>
- Sterne, J. A., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., Cates, C. J., ... Higgins, J. P. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *Bmj*, 366(1), 14898. <https://doi.org/10.1136/bmj.14898>
- Strayer, D. L., Getty, D., Biondi, F., Cooper, J. M., Strayer, L., & Joel, M. (2017). The multitasking motorist the multitasking motorist and the attention economy david. In *Douglas Getty, Francesco Biondi*. Cooper University of Utah.
- Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A systematic review of physiological measures of mental workload. *International Journal of Environmental Research and Public Health*, 16(15), 2716–2723. <https://doi.org/10.3390/ijerph16152716>
- Tsai, Y., Viire, E., Strychacz, C., Chase, B., & Jung, T. (2007). Task performance and eye activity : Predicting behavior relating to cognitive workload. 78(5 Suppl).B176–B185.
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3), 261–293. <https://doi.org/10.3102/10769986030003261>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Waard, D. (1993). *Traffic research*.
- Wang, M., Mehrotra, S., Wong, N., Roberts, S. C., Kim, W., Romo, A., & Horrey, W. J. (2024). Human-machine interfaces and vehicle automation: A review of the literature and recommendations for

- system design, Feedback, and alerts. *Transportation Research Part F: Traffic Psychology and Behaviour*, 107(1), 549–561.
- Weaver, B., & DeLucia, P. (2018). A Systematic Review and Meta- Analysis of Takeover Performance During Conditionally Automated Driving. *Human Factors*. <https://doi.org/10.1177/0018720820976476>
- Wilson, K. M., Yang, S., Roady, T., Kuo, J., & Lenné, M. G. (2020). Driver trust & mode confusion in an on-road study of level-2 automated vehicle technology. *Safety Science*, 130(January), 104845. <https://doi.org/10.1016/j.ssci.2020.104845>
- World Health Organization (WHO). (2023). *Road traffic injuries*. WHO.
- Xu, C., Louw, T. L., Merat, N., Li, P., Hu, M., & Li, Y. (2023). Drivers' gaze patterns when resuming control with a head-up-display: Effects of automation level and time budget. *Accident Analysis & Prevention*, 180(July 2021), 106905. <https://doi.org/10.1016/j.aap.2022.106905>
- Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, 58(1), 1–17. <https://doi.org/10.1080/00140139.2014.956151>
- Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 64(1), 285–307. <https://doi.org/10.1016/j.trf.2019.04.020>
- Zhang, Y., Ma, J., Zhang, C., & Chang, R. (2021). Electrophysiological frequency domain analysis of driver passive fatigue under automated driving conditions. *Scientific Reports*, 11(1), 20348–20349. <https://doi.org/10.1038/s41598-021-99680-4>
- Zhao, R., Liu, Y., Li, T., & Li, Y. (2022). A preliminary evaluation of driver's workload in partially automated vehicles. In *International conference on human-computer interaction* (pp. 448–458). Springer International Publishing.

### Author Biographies

Nicola Vasta is a postdoctoral fellow in the Department of Kinesiology at the University of Windsor. He received his BS and MS from the University of Padova, and his doctorate from the University of Trento.

Francesco N. Biondi is an associate professor in the Department of Kinesiology, Director of the Human Systems Lab, and Research Chair in Human Factors at the University of Windsor. He received his master's in 2011 and doctorate in psychological science in 2015 from the University of Padova.