Ergonomics & Human factors
fade of a discipline

de Winter, J. C.F.; Eisma, Y. B.

DOI
10.1080/00140139.2024.2416553

Publication date
2024

Document Version
Final published version

Published in
Ergonomics

Ergonomics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/terg20

Ergonomics & Human factors: fade of a discipline

J.C.F. de Winter & Y.B. Eisma

To cite this article: J.C.F. de Winter & Y.B. Eisma (23 Oct 2024): Ergonomics & Human factors: fade of a discipline, Ergonomics, DOI: 10.1080/00140139.2024.2416553

To link to this article: https://doi.org/10.1080/00140139.2024.2416553

Published online: 23 Oct 2024.

ERGONOMICS
https://doi.org/10.1080/00140139.2024.2416553

Taylor & Francis
Taylor & Francis Group

TARGET ARTICLE

Ergonomics & Human factors: fade of a discipline

J.C.F. de Winter and Y.B. Eisma

Cognitive Robotics, Delft University of Technology, Delft, the Netherlands

ABSTRACT
In this commentary, we argue that the field of Ergonomics and Human Factors (EHF) has the tendency to present itself as a thriving and impactful science, while in reality, it is losing credibility. We assert that EHF science (1) has introduced terminology that is internally inconsistent and hardly predictive-valid, (2) has virtually no impact on industrial practice, which operates within frameworks of regulatory compliance and profit generation, (3) repeatedly employs the same approach of conducting lab experiments within unrealistic paradigms in order to complete deliverables, (4) suggests it is a cumulative science, but is neither a leader nor even an adopter of open-science initiatives that are characteristic of scientific progress and (5) is being assimilated by other disciplines as well as Big Tech. Recommendations are provided to reverse this trend, although we also express a certain resignation as our scientific discipline loses significance.

Practitioner Summary: This paper offers criticism of the field of Ergonomics. There are issues such as unclear terminology, unrealistic experiments, insufficient impact and lack of open data. We provide recommendations to reverse the trend. This article concerns a critique of EHF as a science, and is not a critique of EHF practitioners.

ARTICLE HISTORY
Received 30 August 2024
Accepted 9 October 2024

KEYWORDS
Human factors; ergonomics; constructs; impact; realism

Introduction

This paper is a commentary on the relevance of the field of Ergonomics and Human Factors (EHF). The immediate reason for this writing is that we observe an increasing number of articles appearing in EHF journals in which the assumptions and terminology used, as well as the value of experimental designs, appear to be taken for granted without being grounded in real problems and meaningfulness. We observe that EHF as a discipline is being taken less and less seriously and is losing strength.

It is not that EHF theories are being tested and refuted in favour of newer or better theories, leading to progress in knowledge. What is happening is more akin to what was described by Meehl (1978), namely that some disciplines gradually wither away because scientists lose interest as the theories being used offer little real substance. Research on technology (including automation, AI) and the application of this technology is currently taking place in many fields and within companies, but often without relying on EHF literature and theories from the last decades.

There exists an internal tension within EHF where, on the one hand, it proclaims to be a science with a progressing knowledge base, while at the same time, EHF scientists themselves are aware that there is still much work to be done, particularly regarding the impact the discipline has. In a recent article by Salmon et al. (in press), the opinions of 18 EHF scientists and practitioners were surveyed. Among other questions, they were asked: 'On a scale of 1 (not having any impact at all) to 10 (achieving its full desired impact), how would you rate the impact that EHF is currently having on the world?' The responses yielded an average of only 4.95 on the mentioned scale of 1 to 10.

We believe that the 4.95 rating may, in fact, be overly optimistic; after all, the experts were evaluating the impact of their *own* discipline and, in some cases, their *own* research programs. Considering the potential bias towards socially desirable responses, the actual influence could be considerably lower. In any case, we see no reason for 'celebrating a 75th anniversary' (Salmon et al., in press). On the contrary, we argue that the EHF field is losing value and credibility.

In line with Meehl's (1978) observations on how research theories can fade away, we have identified five critical pain points that highlight the dissipation of our discipline. Our goal is to present these issues clearly while leaving space for debate. Therefore, instead of

CONTACT J.C.F de Winter j.c.f.dewinter@tudelft.nl Cognitive Robotics, Delft University of Technology, Delft, the Netherlands
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

offering an exhaustive defence, we will support these points with a relatively limited number of references.

We conclude this commentary with several proposed solutions and suggestions. These recommendations are based on the assumption that EHF as a scientific discipline is still sufficiently flexible and adaptive to recover itself. However, the decline of the EHF field might also need to be accepted; such an outcome could even be warranted.

Problem 1: EHF science uses invalid constructs and theories

The field of EHF is characterised by the use of a variety of terminology, such as human error, complacency, workload and situation awareness (SA). The latter two in particular are regarded as the flagship constructs of EHF (eg Vidulich and Tsang 2012). However, these constructs lack internal consistency and offer little predictive value (for earlier criticism of the uncritical use of EHF terminology, see Dekker and Hollnagel 2004; Liu, in press).

For example, seminal EHF works have defined workload in terms of the usage of *mental* resources and as a *human-centred* construct (eg Hart and Staveland 1988). Various textbooks and research articles have further clarified that workload is related to, but *not* the same as, 'task demands' (ie the objective definition of task difficulty), 'effort' (ie how many of one's mental resources are applied/how hard one tries), task performance (ie how well the task is executed), or physical work (eg Yeh and Wickens 1988).

However, the NASA Task Load Index (TLX), the most widely used tool for measuring workload (De Winter 2014; Grier 2015), specifically polls the operator's perceived 'demands' (three of the six items), effort (one item), performance (one item) and physical experience (one item) (see also Matthews, De Winter, and Hancock 2020). This points to a major discrepancy between the theoretical definition of workload and the empirical tools designed to assess it. It appears that researchers select measurement tools that have the appearance of validity (eg due to the name 'NASA') or because of their popularity (ie the Matthew effect), as previously argued by De Winter (2014). There is also little convergence among methods used to measure workload. Various approaches such as questionnaires, reaction times, visual search tasks, or physiological indices have been employed to measure workload, but the scores on these measures correlate only weakly with each other (eg Matthews et al. 2015).

The same kind of argument applies to SA. Measures of SA tests are diverse, ranging from self-reports (eg SART; Taylor 1990) to knowledge-based tests (eg SAGAT; Endsley 1988), but regardless of the measurement method, SA scores correlate only moderately with task performance once selective reporting bias is accounted for (see Bakdash et al. 2020, 2022).

The issue here does not lie with the measurements themselves. There is nothing inherently wrong with asking how demanding someone finds a task, or figuring out whether the operator understands the task environment. The problem lies in the fact that these measurements have been elevated into themes/constructs, giving the pretence of the construct representing some psychological entity and being something more theoretically coherent and predictive-valid than it actually is.

Over the past decade, several 'State of Science' articles have been published in the journal *Ergonomics* by prominent figures in the field, such as a State of Science article on the topic of workload (Young et al. 2015). However, this review is not particularly consistent within itself and even concludes: 'MWL (mental workload) looks to be just as nebulous a concept today as it did three decades ago, and researchers continue to debate over definitions to this date' (p. 11). Apparently, it has not yet occurred to the authors that the topic of workload (and other EHF constructs) might be an emperor with no clothes and could be detrimental to the credibility of the field.

Problem 2: research in human-automation interaction has little to no impact on practice

In the development of products and services, EHF methods such as focus groups, interviews, or usability studies are often used (Proctor and Van Zandt 2017). Handbooks from the 20th century contain guidelines on how interfaces should be designed in terms of colour, light intensity, movement, etc. (eg Boff, Kaufman, and Thomas 1986). These methods and guidelines have proven their worth, that much is true.

Worldwide, thousands of EHF practitioners, designers, and engineers are working to develop better interfaces, products and services. However, there appears to be a large gap between EHF research and practice, as the research findings published by EHF scientists over the past decades seem to have no influence on the design of these interfaces, products and services. In other words, research from the past few decades has had little impact on practice. The best example of EHF's failure in this regard concerns human-automation interaction research. For decades, EHF scientists have been proclaiming that human operators are not well-suited to be supervisors of automation and that loss of situation awareness (see Problem 1), complacency and loss of vigilance are inevitable consequences (eg Bainbridge 1983; Kyriakidis et al. 2019).

A cynical perspective on this is as follows: technology advances indifferently, and this can be perpetually leveraged by EHF scientists under the following narrative: 'New automation technology (eg automated cars, AI, etc.) is being introduced. But beware: decades of EHF studies show that technology has dangerous side effects. Therefore, we are applying for this research grant'.

In reality, automation developers largely ignore the numerous research articles from the EHF domain. For example, despite years of criticism from the scientific EHF community, Tesla has introduced Full Self Driving (FSD), where in the latest software versions, the driver barely needs to intervene and does not have to touch the steering wheel anymore (eg FSD Community Tracker 2024; WholeMars [Whole Mars Catalog] 2024a, 2024b). Although the use of automated driving technology may involve certain risks (which are very often emphasised in the EHF literature), the doomsday predictions of EHF seems to hold little validity. Direct observations and interviews with drivers have revealed some effects that are contrary to those reported in the mainstream EHF literature. For example, Nordhoff et al. (2023) found some positive effects of Tesla's autopilot, such as increased attentiveness ("increased situational awareness") and reduced fatigue, while Fridman et al. (2019) reported that drivers remain "functionally vigilant", meaning they are effectively responding to challenging driving situations. Moreover, it is worth pointing out that Tesla sales or cars equipped with automated driving systems continue to rise annually, with no indication of a significant increase in fatal accidents.

EHF science essentially capitalises on technological progress. Each time new technology emerges, EHF scientists are quick to latch onto it, warning, with a wagging finger, that this new technology is dangerous. The industry is struggling with an overabundance of regulations (Hubbard and Starger 2024), and it is understandable that when possibilities are limited, and lawsuits or bad press are lurking, there is little interest in listening to EHF scientists who 'complain' that technology is fundamentally dangerous, typically without real-world evidence (see Problem 3).

Problem 3: research designs in EHF serve academic productivity, not meaningfulness

We previously made this statement regarding EHF studies on the topic of automated driving (De Winter, Stanton, and Eisma 2021). In short, our argument was that hundreds of studies have already been conducted in driving simulators where drivers had to take control of a Level 3 automated vehicle that, at some point, issued a take-over request. Drivers then typically have 5 to 7 s to intervene and avoid an accident.

The problem with these studies is that they are designed to produce a research paper and lack an identification of the *real* issues. In reality, there exists no Level 3 automation,[1] while automated vehicles with Level 4 or 5 automation (eg Waymo, Lyft; Hu et al. 2023; Li et al. 2023) have long been available. Simultaneously, some automakers are pursuing the development of Level 2 automation as a strategic path forward (Teslamagazine 2024).

In Level 4 automation, a take-over often involves the vehicle coming to a stop and a remote operator solving the problem. In Level 2 automation, on the other hand, the driver typically needs to take over the steering wheel within a very short timeframe (eg within 1 s). In some cases, the automation system deactivates itself with a warning signal, but in many other cases, the driver must act proactively, ie before the automated driving system realises it is in trouble, also called silent take-overs. The importance of silent take-overs is also evident from disengagement reports of test vehicles (Boggs et al., 2020; Chengula et al. 2023). However, these types of more immediate (silent, proactive) take-overs have hardly been studied, suggesting that researchers do not care about them.

The mechanism by which EHF science operates seems to be as follows:

- An experienced researcher secures funding under the default narrative ('Technology is advancing. EHF research is important') (see Problem 2).
- With the secured budget, the experienced researcher (supervisor) recruits PhD students and manages the project, for a period of typically four years.
- The supervisor realises that it is a challenge to successfully complete the project on time. The supervisor feels responsible for meeting all deliverables and would like the PhD students to produce a thesis with at least four journal articles. Efficient action will be necessary. The funding agency and graduate school assist with this; they enforce formal progress meetings and offer 'how to' courses to ensure matters run smoothly and efficiently.
- The supervisor and co-supervisors help devise a plan of multiple experiments, interview studies, or

questionnaire surveys that are considered feasible within the available time frame and lab facilities. A typical feasible study might involve an EHF experiment in a virtual simulation, where participants (often students) each contribute an hour of their time.
- After about four or five years, the PhD students defend their dissertations and the final report is submitted.

The result of this approach is that it creates an *illusion* of success: The theses are delivered within the nominal time frame; the thesis chapters (articles) are of good quality, ideally published in leading journals (and with less luck, in the form of preprints). The PhD students, the supervisors, the university, and the funding agency are all satisfied.[2]

The downside of this systematised approach to EHF research is that its outcomes remains predictable, ticking off the above steps rather than pursuing creative paths or intellectual depth or touching on real problems. Modern EHF research rarely offers a surprising perspective, dives deeply into a problem, or offers a counter-narrative. Members of the dissertation committee, who will not admit this, often only skim through the dissertation, being trapped in a system that is focused on manufacturing bureaucratic success. The senior researchers rarely publish authoritative literature reviews or well-written books. They do not have time for this, and necessarily spend their time on grant writing, supervision, meetings and administrative tasks.

Problem 4: few EHF data have been deposited in open data repositories

Around 2010, the field of psychology was shaken by replication problems. Due to increasing sample sizes and a data-driven attitude of a new generation of researchers, it became clear that many findings, especially in social psychology, did not hold up. In response, there has been an international shift towards open data and transparency, as seen in the Open Science Framework, GitHub and other initiatives (Christensen et al. 2019; Escamilla et al. 2022; Jansen et al. 2024).

However, EHF has generally not embraced these developments. Only a small portion of the papers published in EHF journals adhere to open science principles (McCarley et al. 2023; as also recognised by Ebel et al. 2024). As a result, there is hardly any measurable 'state of science', where researchers would build on each other's results or code. In short, not only in a theoretical and fundamental sense (see Problem 1), but also in an operational sense (Problem 4), EHF is not a flag bearer or even an adopter of these principles.

EHF is closely related to psychophysics and cognitive psychology, and to a lesser extent to social psychology, which is favourable regarding expected replicability (Scholl 2017; Zwaan et al. 2018). Our impression is that EHF research findings should generally replicate reasonably well. Nonetheless, it is still common for non-robust statistical tests or models (eg multiway interactions in an ANOVA, non-independent sampling), *p*-values that barely pass the 0.05 threshold, or selective presentations of findings to be used, which raises doubts about replicability (for typical pitfalls, see eg De Winter and Dodou 2021).

In summary, the concern we express is that although EHF research is not so fragile as to make claims akin to those in acupuncture or clairvoyance (eg Bem 2011), we should not be proud either. More attention should be given to, and inspiration drawn from, other sciences, such as computer science, where cumulative progress is being made through publicly available software and datasets.

Problem 5: EHF science is being subsumed under other disciplines

The field of EHF has its origins in labs in the USA and the UK (Meister 1999; Sheridan 1986). There has, however, been significant globalisation in recent decades. Connected to this, there is diminishing awareness of each other's research. It is becoming more common for innovative EHF-related articles to emerge from adjacent disciplines and from countries that are increasingly prominent in scientific research, such as China (eg Zhang et al. 2024). In other words, there does not seem to be a singular state of science in EHF, but rather an expansion and growing connection with other disciplines such as robotics, computer science, as well as ethics or business and operational science, disciplines that employ different terminologies and theories. These disciplines are, in a sense, overshadowing the traditional EHF field.

Secondly, the field is increasingly being transformed by AI. However, AI ideation and development are not occurring in Europe. The irony is that students and academics alike write their deliverables using tools like ChatGPT (De Winter, Dodou, and Stienen 2023; Kobak et al. 2024), but the servers for data storage and the computations performed by these innovations are located in the USA. The same is true for the outsourcing of computations or data storage, such as in cloud computing or with tools like Microsoft Teams/Outlook. The power resides within the walls of Big Tech, or more precisely, within their data centres.

Many of the EHF scientists concur that AI is a determining factor for the future (Salmon et al. in press).

One quote from Paul Salmon stands out in particular: 'AI is an obvious area where EHF should be taking centre stage to ensure that technologies are safe, ethical, and beneficial to humanity. Our inability to sufficiently impact the evolution of ANI is arguably our most significant failure as a discipline'. This quote is illustrative for two reasons. The first sentence represents the default narrative (see Problem 2), an attempt to latch onto technological developments. The second sentence directly points out the issue that EHF is unlikely to have any meaningful impact. Research on AI safety, for example, is currently being fought out within Big Tech itself, with plenty of discussion involving ethicists, legislators and lobbyist. EHF may be merely observing from the sidelines, with little real expertise and authority in this area.

In summary, there is no longer a 'state of EHF science' with its centre of gravity in the UK or Europe. If there is any cumulative progress, it is happening elsewhere in the world, often outside the domain of science and within the closed communities of Big Tech.

Solutions and suggestions

Given the observations above, the question is how we, as the EHF community, should move forward to become a more credible scientific discipline characterised by a cumulative state of science. We propose to implement the following recommendations.

1. Know the EHF classics and replicate them. If we aim to move beyond the hollow nature of current EHF theory and experiments, there needs to be commitment to more extensive reading and analysis. EHF researchers should reflect more deeply on the classics. This includes foundational works by figures like Paul M. Fitts and Frederick W. Taylor. Additionally, we recommend conducting replication research of published EHF findings. At first glance, one might assume that replication research is, by definition, not innovative. However, we believe this assumption is misguided. Replication research can provide a critical foundation for strengthening (or critiquing) the knowledge base and thereby pave the way for genuine innovation. Moreover, replication research offers valuable insights into the field and its methodologies, as further discussed by Derksen et al. (2024).
2. Determine the real state of EHF science and ask real questions. In addition to knowing and understanding the classics, we need to thoroughly assess the *true* state-of-the-art in EHF, wherever it exists, and determine how to build upon it. We should not rely anymore on superficial literature reviews, which unfortunately are commonplace, but instead, we must go *deeper*. Individual researchers should commit to thoroughly reading and analysing at least several dozen papers (including the associated computer code) before conducting a new EHF experiment. Additionally, EHF researchers must look beyond what has already been done, like take-over studies, and think outside the box, aiming to identify the *real* problems and needs in the world. One strategy to assess the relevance of a research question or method could be to present it to practitioners or the general public and gauge their reactions (or to let the general public generate the research questions, e.g., Dutch Research Agenda; Dutch Research Council 2024).
3. Improve and justify one's methods and measures. EHF researchers should better justify their assumptions and describe whether their models and measurement methods are valid. This especially means that the uncritical use and adoption of terminology such as 'situation awareness' needs to be prevented in favour of a detailed description of one's measurement methods. We therefore advocate for the adoption of an operationalist philosophy within EHF to prevent constructs from being (mis)used as if they were causal entities. Additionally, to improve reproducibility, EHF scientists should focus on open data practices, with increased sample sizes.

The dilemma is that within current funding programs, it is hardly possible to devote time to extensive reading, analysis, replication, or in-depth exploration. Early stage researchers are required to produce a certain number of papers within four years' time. While the *average* quality (Dechartres et al. 2017) and quantity (Bornmann, Haunschild, and Mutz 2021) of research papers have presumably improved, the diversity of ideas, exceptionalism and impact of EHF science on practice remain low. Added to this, funding agencies, universities, departments and sections demand meetings, get-togethers, deliverables, demonstration events and impose administrative structures that consume considerable amounts of time. It seems we are increasing demands on each other to entertain rather than to encourage one another to study, read and engage deeply with the subject matter. It nowadays requires considerable perseverance and contrarianism to set aside time to study literature or to engage in a substantive, in-depth discussion about eg research hypotheses or data. We have even spoken to several EHF professors who admitted that they are no longer up-to-date with the state-of-the-art and effectively do not read any articles or analyse datasets anymore.

4. Nurture curiosity; refuse bureaucracy. One way out of the above dilemma is to avoid bureaucracy while finding more time for curiosity. Several strategies may exist here:

   a. EHF scientists, when they want, should be able to stand up and refuse to participate in activities that do not stimulate creativity or deep debate, or otherwise contribute to a better science. We must also stand up for core values such as knowledge generation, free debate over research hypotheses and providing honest feedback on ideas. It is quite challenging to find time for this amidst all the top-down imposed procedures and self-initiated events. A change of mindset and of core values may be required.
   b. EHF scientists should better distribute work through team-based science. In a research team, some members could focus on management, others on deeper literature knowledge and others on software development and experiment execution. All these contributors deserve authorship positions according to the CRediT criteria (CRediT 2024; Holcombe 2021). The requirement that a PhD student must always be the first author on a certain number of articles is restrictive in this context. Of course, teams do not need to consist of hundreds of authors, as is sometimes seen (eg in molecular biology; Adams et al. 2021), but a team should be able to form an organic unit with the purpose of conducting cumulative science, not necessarily to 'produce' PhD graduates and project deliverables.
   c. Funding agencies and universities should focus on their core tasks, namely allocating funds, monitoring the legitimacy of expenditures and preventing unsafe or corrupt practices. They should be *less* involved in overseeing or encouraging step-by-step plans, deliverables, utilisation requirements, or progress meetings. These recommendations for a reduced role of funders have been previously discussed (Edwards and Roy 2017; Ioannidis 2011).
   d. There should be room for self-funded external researchers within EHF; individuals with good ideas who wish to join a university unpaid to develop these ideas, thereby conducting their work at the university with *pure* curiosity-driven motives (as they receive no salary and the work is self-initiated).

5. Learn from EHF research that does have an impact. Valuable insights can possibly be gained from the few EHF researchers who do claim they have an influence in practice. For example, Salmon et al. (in press) reported that in certain sectors like rail, EHF is indeed embedded and impactful. A closer examination of this claim suggests that in the UK there is a close collaboration between railway organisations and universities (although it can be debated to what extent this still qualifies as EHF *science* and is not a partnership aimed at directly improving products and services). In summary, although this commentary has provided much criticism of the EHF as a whole, it is possible that in subdomains, EHF science is developed in a meaningful and impactful way. There is potential to learn from this and build upon it.

Naturally, the present paper can be critiqued. We have presented this manuscript to a number of colleagues, and they generally agreed with our assertions. However, they pointed out that some of the claims we make, such as those concerning academic productivity and superficiality, are not unique to EHF but applicable to other scientific disciplines as well. Our response to this is that it is true that the current commentary may also be relevant to other fields. However, EHF is under particular scrutiny because its raison d'être lies in the study of technology. It is therefore especially important that experimental results can be translated back into practice. It is not without reason that the journal *Ergonomics* includes a mandatory 'practitioners summary'[3] and the journal *Human Factors* requires a mandatory 'Application' sentence in the abstract, ie attempts to make clear that/how the work is practically relevant. Another point of criticism we received is that this paper has a somewhat apocalyptic tone, and that with the increasing introduction of technology, such as AI on smartphones, EHF is actually becoming *more* relevant, not less. This may be true, but our point is *not* that humans or 'the human factor' are becoming less important. Our point is that EHF as a research field seems to be losing strength, with other disciplines gradually taking over, disregarding EHF terminology and history in the process.

Instead of turning the tide, a legitimate position is that it does not matter whether EHF weakens or even fades away entirely as a discipline, as science is ultimately a self-organising system. We consider it plausible that EHF will be gradually assimilated into larger and more pragmatic research fields, such as computer science. We see the decline of EHF as a justified outcome. EHF has spent decades talking to itself, clinging to certain terminology or theories and has built too little of a genuine knowledge base, or made sufficient efforts to understand the knowledge base of other disciplines such as machine learning and data science (eg Hannon et al. 2020).

Given the developments in end-to-end neural networks, it is conceivable that understanding the human factor by means of explanatory constructs will become increasingly less important. Neural networks will be able to support humans optimally as a black box, without the need for any underlying EHF theory, model, or construct. An example of this can be seen in the latest generation of automated cars, as well as online platforms like ChatGPT, X (formerly Twitter) and Facebook. These platforms are valued by their users, and the training and fine-tuning of the underlying neural models are carried out on highly pragmatic grounds, ie optimised to ensure maximum user satisfaction or profit, as determined by mouse clicks and preference choices. In this regard, nothing more is needed.

Notes

1. We do not count the Level 3 automation of Mercedes-Benz (2023) and similar Automated Lane Keeping Systems (ALKS) here (UNECE 2021). The operational design domain of these systems is so narrow that it is more of a feature enabling automated driving in traffic jams on selected roads.
2. The supervisors, if possible, may help their students achieve a *cum laude* distinction or another type of award, a concept that further perpetuates the illusion.
3. This also applies to the current paper; in the submission of the manuscript, it was returned to us because we had not included a practitioner summary.

Ethical approval

Not applicable.

Disclosure statement

The views expressed in this article are those of the authors only and do not necessarily reflect the views of any affiliated organization or funding agency.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

J.C.F. de Winter http://orcid.org/0000-0002-1281-8200
Y.B. Eisma http://orcid.org/0000-0003-3437-2761

References

Adams, J., D. Pendlebury, R. Potter, and M. Szomszor. 2021. "Multi-Authorship and Research Analytics." Institute for Scientific Information. Accessed August 27, 2024. https://clarivate.com/wp-content/uploads/2021/02/ISI_-Multiauthorship_Global_Research_Report.pdf

Bainbridge, L. 1983. "Ironies of Automation." *Automatica 19* (6): 775–779. doi:10.1016/0005-1098(83)90046-8.

Bakdash, J.Z., L.R. Marusich, K.R. Cox, M.N. Geuss, E.G. Zaroukian, and K.M. Morris. 2022. "The Validity of Situation Awareness for Performance: A Meta-Analysis." *Theoretical Issues in Ergonomics Science 23* (2): 221–244. doi:10.1080/1463922X.2021.1921310.

Bakdash, J.Z., L.R. Marusich, J.B. Kenworthy, E. Twedt, and E.G. Zaroukian. 2020. "Statistical Significance Filtering Overestimates Effects and Impedes Falsification: A Critique of Endsley (2019)." *Frontiers in Psychology 11*: 609647. doi:10.3389/fpsyg.2020.609647.

Bem, D.J. 2011. "Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect." *Journal of Personality and Social Psychology 100* (3): 407–425. doi:10.1037/a0021524.

Boff, K., L. Kaufman, and J. Thomas, eds. 1986. *Handbook of Perception and Performance*. New York: Wiley.

Boggs, A.M., R. Arvin, and A.J. Khattak. 2020. "Exploring the who, what, when, where, and why of automated vehicle disengagements." *Accident Analysis & Prevention* 136: 105406. doi:10.1016/j.aap.2019.105406.

Bornmann, L., R. Haunschild, and R. Mutz. 2021. "Growth Rates of Modern Science: A Latent Piecewise Growth Curve Approach to Model Publication Numbers from Established and New Literature Databases." *Humanities and Social Sciences Communications 8* (1): 224. doi:10.1057/s41599-021-00903-w.

Chengula, T.J., J. Mwakalonge, G. Comert, J. Perkins, and F. Ngeni. 2023. "The Use of Topic Modeling and Categorical Correlations to Unveil Autonomous Vehicle Disengagement Patterns." *SSRN*. doi:10.2139/ssrn.4666730.

Christensen, G., Z. Wang, E.L. Paluck, N. Swanson, D.J. Birke, E. Miguel, and R. Littman. 2019. "Open Science Practices Are on the Rise: The State of Social Science (3S) Survey." *CEGA Working Paper Series* 106. doi:10.31222/osf.io/5rksu.

CrediT. 2024. "Contributor Role Taxonomy." Accessed August 27, 2024. https://credit.niso.org

Dechartres, A., L. Trinquart, I. Atal, D. Moher, K. Dickersin, I. Boutron, E. Perrodeau, D.G. Altman, and P. Ravaud. 2017. "Evolution of Poor Reporting and Inadequate Methods over Time in 20 920 Randomised Controlled Trials Included in Cochrane Reviews: Research on Research Study." *BMJ (Clinical Research ed.) 357*: j2490. doi:10.1136/bmj.j2490.

Dekker, S., and E. Hollnagel. 2004. "Human Factors and Folk Models." *Cognition, Technology & Work* 6 (2): 79–86. doi:10.1007/s10111-003-0136-9.

Derksen, M., S. Meirmans, J. Brenninkmeijer, J. Pols, A. de Boer, H. van Eyghen, S. Gayet, R. Groenwold, D. Hernaus, P. Huijnen, N. Jonker, R. de Kleijn, C.F. Kroll, A.-M. Krypotos,

N. van der Laan, K. Luijken, E. Meijer, R.S.A. Pear, R. Peels, R. Peeters, C.C.S. Rulkens, C. Scholz, N. Smit, R. Stapel, and J. de Winter. 2024. "Replication Studies in The Netherlands: Lessons Learned and Recommendations for Funders, Publishers and Editors, and Universities." *Accountability in Research*. doi:10.1080/08989621.2024.2383349.

De Winter, J.C.F. 2014. "Controversy in Human Factors Constructs and the Explosive Use of the NASA TLX: A Measurement Perspective." *Cognition, Technology & Work 16* (3): 289–297. doi:10.1007/s10111-014-0275-1.

De Winter, J.C.F., and Dodou, D. 2021. "Pitfalls of Statistical Methods in Traffic Psychology." In *International Encyclopedia of Transportation*, edited by R. Vickerman, 87–95. Amsterdam, The Netherlands: Elsevier. doi:10.1016/B978-0-08-102671-7.10665-7.

De Winter, J.C.F., D. Dodou, and A.H.A. Stienen. 2023. "ChatGPT in Education: Empowering Educators through Methods for Recognition and Assessment." *Informatics 10* (4): 87. doi:10.3390/informatics10040087.

De Winter, J.C.F., N.A. Stanton, and Y.B. Eisma. 2021. "Is the Take-over Paradigm a Mere Convenience?" *Transportation Research Interdisciplinary Perspectives 10*: 100370. doi:10.1016/j.trip.2021.100370.

Dutch Research Council. 2024 Dutch Research Agenda. https://www.nwo.nl/sites/nwo/files/media-files/nwa_-_all_factsheets_en.pdf

Ebel, P., P. Bazilinskyy, M. Colley, C.M. Goodridge, P. Hock, C.P. Janssen, H. Sandhaus, A. Ramakrishnan Srinivasan, and P. Wintersberger. 2024. "Changing Lanes toward Open Science: Openness and Transparency in Automotive User Research." In *Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 94–105. Stanford, CA: ACM. doi:10.1145/3640792.3675730.

Edwards, M.A., and S. Roy. 2017. "Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition." *Environmental Engineering Science 34* (1): 51–61. doi:10.1089/ees.2016.0223.

Endsley, M.R. 1988. "Situation Awareness Global Assessment Technique (SAGAT)." In *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, 789–795. Dayton, OH: IEEE. doi:10.1109/NAECON.1988.195097.

Escamilla, E., M. Klein, T. Cooper, V. Rampin, M.C. Weigle, and M.L. Nelson. 2022. "The Rise of GitHub in Scholarly Publications." In *Linking Theory and Practice of Digital Libraries. TPDL 2022*, edited by G. Silvello, O. Corcho, P. Manghi, G.M. di Nunzio, K. Golub, N. Ferro, and A. Poggi, 187–200. Cham: Springer. doi:10.1007/978-3-031-16802-4_15.

Fridman, L., D.E. Brown, J. Kindelsberger, L. Angell, B. Mehler, and B. Reimer. 2019. "Human Side of Tesla Autopilot: Exploration of Functional Vigilance in Real-World Human-Machine Collaboration." Accessed August 27, 2024. https://web.archive.org/web/20190407172429/https://hcai.mit.edu/tesla-autopilot-human-side.pdf

FSD Community Tracker. 2024. "Tesla FSD Tracker." Accessed August 27, 2024. https://teslafsdtracker.com/Main

Grier, R.A. 2015. "How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 59* (1): 1727–1731. doi:10.1177/1541931215591373.

Hannon, D., E. Rantanen, B. Sawyer, A. Hughes, K. Darveau, R. O'Donnell, J. Intriligator, and J.D. Lee. 2020. "The Education of the Human Factors Engineer in the Age of Data Science." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 64* (1): 480–484. doi:10.1177/1071181320641109.

Hart, S.G., and L.E. Staveland. 1988. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In *Human Mental Workload*, edited by P.A. Hancock and N. Meshkati, 139–183. North-Holland: Elsevier. doi:10.1016/S0166-4115(08)62386-9.

Holcombe, A. 2021. "Authorship: Giving Credit Where It's Due." Accessed August 27, 2024. https://www.apa.org/pubs/journals/resources/publishing-tips/giving-credit

Hu, X., Z. Zheng, D. Chen, and J. Sun. 2023. "Autonomous Vehicle's Impact on Traffic: Empirical Evidence from Waymo Open Dataset and Implications from Modelling." *IEEE Transactions on Intelligent Transportation Systems 24* (6): 6711–6724. doi:10.1109/TITS.2023.3258145.

Hubbard, W., and C.P. Starger. 2024. "The Collision Course between Outdated State Laws and Automated Vehicles." *SSRN*. doi:10.2139/ssrn.4916418.

Ioannidis, J.P.A. 2011. "Fund People Not Projects." *Nature 477* (7366): 529–531. doi:10.1038/477529a.

Jansen, Y., J.B. Vornhagen, O. Iarygina, K. Salehzadeh Niksirat, L. Besançon, P. Dragicevic, J. Gori, and C. Wacharamanotham. 2024. "The Many Ways of Being Transparent in Human-Computer Interaction Research." *OSF*. doi:10.31219/osf.io/2wze6.

Kobak, D., R.G. Márquez, E.Á. Horvát, and J. Lause. 2024. "Delving into ChatGPT Usage in Academic Writing through Excess Vocabulary." arXiv. doi:10.48550/arXiv.2406.07016.

Kyriakidis, M., J.C.F. de Winter, N. Stanton, T. Bellet, B. van Arem, K. Brookhuis, M.H. Martens, K. Bengler, J. Andersson, N. Merat, N. Reed, M. Flament, M. Hagenzieker, and R. Happee. 2019. "A Human Factors Perspective on Automated Driving." *Theoretical Issues in Ergonomics Science 20* (3): 223–249. doi:10.1080/1463922X.2017.1293187.

Li, G., Y. Jiao, V.L. Knoop, S.C. Calvert, and J.W.C. van Lint. 2023. "Large Car-following Data Based on Lyft Level-5 Open Dataset: Following Autonomous Vehicles vs. human-Driven Vehicles.." In *Proceedings of the 2023 IEEE 26th International Conference on Intelligent Transportation Systems*, 5818–5823. Bilbao, Spain: IEEE. doi:10.1109/ITSC57777.2023.10422574.

Liu, P. in press. "Reflections on automation complacency." *International Journal of Human–Computer Interaction*. doi:10.1080/10447318.2023.2265240.

Matthews, G., J. De Winter, and P.A. Hancock. 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. doi:10.1080/1463922X.2018.1547459.

Matthews, G., L.E. Reinerman-Jones, D.J. Barber, and J. Abich, IV. 2015. "The Psychometrics of Mental Workload: Multiple Measures Are Sensitive but Divergent." *Human Factors 57* (1): 125–143. doi:10.1177/0018720814539505.

McCarley, J.S., L.E. Rose, S. Fischer, M.S. Haney, A. Hyk, K. Kuhn, B. Meredith, J.R. Moussaoui, F. Munoz Gomez Andrade, E.M. Pietrok, and S.L. Silberstein. 2023. "Open Science Practices in The Journal Human Factors: 2017–2022." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 67* (1): 1831–1836. doi:10.1177/21695067231192276.

Meehl, P.E. 1978. "Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology." *Journal of Consulting and Clinical Psychology 46* (4): 806–834. doi:10.1037/0022-006X.46.4.806.

Meister, D. 1999. "The Formal History of HFE." In *The History of Human Factors and Ergonomics*, edited by D. Meister, 146–182. Mahwah, NJ: Lawrence Erlbaum Associates. doi:10.1201/9781315276069-4.

Mercedes-Benz. 2023. "Certification for SAE Level 3 System for U.S. Market." Accessed August 27, 2024. https://group.mercedes-benz.com/innovation/product-innovation/autonomous-driving/drive-pilot-nevada.html

Nordhoff, S., J.D. Lee, S.C. Calvert, S. Berge, M. Hagenzieker, and R. Happee. 2023. "(Mis-)Use of Standard Autopilot and Full Self-Driving (FSD) Beta: Results from Interviews with Users of Tesla's FSD Beta." *Frontiers in Psychology 14*: 1101520. doi:10.3389/fpsyg.2023.1101520.

Proctor, R.W., and T. van Zandt. 2017. "Research Methods in Human Factors." In *Human Factors in Simple and Complex Systems*, edited by R.W. Proctor and T. van Zandt, 25–52. CRC Press. doi:10.1201/9781420063608-6.

Salmon, P.M., C. Burns, S. Broadbent, S. Chari, R. Clay-Williams, P.A. Hancock, D.P. Jenkins, A. Mills, K. Parnell, K.L. Plant, G.J.M. Read, S. Sharples, S. Shorrock, N.A. Stanton, G.H. Walker, C. Williams, P. Waterson, and M.S. Young. in press. "The Chartered Institute of Ergonomics and Human Factors at 75: Perspectives on Contemporary Challenges and Future Directions for Ergonomics and Human Factors." *Ergonomics*. doi:10.1080/00140139.2024.2378355.

Scholl, B.J. 2017. "Reliability in Psychology: Means versus Ends." *APS Observer 30*: 38–39. https://perception.yale.edu/papers/17-Scholl-APSObserver.pdf

Sheridan, T.B. 1986. "Forty-Five Years of Man–Machine Systems: History and Trends." In *Proceedings of the 2nd IFAC/IFIP/IFORS/IEA Conference*, Varese, Italy, 1–9. Oxford, England: Pergamon Press Publisher. doi:10.1016/B978-0-08-032566-8.50006-0.

Taylor, R.M. 1990. "Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design." In *Situational Awareness*, edited by R.M. Taylor, 111–128. London, UK: Routledge. doi:10.4324/9781315087924-8.

Teslamagazine. 2024. "FSD WEL OF NIET KOPEN? *Alles* over FSD Supervised [FSD TO BUY OR NOT TO BUY? Everything about FSD Supervised]." August 17. https://www.teslamagazine.nl/fsd_wel_of_niet_kopen-528836.html

UNECE. 2021. "UN Regulation on Automated Lane Keeping Systems (ALKS) Extended to Trucks, Buses and Coaches." Accessed August 27, 2024. https://unece.org/sustainable-development/press/un-regulation-automated-lane-keeping-systems-alks-extended-trucks

Vidulich, M.A., and P.S. Tsang. 2012. "Mental Workload and Situation Awareness." In *Handbook of Human Factors and Ergonomics*, edited by G. Salvendy, 4th ed., 243–273. New York, NY: Wiley. doi:10.1002/9781118131350.ch8.

WholeMars [Whole Mars Catalog]. 2024a. "Raw 1x: Hands Free Tesla FSD 12.4.1: West Hollywood to Palos Verdes at Night with Zero Interventions [Video]." YouTube. June 25. https://www.youtube.com/watch?v=7ttRMMoDxOg

WholeMars [Whole Mars Catalog]. 2024b. "Tesla FSD 12.5.1.3 Drives One Hour Through San Francisco with Zero Interventions & My Commentary [Video]." YouTube. August 11. https://www.youtube.com/watch?v=4RZfkU1QgTI

Yeh, Y.-Y., and C.D. Wickens. 1988. "Dissociation of Performance and Subjective Measures of Workload." *Human Factors: The Journal of the Human Factors and Ergonomics Society 30* (1): 111–120. doi:10.1177/001872088803000110.

Young, M.S., K.A. Brookhuis, C.D. Wickens, and P.A. Hancock. 2015. "State of Science: Mental Workload in Ergonomics." *Ergonomics 58* (1): 1–17. doi:10.1080/00140139.2014.956151.

Zhang, K., S. Wang, N. Jia, L. Zhao, C. Han, and L. Li. 2024. "Integrating Visual Large Language Model and Reasoning Chain for Driver Behavior Analysis and Risk Assessment." *Accident, Analysis and Prevention 198*: 107497. doi:10.1016/j.aap.2024.107497.

Zwaan, R.A., D. Pecher, G. Paolacci, S. Bouwmeester, P. Verkoeijen, K. Dijkstra, and R. Zeelenberg. 2018. "Participant Nonnaiveté and the Reproducibility of Cognitive Psychology." *Psychonomic Bulletin & Review 25* (5): 1968–1972. doi:10.3758/s13423-017-1348-y.