Road-safety-II
Opportunities and barriers for an enhanced road safety vision

Papadimitriou, Eleonora; Pooyan Afghari, Amir; Tselentis, Dimitrios; van Gelder, Pieter

DOI
10.1016/j.aap.2022.106723

Publication date
2022

Document Version
Final published version

Published in
Accident Analysis and Prevention

Accident Analysis and Prevention 174 (2022) 106723

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Short Communication

Road-safety-II: Opportunities and barriers for an enhanced road safety vision

Eleonora Papadimitriou *, Amir Pooyan Afghari, Dimitrios Tselentis, Pieter van Gelder

Delft University of Technology, Faculty of Technology, Policy & Management, Section Safety & Security Science, Jaffalaan 5, 2628 BX Delft, the Netherlands

ARTICLE INFO

Keywords:
Safety-II
Road safety
Traffic conflicts
Driver behaviour

ABSTRACT

Road safety research is largely focused on prediction and prevention of technical, human or organisational failures that may result in critical conflicts or crashes. Indicators of traffic risk aim to capture the passage to unsafe states. However, research in other industries has shown that it is meaningful to analyse safety along the whole spectrum of behaviours. Knowing the causes and patterns of "successful" interactions, rather than failures, could give new insights on the complexity of the system and the adaptability and resilience of its users in handling the inherent risks. The concept is known as Safety-II and has been extensively explored in the aviation, healthcare and process engineering domains. In this paper, we explore a new Safety-II paradigm for road safety research. We briefly review Safety-II applications in other sectors. We then present a Safety-II model for road safety, by means of an inverse version of Hyden's "safety pyramid". Furthermore, we discuss a number of key road safety goals, theories, analysis methods and data sources and map them into a tentative taxonomy of Safety-I and Safety-II applications. It is concluded that there can be opportunities and benefits from adopting this new mindset, in order to complement existing approaches.

1. Introduction

Road safety research is largely focused on prevention of human error, as the primary cause of roadway crashes. In the Safe Systems approach, the shared responsibility between users, operators and authorities is aiming at prevention of failures or the correction or tolerance of errors. This has been expressed in a number of safety visions and management programmes, at the international or national level e.g. the EC Vision Zero (European Union, 2019) and the UN Sustainable Development target 3.6, the Dutch Sustainable Safety Vision (SWOV, 2018).

In the recent years, a more proactive approach to road safety management is promoted. Instead of looking at historical crash data to identify crash causes and patterns, research is looking into critical events and near misses, through studying traffic conflicts (Arun et al. 2021). Driver assistance and dynamic traffic management systems are also aiming to learn from real-time data, in order to prevent errors and conflicts. Connected and automated vehicles vision to prevent and correct these errors and failures by removing the human component from the system, and optimizing the technical component (vehicle automation and connectivity, digital infrastructures, V2X communication etc.).

The above examples are not exhaustive, however they are representative of what can be understood as an underpinning common rationale: the study of critical events is aiming to learn about the causes of incidents before they occur. This, however, may not present a complete picture of the reality on the roads. First of all, humans learn from their mistakes, and therefore preventing failures and errors may deprive them of the required skills to adequately handle critical events should those failures and errors occur. For example, inexperienced drowsy drivers are less likely to anticipate traffic hazards, whereas more experienced drowsy drivers develop this hazard anticipation over time (Smith et al., 2009). In addition, road users can perform positive behaviours (or "successes"), which may reduce traffic risk, and are not necessarily the opposite of failures or errors. For example, electric scooters who give hand signals while turning, allow drivers to anticipate their movement and are less likely to be involved in potential conflicts (Uluk et al., 2020); yet, not giving a hand signal is not an error or a failure.

At the same time, existing definitions of risk in road safety cannot fully reflect these characteristics of adaptive human behaviour, because they are focused on two states for safety, *safe* or *unsafe*. In this paper, we suggest a shift of mindset in road safety research, focusing not only on "what can go wrong", but also on "how to keep things going right". The

* Corresponding author.
E-mail address: e.papadimitriou@tudelft.nl (E. Papadimitriou).

https://doi.org/10.1016/j.aap.2022.106723
Received 15 March 2022; Received in revised form 26 May 2022; Accepted 27 May 2022
Available online 14 June 2022

concept is not new; it has been popular in aviation safety, but also in broader safety related disciplines, such as process engineering, occupational safety and healthcare –referred to as Safety-II.

The objective of this research is to assess the potential for a Safety-II approach in road safety. We propose that road safety should not be seen as a dual state, but as a continuum along which road users could behave differently and develop mobility skills from safer parts of this continuum to deal with less safe parts.

This paper starts with the presentation of the concept, and its characteristics compared to the "traditional" approach – referred to as Safety-I (section 2). It summarizes relevant findings from other disciplines on the strengths and weaknesses of the concept (section 3). Finally, it proposes a set of context-specific applications and explores the potential opportunities and benefits from introducing it in road safety research (section 4).

2. The Safety-II concept

Safety-II was introduced in 2014 by Erik Hollnagel, as an alternative paradigm to managing safety in various domains (Hollnagel, 2014). While originally intended for the healthcare sector, it was quickly endorsed by the aviation industry and the chemical industry. Hollnagel et al. (2015) define the goal of Safety-II as "how to keep things going right", as opposed to focusing on "what can go wrong" (Safety-I). Traditionally, Safety-I management is looking into human, technical and organisational factors that cause accidents, by investigating previous events, analysing historical data and decomposing the system into components, on which failures can be attached. The main goal is to either eliminate failures, or keep their number "as low as reasonably practicable" (Hollnagel, 2014).

In this context, humans are considered the main cause of accidents, however it has been argued that this is a no longer meaningful focus: first, the mechanisms of human 'failures' are far more complex than that of technical failures. Second, systems are becoming increasingly complex and therefore harder to decompose. In most modern systems, humans engage in a complex interaction with infrastructure, technology, environment, other humans etc., while following existing regulations, rules and norms that govern the system. Therefore, the Safety-II approach emerged from the need to understand safety in complex socio-technical systems.

Safety-II is further based on the principle that, in most systems, failures are from extremely rare to occasional, whereas successes are abundant. Therefore, not only there are learnings to be uptaken by the successful system operation, but also the variability in the system operation can be exploited to ensure safety, rather than be discarded.

It is thereby stressed that "successes" are based on the skills, flexibility and adjustment potential of humans in the increasingly variable conditions that are experienced in routine operations (Sujan et al., 2017), and less so to the pertinence of rules and regulations or the compliance of humans to them. Hollnagel et al. (2015) note that safety management is often made under the assumption of operations taking place in an ideal or "work-as-imagined" mode, and largely ignoring the value of studying the "work-as-done", particularly in terms of successes in daily practice. Martinetti et al. (2019) note that, as a consequence, the aim of Safety-I is to standardise safety rules and procedures, while Safety-II aims to learn from anomalies and deviations in a visioning towards evolving rules and procedures.

The idea is directly linked to resilience engineering, which assumes that both positive and adverse outcomes are based on the daily performance adjustments of human operators. Quoting Shorrock and Licu (2014), "*we need to understand the variability that we need and want, and the variability that we do not want*". As a result of this rationale, safety is seen as a continuum entailing the wide variability of system performance. Provan et al. (2020) underline that Safety-I is based on the notions of control and 'centralised' safety management, while Safety-II deviates from that model towards adaptive and decentralised safety management.

Finally, it should be underlined that Safety-I and Safety-II concepts are considered complementary by most researchers (e.g. Ham, 2021); both are valuable in specific contexts and can give useful insights. It is argued however that Safety-II should be used when the system's variability of conditions exceeds a certain threshold, because it is found that humans exhibit the highest resilience and adaptability (Kubo & Nakanishi, 2019).

3. Experiences from Safety-II applications in various domains

A literature search was carried out in Scopus with the search term "Safety II" (in abstract, title, keywords). The search returned 235 'hits', and the items were screened for relevance on the basis of the title and the abstract. Given that the experiences from the aviation sector were underrepresented in the scientific literature, several additional sources (reports, communication briefs and other informal publications) were retrieved by means of a web search on Safety-II in aviation. Moreover, certain road safety studies may have taken a Safety-II approach without using the term as such, therefore an additional search was made by means of broader terms such as "optimal driving", "white spots OR zero-fatality-roads" etc. Taking into account accessibility and relevance of full texts, eventually, 39 articles were selected, which can be broadly categorized as follows: 17 from the transport sector, 7 from the chemical industry, construction etc., 14 from the health & occupational safety sector, and 4 of a general scope on Safety-II.

A detailed review of these studies is beyond the scope of this short communication. We hereby summarise the main insights from the experiences in other domains, as follows: In several cases, individual and technical factors tend to be approached with Safety-I, while organisational and safety culture factors are more often understood as Safety-II, e.g. in Qiao et al. (2021), Wang et al. (2020). Nævestad et al. (2021) reviewed safety culture factors in the transport and pertroleum industries and underline the potential for positive contribution of the human element, as well as for more flexible regulatory and organisational models, in enhancing safety culture.

Safety-II methods used to identify and feed-forward what went wrong and what went well in various operations include questionnaire surveys (e.g. Wang et al., 2020), focus groups (e.g. Wahl et al., 2020), and in-depth interviews (e.g. Saward & Stanton, 2018, McCarthy, 2020); these were found to enable the sharing of knowledge and learning-by-doing experiences in several domains, including maritime and aviation. The use of empirical observations or simulator experiment data is also used in some studies to monitor the operator behaviour. Intelligent applications for operator monitoring and their integration with training material have also been used to identify positive behaviours, e.g. in Qiao et al. (2021) in maritime.

Formal methods to detect anomalies in Safety-II mostly studies include qualitative and descriptive methods, such as the FRAM (functional reasonance analysis method) (Danial et al. (2021); semi-quantative and quantitative methods may include clustering techniques to identify performance patterns (e.g. Patriarca et al. (2017), or regression trees to identify both failures and successes e.g. (Ham & Park, 2020). The lack of concrete quantitative applications is one of the main criticisms received by Safety-II e.g. in Cooper (2020) regarding occupational safety.

Researchers underline that the digitalization of transport operations and the abundance of Big Data and related analytics tools provide new opportunities to apply the Safety-II idea specifically in transport e.g. Walker (2017) for aviation, Parkinson and Bamford (2016) for railways. However, new metrics to measure the number of "successes" and positive outcomes need to be explicitly identified in each domain.

In road safety, there has been a small number of studies that have taken a Safety-II perspective, without explicitly using the term. For instance, from the road infrastructure perspective, the European Road Federation (ERF) White Roads project (https://erf.

be/projects/white-roads/) analysed road sections of Trans-European Road Networks with zero fatalities over the last 5 years, followed by a similar study on Spanish motorways (de la Peña-González & Zaragoza-Ramírez, 2015).

A limited number of studies has also taken a Safety-II approach from a driver behaviour perspective. Shinar et al. (2001) studied the associations between demographic characteristics and safe driving behaviours regarding seat belt use, speeding and alcohol impairment while driving. Safe driving behaviour characteristics such as safe distance and driving under the speed limit have also been studied in Mazureck & van Hattem (2006). Safe driver profiles have also been explored by Tselentis et al. (2019) in comparison to other more risky practices.

Recently, safe driving behaviour recognition such as smooth steering manoeuvring has been researched as an approach to empower autonomous driving technologies (Farag, 2019). To this end, approaches were also developed to separate typical (safe) driving behaviour from anomalous (unsafe) ones (Tejada et al., 2020), in order to identify safe driving patterns from human behaviour and adopt them to improve the safety of autonomous vehicles.

The above review showed that the application of Safety-II in road safety has not been explicitly examined so far. In the following sections, we demonstrate that the nature of road risk on the one hand, and the availability of data and tools on the other hand, make road safety a promising field to benefit from the Safety-II paradigm.

4. A framework for Road-Safety-II

4.1. Enhanced road safety goals

In order to highlight the potential for Safety-II approaches in road safety, we use the pyramid of road safety outcomes (Hydén, 1987), which largely forms the basis of road safety research in the current state of the art (Fig. 1). This pyramid indicates that the actual crashes occurring are only the ‘tip of the iceberg’, and much more insights may be obtained by analysing the causes of serious incidents (near-misses) or slight incidents (conflicts). This is based on the assumption that in several – yet not all – contexts, conflicts correlate very well with crashes. At the bottom of this pyramid are the undisturbed interactions, which consist the vast majority of traffic movements in the system.

We thereby suggest that a Safety-II approach would result from inversing Hyden’s pyramid, by considering near-misses and conflicts as successful operations that control or mitigate risk, rather than failures that increase risk (see Fig. 1). Moreover, Road-Safety-II would put strong focus on the large yet seldom explored ‘undisturbed interactions’ layer, by considering it as a manifestation of the management of the inherent risks of the traffic system during every day driving, rather than an area of uninteresting ‘noise’. Finally, the ‘undisturbed interactions’ layer would offer the potential of identifying and understanding conditions in which the road system and its users exhibit an optimal safe performance.

In that sense, Safety-II can be considered to have two main research goals: (i) to learn from the whole spectrum of traffic behaviour, and (ii) to identify optimal safe traffic behaviour.

4.1.1. *Learning from the whole spectrum of traffic behaviour*

There may be significant benefits in learning from the flexibility and adaptability of traffic participants in keeping the system safe. Road rules and regulations are primarily established to reduce the number of conflicts. However, a paradoxical consequence of this low number of conflicts – referred to as “the paradox of safety and risk” (James, 1980) – is that road users are less exposed to risky manoeuvres and therefore, they may not develop safe human driving patterns with the required skills to handle these manoeuvers. Alternatively, road users may acquire such skills if they are exposed (in an unharmful way such as via simulators or virtual reality) to a chaotic environment over and over. Moreover, it is often the case that experimental road safety evaluations result in non-significant effects of measures over time, because (among other things) road users tend to adapt their behaviour to the introduction of measures, therefore “*measures tend to erode as they become more commonly used*” (quoting Elvik, 2021).

In addition, not all types of road users’ behaviours are negative and risky. Drivers who timely adhere to advisory speed limit on highway ramps are less likely to be involved in certain types of crashes (Lee and Abdel-Aty, 2009). Pedestrians who let a speeding car pass, even if they have priority, are less likely to be involved in risky events. None of these two behaviours’ ‘opposite’ would be seen as a failure or an error that could be prevented by discouragement. These behaviours are simply successes or positive behaviours that could be encouraged among road users once identified.

Positive behaviours may also be used to further improve the safety of autonomous driving technologies (Farag, 2019, Tejada et al., 2020)). It is shown that it is feasible to teach entire driving tasks such as lane and road following using advanced technologies and sophisticated methodologies. Another example is newer motor insurance industry concepts such as the Pay-How-You-Drive (PAYD) scheme, which are considering both risky and safe driving characteristics to ‘score’ drivers and adjust their insurance premiums (Tselentis et al., 2017).

The current Safety-I road safety mindset also considers a linear one-way effect of external factors on risk; it is assumed that by removing a risk factor, the level of risk decreases as much as it would increase when that risk factor is present, while in reality interactions may be much more complex. For example, while mobile phone distraction is associated with increased risk, distracted drivers tend to reduce their speed, increase their headway, or scan their environment more often as a result of risk-compensating behaviour (Oviedo-Trespalacios et al., 2020). This

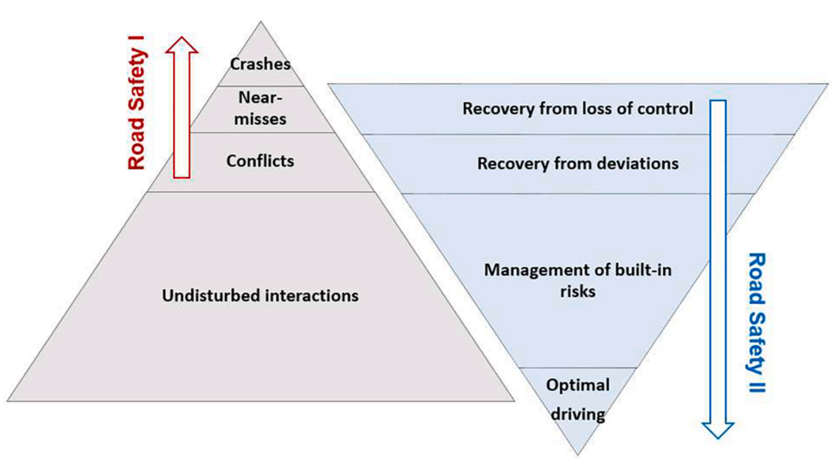

Fig. 1. Road-Safety-I and Road-Safety-II models.

is in line with the risk homeostasis theory, which proposes that, for any activity, people accept a particular level of subjectively evaluated risk to their health and safety in order to gain from a range of benefits associated with that activity (Wilde, 1982). Moreover, while automated vehicles may increase safety on the roads by detecting pedestrians in advance of most fatal collisions (Combs et al., 2019), fully automated vehicles will be risk-averse and thus pedestrians will be able to behave with impunity when interacting with these vehicles, due to the secure knowledge that they will yield (Afghari et al, 2021).

Hauer (2021) demonstrates the biases that can arise in crash causation investigations when only the presence of certain factors, and not also the absence of certain others, is taken into account; the author argues that road user behaviour should not be compared to a certain norm or expectation, but to a more sensitive and context-specific framework that will be highly variable.

These complexities exist in the effects of engineering design on risk too, because design promotes human behaviour. Therefore, infrastructure and vehicle safety factors can also be better understood through the observation of human behaviour in all contexts, and not only in critical situations. For example, there are mixed findings about the effects of horizontal curves on the likelihood of motor vehicle crashes; some studies have found that sharper radius increases this likelihood (Geedipally et al., 2019) whereas other studies have found an opposite effect (Schneider et al., 2010). The unobserved heterogeneity in the behaviour of road users amplifies the complexity of such interaction between risk and its contributing factors. In addition, the pattern of behaviours may change across time due to changes in external factors over time (Mannering, 2018). While many advanced methodologies have been introduced to capture such a between-individual heterogeneity (Mannering et al., 2016), their aim has been to predict more accurately the occurrence of risk, while the possibilities to study and disentangle this heterogeneity itself as a source of knowledge has received little attention.

4.1.2. Identifying optimal safe behaviour

The tip of the Safety-II inverse iceberg is the observation of optimal safe behaviour. From a conceptual viewpoint, "optimal driving" can be defined as the limitation or minimization of those driving behavioural characteristics that increase road risk. It should be understood as different from an "excellent" or "perfect" behaviour, because the latter cannot be consistently achieved by human road users; because of the increased complexity and variability of driving conditions, human drivers are mostly able to "optimise" their behaviour under their multiple objectives and system constraints.

The question of optimal driving has been the focus on several studies on automated vehicles safety, in which two broad approaches have been examined (Papadimitriou et al., 2022): (i) a data-driven approach, in which machine learning is used to identify driving patterns associated with low risk on the basis of naturalistic driving data, vs. (ii) an expert-based approach, resulting from knowledge, experience and consensus of experts. For Road-Safety-II, the first option would be preferable, so that the identification of optimal driving is a distinct task within the analysis of the whole spectrum of behaviour.

Such a data-driven approach would use the common metrics of driver safety performance, e.g. speeding, lateral position, headway, the number of harsh events (i.e. acceleration, braking and cornering), mobile phone use. For example, a risky driver is one for whom we observe a higher number of harsh accelerations or speed limit violations and therefore a driver who performs significantly less such events is closer to "optimal driving" (Tselentis et al., 2019).

From an analytics perspective, the outlier behaviours, both safe and risky, can be detected by observing the tails of the distributions of these driving metrics. This is a common Safety-II hypothesis adopted by Hollnagel et al. (2015), assuming that, while one tail corresponds to 'things that go wrong', the other one indicates "positive surprises" resulting from timeliness, excellence or innovation. The detection of these positive outliers of driving behaviour could lead to the identification of a new threshold for a certain metric, and therefore to the definition of an optimal driving benchmark.

Of course, this benchmark will differ depending on the road type, the weather conditions, the drivers' sample characteristics (e.g. country, age distribution). Therefore, this bottom-up exercise should be performed for various scenarios, and for several combinations of driving metrics, not limited to those mentioned above. Recent studies in this direction (e.g. Tejada et al., 2020) suggest a multivariate characterization of safe driving, taking into account both proximity metrics and kinematic variables. There may be additional context-specific variables, as well as qualitative information that could be added in this respect. Moreover, not all distributions will have two tails, and therefore the shape for each behavioural aspect examined is itself a first step to understanding how and whether optimal driving can be defined.

4.2. Taxonomy of Road-Safety-I and Road-Safety-II

A tentative classification of the features of the two models is shown in Table 1. Particular focus is put on the pertinence and relevance of basic road safety theories, methodologies and data sources. Regarding the theories, the Poisson-theorem or the Extreme Value theory for modelling rare events have long been the foundations of predicting crashes and other undesirable incidents in the traffic system (e.g. Papadimitriou et al., 2013; Songchitruksa & Tarko, 2006; Afghari et al., 2018). These models aim to quantify exceedance probabilities of thresholds or boundaries of a "safe envelope", such as the operation boundary, the controllable boundary, the viable boundary (Fig. 2), each one of them indicating a deviation from normal operations. For example, the recent i-Dreams naturalistic driving study is developing a 'safety tolerance zone' concept with 3 phases (normal driving, danger phase and avoidable crash phase) (Michelaraki et al., 2022).

Table 1
Taxonomy of Safety-I and Safety-II objectives and methods for road safety.

<table>
<tr><th></th><th>Safety-I</th><th>Safety-II</th></tr>
<tr><td>Objectives</td><td>Reactive/Proactive</td><td>Proactive</td></tr>
<tr><td></td><td>Prevent failures/errors</td><td>Maintain "safe" status</td></tr>
<tr><td></td><td colspan="2">Detection of anomalies and deviations</td></tr>
<tr><td></td><td>Learn from past accidents</td><td>Learn from (successful) everyday operations</td></tr>
<tr><td></td><td>Identify causes (linear causality)</td><td>Understand complexity</td></tr>
<tr><td>Common assumptions</td><td>Causes are linearly related to impacts</td><td>Causes and impacts can be related in diverse context-specific ways</td></tr>
<tr><td></td><td>Conflicts and near-misses are pre-cursors of crashes</td><td>Conflicts and near-misses are positive outcomes</td></tr>
<tr><td>Relevant theories</td><td>Theory of planned behaviour</td><td>Risk compensation, behavioural adaptation</td></tr>
<tr><td></td><td>Rare events/Poisson statistics</td><td></td></tr>
<tr><td></td><td>Rare events/Extreme value theory</td><td></td></tr>
<tr><td>Common methods</td><td>Crash Prediction Models</td><td></td></tr>
<tr><td></td><td>Crash severity models</td><td></td></tr>
<tr><td></td><td colspan="2">Discrete choice (outcome) models</td></tr>
<tr><td></td><td colspan="2">Structural Equation Models (SEM)</td></tr>
<tr><td></td><td colspan="2">Bayesian probabilistic modelling</td></tr>
<tr><td></td><td colspan="2">Artificial Intelligence/Machine Learning</td></tr>
<tr><td>Data</td><td>Crash statistics</td><td></td></tr>
<tr><td></td><td colspan="2">Attitudinal/behavioural data</td></tr>
<tr><td></td><td colspan="2">Traffic conflicts & SMoS (simulation, naturalistic & observational studies)</td></tr>
<tr><td></td><td colspan="2">Big data, crowdsourcing</td></tr>
<tr><td>Outputs</td><td>Definition of unacceptable risk level</td><td>Definition of desirable risk level</td></tr>
<tr><td></td><td>Crash prediction</td><td>Identification of optimal driving</td></tr>
<tr><td></td><td>Driver error prediction and mitigation</td><td>Driver education, information and alert</td></tr>
<tr><td></td><td>Standardization of measures</td><td>Learnings and evolution of measures</td></tr>
</table>

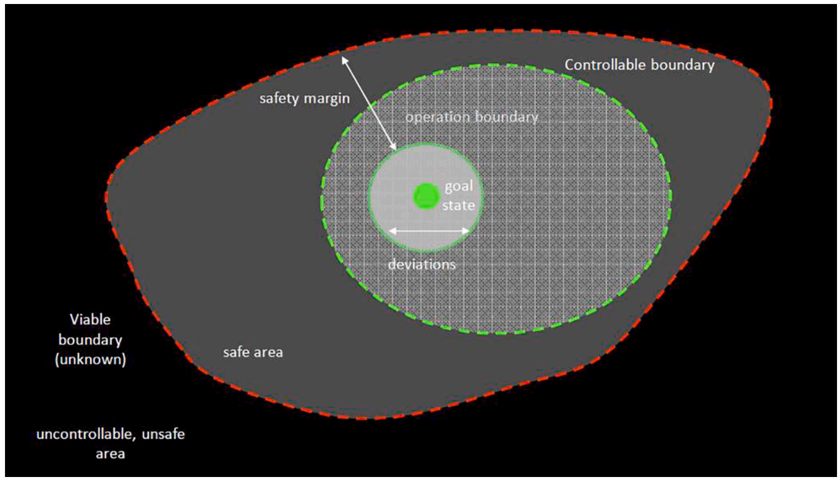

Fig. 2. The use of multiple boundaries in a "safe envelope".

At the same time, human factors based theories, e.g. the theory of planned behaviour (e.g. Parker et al., 1992) or Fuller's task capability model (Fuller, 2000) are similarly oriented towards the understanding of causes of errors and failures. On the other hand, the theories of risk compensation or behavioural adaptation (Summala, 1997), that are applied in risk moderation hypotheses, are very suitable in a Safety-II context as they reflect a "keeping the system safe" approach.

As regards methods and data, earlier road safety research was based on crash statistics (historical data). In the last decades, traffic conflicts and surrogate measures of safety (SMoS) were adopted as the basis of more proactive risk assessment (Arun et al., 2021). The relevant data come from driving simulator studies or other experimental setups, as well as from naturalistic driving studies. Moreover, a wealth of data are available from in to vehicle sensors, road user portable devices or wearables, and other Big Data that are collected by various service providers (Stylianou et al., 2019).

While this data cover the 'undisturbed interactions' system states, they are mostly used for crash risk prediction, investigation of the relationship between traffic flow and safety, and ADAS (Advanced Driver Assistance Systems) development. For instance, most in-vehicle systems constantly monitor the vehicle/driver, but are designed to generate risk alerts and the data provided by these devices are typically event-based. As Walker (2017) points out, the introduction of Flight Data Recorders (FDR) in aviation led to an abundance of data on successful handling of critical operations, while the emergence of AI and big data analytics shed new light to the determinants of safety in aviation. Given the recent developments in road safety data availability through vehicle, user or infrastructure sensors, the road sector can test similar safety scenarios.

Table 1 is certainly not exhaustive on road safety methods, theories and data; it includes a number of common and representative features of road safety research in order to demonstrate the main particularities and potential synergies of the two models. Overall, it can be said that, while Road-Safety-I and Road-Safety-II differ in terms of objectives, main assumptions and final outputs, they can be both tested and implemented by means of the same data sources and quantitative techniques.

For instance, Extreme Value theory and discrete choice models can be used to predict optimal vehicle headways, instead of critical ones, and test whether the determinants and their effects are similar in sign and magnitude. Structural equation models are also pertinent for Safety-II analyses, by aiming to model the complex interactions of road, vehicle and individual constructs within the 'undisturbed interactions' layer. Big Data on driving behaviour can be analysed with ML techniques, for profiling the whole population of drivers and identifying behavioural patterns among them, with the aim of identifying compensation mechanisms, driver adaptability and other resilience patterns.

4.3. Opportunities and barriers

There are several opportunities from developing Safety-II research applications in road safety. First of all, the analysis of the state of the art shows that there is a large number of suitable and promising existing data sources and methods to support Road-Safety-II; further research and specific applications can be implemented rather directly, by means of a shift of focus in the objectives.

Second, the Safety-II approach has the potential to tackle different types of urgent road safety problems. On the one hand, in most industrialised countries there is a stagnation in road fatalities, and a more complete understanding of the complexity of the system is needed. On the other hand, the number of fatalities in low-to-middle-income countries remains unacceptable, while the traffic and behavioural patterns in these settings have not been explored in-depth. In both cases, Safety-II can reveal new traffic interaction mechanisms that are hidden in the broad "undisturbed interactions" or "slight conflicts" layers of the pyramid and enhance our knowledge on the capabilities and limitations of traffic participants.

Third, the Safety-II goals can assist in the development of more trustworthy road safety technology applications, e.g. setting enhanced safety goals for ADAS, developing new AI algorithms for automated driving, that are more human-mimic and thus more predictable.

At the same time, there are a number of technical and organisational barriers that need to be addressed. The Safety-II approach was initially developed for very rare safety incidents and accidents that concern a specific task or operation – this may be more relevant to industrial, aviation and maritime accidents, as well as in healthcare. It has been mostly tested in operations involving trained professionals. Road driving involves an infinite number of (often simultaneous) tasks in a dynamic environment with very small time scales. Moreover, 'everybody' is a driver, while the road safety sector is relatively loosely regulated compared to the other domains. However, the availability of data and methods enabling the testing of Safety-II application can ensure an efficient testing in the road safety domain.

5. Conclusions

This paper makes an exploratory analysis of the potential of Safety-II approach in road safety, and proposes a conceptual model, and a set of research goals, methods and data from the current state of the art for its testing and pilot implementation. In the next steps of this research, we will test the proposed road safety goals with actual data in order to explore the added value of the concept in concrete road safety issues.

Our study has some limitations, and further research is needed to reveal the full potential of a Road-Safety-II model. The much more extensive experiences of the application of the concept in the healthcare and occupational safety domains need to be reviewed, as this was out of the scope of the present paper. Moreover, a broader consultation is needed to identify additional road safety theories, models and data sources that have the potential to be exploited for Safety-II analyses.

In this paper we demonstrate that focusing solely on risk identification is not sufficient. The identification of good practice in preventing or mitigating traffic risk needs to be complemented with the identification of good practice in keeping the traffic system safe, through the analysis of drivers' experience, adaptability and resilience in handling traffic interactions and recovering from deviations. This will lead to a much more complete and clear depiction of 'how the system works' in all its complexity. It will also broaden the research on safety culture, by explicitly identifying the aspects that need to be more strongly 'defended'.

The road safety visions can be similarly enhanced. The Safety-II approach can provide new insights on the factors that contribute to safe behaviours and the road users' practices in staying safe in traffic, and allow to identify new ideas and areas of intervention for halving traffic fatalities by 2030 (European Commission, 2019). Given the urgency in addressing the ongoing road safety 'pandemic', Safety-II seems a promising and meaningful new research direction, in order to complement and strengthen the important existing efforts in the field.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work has been partially funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101027222.

References

Afghari, A.P., Washington, S., Haque, M.M., Li, Z., 2018. A comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk. Analyt. Methods Accid. Res. 18, 1–14.
Afghari, A.P., Papadimitriou, E., Li, X., Kaye, S.-A., Oviedo-Trespalacios, O., 2021. How much should a pedestrian be fined for intentionally blocking a fully automated vehicle? A random parameters beta hurdle model with heterogeneity in the variance of the beta distribution. Analyt. Methods Accid. Res. 32 (2021), 100186.
Arun, A., Haque, M.M., Bhaskar, A., Washington, S., Sayed, T., 2021. A systematic mapping review of surrogate safety assessment using traffic conflict techniques. Accid. Anal. Prevent. 153, 106016.
Combs, T.S., Sandt, L.S., Clamann, M.P., McDonald, N.C., 2019. Automated vehicles and pedestrian safety: exploring the promise and limits of pedestrian detection. Am. J. Prev. Med. 56 (1), 1–7.
Cooper M.D. (2020). The Emperor has no clothes: A critique of Safety-II. Article In Press, Corrected Proof, Safety Science 105047.
Danial, S.N., Smith, D., Veitch, B., 2021. A method to detect anomalies in complex socio-technical operations using structural similarity. J. Mar. Sci. Eng. 9, 212.
de la Peña-González, E., Zaragoza-Ramírez, A. (2015) Model for the identification and characterization of road sections without accidents (white sections). Application in the Spanish toll motorways network | [Modelo de identificación y caracterización de tramos de carreteras sin accidentes (tramos blancos). Aplicación a la red de autopistas de peaje españolas] Informes de la Construccion 67(540), e117.
Elvik, R., 2021. Why are there so few experimental road safety evaluation studies: could their findings explain it? Accid. Anal. Prevent. 163, 106467.
European Union. 2019. EU Road Safety Policy Framework 2021-2030 – Next Steps towards "Vision Zero".
Farag, W., 2019. Cloning safe driving behavior for self-driving cars using convolutional neural networks. Recent Patents Comput. Sci. 12 (2), 120–127.
Fuller, R., 2000. The task-capability interface model of the driving process. Recherche Transports Sécurité 66, 47–59.
Geedipally, S.R., Pratt, M.P., Lord, D., 2019. Effects of geometry and pavement friction on horizontal curve crash frequency. J. Transp. Safety Security 11 (2), 167–188.
Ham, D.-H., Park, J., 2020. Use of a big data analysis technique for extracting HRA data from event investigation reports based on the Safety-II concept. Reliab. Eng. Syst. Saf. 194, 106232.
Ham, D.-H., 2021. Safety-II and resilience engineering in a nutshell: an introductory guide to their concepts and methods. Safety Health Work 1, 10–19.
Hauer, E., 2021. Crash causation and prevention. Accid. Anal. Prev. 143, 105528.
Hollnagel, E., 2014. Safety-I and Safety-II: The Past and Future of Safety Management. Ashgate Publishing Ltd.
Hollnagel E., Wears R.L., Braithwaite J. (2015). From Safety-I to Safety-II: A White Paper. The Resilient Health Care Net: Published simultaneously by the University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia.
Hydén, C. (1987). The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique. Bulletin Lund Institute of Technology, Department, (70).
James, T., 1980. The paradox of safety and risk. J. Exp. Educ. 3 (2), 20–23. https://doi.org/10.1177/105382598000300204.
Kubo N., Nakanishi M. (2019) In What Conditions Do People Adopt "Resilient" Behavior for Safety?. In: Bagnara S., Tartaglia R., Albolino S., Alexander T., Fujita Y. (eds) Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). IEA 2018. Advances in Intelligent Systems and Computing, vol. 819. Springer, Cham. https://doi.org/10.1007/978-3-319-96089-0_54.
Lee, C., Abdel-Aty, M., 2009. Analysis of crashes on freeway ramps by location of crash and presence of advisory speed signs. J. Transp. Safety Security 1 (2), 121–134.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analyt. Methods Accid. Res. 11, 1–16.
Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Analyt. Methods Accid. Res. 17, 1–13.
Martinetti, A., Chatzimichailidou, M.M., Maida, L., van Dongen, L., 2019. Safety I-II, resilience and antifragility engineering: a debate explained through an accident occurring on a mobile elevating work platform. Int. J. Occup. Safety Ergon. 25 (1), 66–75.
Mazureck, U., van Hattem, J., 2006. Rewards for safe driving behavior: influence on following distance and speed. Transp. Res. Rec. 1980 (1), 31–38.
Michelaraki, E., Katrakazas, C., Afghari, A.P., Papadimitriou, E., Al Haddad, C., Yang, K. Antoniou, C., Brijs, T., Yannis, G. (2022). Modeling the concept of a Safety Tolerance Zone: State-of-the-art and proposed alternatives. In the Proceedings of the Road Safety and Simulation Conference, Athens June 2022.
McCarthy P. (2020) The Application of Safety II in Commercial Aviation – The Operational Learning Review (OLR). In: Harris D., Li WC. (eds) Engineering Psychology and Cognitive Ergonomics. Cognition and Design. HCII 2020. Lecture Notes in Computer Science, vol. 12187. Springer, Cham.
Storesund, N.T.-O., Hesjevoll, I., Elvik, R., 2021. How can regulatory authorities improve safety in organizations by influencing safety culture? A conceptual model of the relationships and a discussion of implications. Accid. Anal. Prevent. 159, 106228.
Qiao, W., Liu, Y., Ma, X., Lan, H., 2021. Cognitive gap and correlation of Safety-I and Safety-II: a case of maritime shipping safety management. Sustainability 2021, 13.
Oviedo-Trespalacios, O., Afghari, A.P., Haque, M.M., 2020. A hierarchical Bayesian multivariate ordered model of distracted drivers' decision to initiate risk-compensating behaviour. Analyt. Methods Accid. Res. 26, 100121.
Papadimitriou, E., Farah, H., Van de Kaa, G., Santoni de Sio, F., Hagenzieker, M., Van Gelder, P.H.A.J.M., 2022. Towards common ethical and safe 'behaviour' standards for automated vehicles. Article in Press. Accid. Anal. Prev.
Papadimitriou, E., Yannis, G., Bijleveld, F., Cardoso, J.L., 2013. Exposure data and risk indicators for safety performance assessment in Europe. Accid. Anal. Prev. 60, 371–383.
Parker, D., Manstead, A.S.R., Stradling, S.G., Reason, J.T., Baxter, J.S., 1992. Intention to commit driving violations: an application of the theory of planned behavior. J. Appl. Psychol. 77 (1), 94–101.
Parkinson H.J., Bamford G.J. (2017) Big Data and the Virtuous Circle of Railway Digitization. In: Angelov P., Manolopoulos Y., Iliadis L., Roy A., Vellasco M. (eds) Advances in Big Data. INNS 2016. Advances in Intelligent Systems and Computing, vol. 529. Springer, Cham.
Patriarca, R., Di Gravio, G., Costantino, F., 2017. Resilience engineering to assess risks for the air traffic management system: a new systemic method. Int. J. Reliab. Saf. 10 (4).
Provan, D.J., Woods, D.D., Dekker, S.W.A., Rae, A.J., 2020. (2020) Safety II professionals: how resilience engineering can transform safety practice. Reliab. Eng. Syst. Saf. 195, 106740.
Saward, J.R.E., Stanton, N.A., 2018. Individual latent error detection: simply stop, look and listen. Saf. Sci. 101 (2018), 305–312.
Schneider IV, W.H., Savolainen, P.T., Moore, D.N., 2010. Effects of horizontal curvature on single-vehicle motorcycle crashes along rural two-lane highways. Transp. Res. Rec. 2194 (1), 91–98.
Songchitruska, P., Tarko, P., 2006. The extreme value theory approach to safety estimation. Accid. Anal. Prev. 38, 811–822.
Shinar, D., Schechtman, E., Compton, R., 2001. Self-reports of safe driving behaviors in relationship to sex, age, education and income in the US adult driving population. Accid. Anal. Prev. 33 (1), 111–116.
Shorrock, S., Licu, T. (2014). Moving toward Safety-II: Ensuring things go right. The Ergonomist. March.
Smith, S.S., Horswill, M.S., Chambers, B., Wetton, M., 2009. Hazard perception in novice and experienced drivers: the effects of sleepiness. Accid. Anal. Prev. 41 (4), 729–733.
Stylianou K., Dimitriou L., Abdel-Aty M. (2019). Big Data and Road Safety: A Comprehensive Review. In: Antoniou C., Dimitriou L., Pereira F. (Eds), Mobility Patterns, Big Data and Transport Analytics. Elsevier.
Sujan, M.A., Huang, H., Braithwaite, J., 2017. Learning from incidents in health care: critique from a Safety-II perspective. Saf. Sci. 99, 115–121.
Summala, H. (1997). Hierarchical model of behavioural adaptation and traffic accidents. Traffic and transport psychology. Theory and application.

SWOV, 2018. Sustainable Safety 3rd edition – The Advanced Vision for 2018–2030 - Principles for Design and Organization of a Casualty-Free Road Traffic System. SWOV, The Hague.

Tejada, A., Manders, J., Snijders, R., Paardekooper, J. P., & de Hair-Buijssen, S. (2020, September). Towards a Characterization of Safe Driving Behavior for Automated Vehicles Based on Models of "Typical" Human Driving Behavior. In 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC) (pp. 1-6). IEEE.

Tselentis, D.I., Yannis, G., Vlahogianni, E.I., 2017. Innovative motor insurance schemes: a review of current practices and emerging challenges. Accid. Anal. Prev. 98, 139–148.

Tselentis, D.I., Vlahogianni, E.I., Yannis, G., 2019. Driving safety efficiency benchmarking using smartphone data. Transp. Res. Part C: Emerg. Technol. 109, 343–357.

Uluk, D., Lindner, T., Palmowski, Y., Garritzmann, C., Göncz, E., Dahne, M., Möckel, M., Gerlach, U.A., 2020. E-scooter: erste Erkenntnisse über Unfallursachen und VerletzungsmusterE-scooter: initial knowledge about causes of accidents and injury patterns. Notfall+ Rettungsmedizin 23 (4), 293–298.

Wahl, A., Kongsvik, T., Antonsen, S., 2020. Balancing Safety I and Safety II: Learning to manage performance variability at sea using simulator-based training. Reliab. Eng. Syst. Saf. 195 (2020), 106698.

Walker, G., 2017. Redefining the incidents to learn from: Safety science insights acquired on the journey from black boxes to Flight Data Monitoring. Saf. Sci. 99 (2017), 14–22.

Wang, F., Tian, J., Lin, Z., 2020. Empirical study of gap and correlation between philosophies Safety-I and Safety-II: a case of Beijing taxi service system. Appl. Ergon. 82 (2020), 102952.

Wilde, G.J.S., 1982. The theory of risk homeostasis: implications for safety and health. Risk Anal. 2 (4), 209–225.