12 Improving Safety by Learning from Automation in Transport Systems with a Focus on Sensemaking and Meaningful Human Control

Å. S. Hoem
Norwegian University of Science and Technology

S. O. Johnsen
SINTEF

K. Fjørtoft and Ø. J. Rødseth
SINTEF Ocean

G. Jenssen and T. Moen
SINTEF

CONTENTS

INTRODUCTION

There is an increase in the use of automation and autonomous solutions within transportation. According to *The Oxford Dictionaries*, autonomy is the right or condition of self-government, and the freedom from external control or influence. Many researchers (Relling et al., 2018) have discussed that the term is used differently in colloquial language than in the technical definition and that it is interpreted in different ways across industries. In this chapter, we emphasise that autonomy does not necessarily mean absence of human interaction. Often there is a strong need to design how humans can make sense of automation failures and enact meaningful human control.

Automated systems operate by clear repeatable rules based on unambiguous sensed data. An autonomous system can be a set of automated tasks, with interactions with several sub-systems and/or humans, with a specific degree/level of autonomy. Autonomous systems obtain data about the unstructured world around them, process the data to generate information and generate alternatives and make decisions in the face of uncertainty. Systems are not necessarily either fully automated or fully autonomous but often fall somewhere in between (Cummings, 2019). For example, transportation can have different modes during a sea voyage. Outside the harbour, in heavy traffic, it can be closely operated either by the remote control centre (RCC) or a captain/driver, while in open waters with low traffic it can be controlled by the computers or the autonomous system. Within the road traffic segment, the Society of Automotive Engineers (SAE) has defined a taxonomy on the levels of automation describing the expectations between automated systems and the human operator (SAE, 2018). This is summarised in Table 12.1 below.

The levels apply to the driving automation feature(s) that are engaged in any given instance of operation of an equipped vehicle. As such, a vehicle may be equipped with a driving automation system that is capable of delivering multiple driving automation features that perform at different levels. The level of driving automation exhibited in any given instance is determined by the feature(s) that are engaged (SAE, 2018). Hence, autonomy is different across application areas; it varies over time and is affected by the context.

To get a better overview and understanding, we start by looking at experiences gained from ongoing research and/or industry projects in the four transportation

TABLE 12.1
Levels of Automation – Simplified Description from SAE J3016 (2018).

LoA	Humans in control	Automation in control	Examples of automated features
0: No driving automation	All operations	No automated task. Warns; protect	Blind-spot monitoring and lane-departure warning
1: Driver assistance	All operations	Single automated systems: assists	Adaptive cruise control (ACC)
2: Limited assist; auto throttle	Drives in-the-loop	Guides	Automated lane centring combined with ACC
3: Assist, tactical; supervised	On-the loop human monitors all time	Manage movement within defined limits	"Traffic jam chauffeur"
4: Automated assist strategic	Out-of-loop asked by system	Operates, but may give back control	Self-driving mode with geofencing
5: Autonomous	Completely out-of-loop	Operates with graceful degradation	None are yet available to the general public

domains: road, sea, rail and air. Through these case studies, we aim to explore safety, security, sensemaking and the human control of autonomous transport systems. We have adopted the term "meaningful human control" from discussion and debates from another area (lethal autonomous weapon systems; Cummings, 2019). The term addresses the concerns of a "responsibility gap" for harms caused by these systems, i.e. humans, not computers and their algorithms should ultimately remain in control of, and thus morally responsible for, relevant decisions about military operations. The same concern must be the result of autonomous systems in transportation, i.e. humans (supported by computers and algorithms) should ultimately remain in control and responsible for relevant decisions. The responsibility may be on the designer and producer of the autonomous systems, as Volvo and Mercedes Benz have stated for their autonomous cars (Chinen, 2019, p. 109).

BACKGROUND: SAFETY OF AUTONOMOUS SYSTEMS

Safety is commonly defined as freedom from unacceptable risk (Hollnagel et al., 2008). For autonomous transportation to become a success, It must prove to be at least as safe and reliable as today's transport systems. By some, it is claimed that increased safety will be achieved by reducing the likelihood of human error when introducing more autonomy (Ramos et al., 2018). However, autonomy may create new types of accidents that before were averted by the human in control, as demonstrated by the Tesla fatal accident with Joshua Brown, NTSB (2017). Besides, the introduction of new technology will create new accident types, as explained by Porathe et al. (2018), Teoh and Kidd (2017), and Endsley (2019). The main safety challenges for autonomous systems are unexpected incidents not foreseen by automation, cybersecurity

threats, technological changes (with increased complexity and couplings), poor sensemaking, lower possibility for meaningful human control (Human not in the loop) and limited learning from accidents.

The term "Human in the loop" means that the human is a part of the control loop, i.e. that the human receives information and can influence other parts of the chain of events (Horowitz and Scharre, 2015). A key issue is the ability of the actors to make sense of the situation. In our study, we define sensemaking in a pragmatic context as a continuous process of interpreting cues to establish situational awareness in a social context, as described in Kilskar et al. (2020).

When trying to scope risks of autonomous systems, we must include regulation, risk governance, organisational framework, interfaces between humans and the autonomous system, and the available infrastructure (software components and cyber-physical systems) to build a sense of the situation for humans and the automated system (Johnsen et al., 2019).

Autonomous systems are socio-technological systems. Hence, a holistic approach is necessary, rather than a reductionist approach looking at the system as isolated processes and components. We lack statistical evidence for the probability of accidents with autonomous transportation systems. However, several actors have started pilots with different levels of autonomy within different transport modes. There is a need to collect and systemise experiences from these. The following sections present a review of experiences from different transport modes. The main objective has been to gather experiences and status on different transport domains and to learn between the modes, by asking the following research questions:

1. What are the major safety and security challenges of autonomous industrial transport systems?
2. What can the various transport modes learn from each other regarding safety and security related to sensemaking and meaningful human control?
3. What are the suggested key measures related to organisational, technical and human issues?

FINDINGS

Autonomy at Sea

Several countries have developed test areas for testing Maritime Autonomous Surface Ships (MASS). The International Maritime Organisation (IMO) currently uses the term MASS for any vessel that falls under provisions of IMO instruments and which exhibits a level of automation that is currently not recognised under existing instruments. There are already several small-size unmanned and autonomous maritime crafts which have been engaged in surface navigation, scientific activities, underwater operations and specific military activities.

In Norway, three national testing areas have been established, with supporting infrastructure, with the aim to test out MASS in the same area as conventional ships. Norwegian Forum for Autonomous Ships (NFAS, 2020) is a network established for sharing experiences and research within the subject of autonomous ships, with

the International Network for Autonomous Ships (INAS, 2020) as an extension of NFAS outside Norway. The research centre for Autonomous Marine Operations and Systems (AMOS, 2020) at NTNU was established in 2013 as a multidisciplinary centre for autonomous marine operations and control systems.

More extensive research projects, such as AAWA (2020), MUNIN (2020), Autosea (2020), Autoship (2020) and IMAT (2020), focus on specific concepts where unmanned, autonomous or smart ships are explored and tested. The world's first fully electric and autonomous container ship, Yara Birkeland (2020), is under construction. The ship is now planned to be in operation by 2022, earlier planned to start in 2020, and centres are scheduled to handle all aspects of remote and autonomous operation to ensure safety.

A newly established company, Zeabuz (2020), will test prototypes of an autonomous electric ferry system for urban waterways. Limited information is given about the concept other than it will be self-driving and electric. The remote and autonomous operational aspect of an RCC is not mentioned, but a remote support center is planned to operate in the initial phase.

Most of the projects above are in the initial stages with limited operational experience. Most safety concerns are related to the reliability of sensors and technical equipment and their ability to handle different situations.

Experiences Related to Safety Challenges

In operation, MASS have only been tested in small scale without an interface for human supervision or control. We have examples of safety issues during early testing of autonomous technology (software and hardware) local in Norway in Trondheimsfjorden, with the small-scale version of the passenger ferry *AutoFerry*. One example is loss of control due to a technical failure, a so-called fallout, of the dynamic positioning system which made *AutoFerry* run into the harbour. However, there is no systematic data collection of failures or unforeseen events, and this is not a requirement from the Norwegian Maritime Authority (NMA) at present. Though, a Preliminary Hazard Analysis (PHA) has been carried out for the operation of the *AutoFerry* (Thieme et al., 2019), the main hazards were software failure; failure of internal and external communication systems; traffic in the channel (especially kayaks, difficult to discover); passenger handling and monitoring; and weather conditions. The practical challenges encountered in the ferry project were also listed. These challenges are related to available risk analysis methods and data, determining and establishing an equivalent safety level, and some of the prescriptive regulations currently in use by NMA. At present (start 2021) the *AutoFerry* project lacks an established plan on who should operate the ferry and how to intervene especially during emergencies. The human operator is said to be in the loop and able to intervene from an RCC. However, none of the projects have developed such a centre or made detailed plans for their operation so far. In the reviewed projects, the focus has been on technology development.

A literature review on risk identification methods for MASS (Hoem, 2019) identifies the uncertainty of the operational mode and context of the MASS operation (i.e. operational domain) to be a major challenge when identifying operational hazards and risks. There is a need to define what conditions the ship is designed to operate under. Rødseth (2018) proposed to use the "operational design domain" from SAE J3016 (2018) to define the context, i.e. the operational domain with its complexity.

This term is further described as an operational envelope (Fjørtoft and Rødseth, 2020). An operational envelope defines precisely what situation the MASS must be able to handle by assigning responsibilities to the human operators and the automation. It defines conditions of operations, describes the characteristics and requirements of the system and enables the design of Human–Autonomy Interface (HAI), based on specific task analysis, safety-critical tasks and challenges of sensemaking.

Several different guidelines are developed for autonomous shipping. IMO has published an Interim Guideline for MASS trials which aims to assist authorities and relevant stakeholders to perform autonomous tests. It includes risk management, how to comply with existing rules and regulations, safe manning, the human element and HMI, infrastructure, trial awareness, and communication and information sharing.

Lessons Learned from Autonomy at Sea

Based on the preliminary testing and risk analysis, it is evident that MASS is a system of systems, depending on local sensor systems, automated port services, communication with RCC, other autonomous ships, conventional ships, Vessel Traffic Centres (VTS) and similar. These interactions are critical factors and should be addressed in design and operations. The degree of autonomy varies and is affected by the complexity of the operation. A MASS will operate in phases with transitions between human control and automation control. A well-defined operational envelope is key for addressing safety issues and carrying out a risk assessment. Potential hazards within each transition must be identified with fallback procedures in place, with focus on the sensemaking process and how humans should enter the control loop.

Challenges related to communicating the intent of a MASS in interactions between autonomous, unmanned ships and manned ships are addressed by Porathe (2019). The authors argue for "automation transparency" and methods allowing other seafarers to "look into the mind" of the autonomous ship, to see if they themselves are detected, and the present intentions of the MASS, i.e. sensemaking among all actors. This can be done by sharing information about the intention, what the automation knows about its surroundings, what other vessels are observed by its sensors and similar by a live chart screen accessible on-line through a web portal by other vessels, VTS, coastguard, etc. Such a common system could be the responsibility of the VTS and should be specified as a requirement for the operational design domain and the operational envelope.

In a guideline from the Bureau Veritas (2019), several hazards are listed as important: voyage, navigation, object detection, communication, ship integrity, machinery and related to systems, cargo and passenger management, remote control and security. Within each of them, a list of factors is mentioned. Using this, Hoem et al. (2019) identified a list of hazards comparing autonomous and manned ships. The scenarios were focussed on the following differentiating factors: fully unmanned, constrained autonomy, RCC, higher technical resilience and improved voyage planning. The paper gave a draft attempt to classify risk factors that can either be characterised as new types of incidents caused by technology, what is most characterised in regard to today's incidents in shipping and if the incidents are averted by crew today. As an example, the category fully unmanned points to a higher risk for technical failure but may improve some of today's operators' errors caused by poor design and

lack of good human factor engineering practice. Important factors moving forward are robust sensor quality, redundancy on key technology and good education for land-based operators that support sensemaking and build situational awareness. It is likely that humans are not continuously monitoring one vessel at a time but will be needed to supervise and intervene when necessary. For a constrained autonomous vessel, the paper pointed to the need for better HAI due to the need of time to support sensemaking and get situational awareness before action.

Autonomy in Air

Automation and autonomy in aviation have been implemented since World War II, where functions have been systematically automated and the manning has been systematically reduced. Incidents due to automation happen, but aviation safety (commercial passenger traffic) is extremely high.

In addition to increased automation in manned flights, the use of drones or unmanned aerial systems (UAS) has risen significantly in the last years. Examples of use are:

- Photography and video recording to support information and crisis management
- Inspection of (critical) components to improve safety, avoid human exposure, reduce costs or improve quality
- Detection and survey of environmental issues, such as gas emissions, ice detection in sea, overview and control of pollution
- Logistics – delivery of critical components or supplies (such as medicine, blood)

Safety Challenges

Manned flights have a high level of safety, issues have often been a result of poor sensemaking and poor situational awareness of the crew. The reliability of the technical equipment is high. Automation accidents have happened lately where guidelines during design and certification have not been followed. This was the case in the Boeing 737 MAX fatal crashes (Cruz and de Oliveira Dias, 2020). After analysing the accidents, Endsley (2019) recommended ensuring compliance with human factors design standards and support for human factors assessment in aircraft testing and certification.

Safety challenges in UAS differ from the challenges in manned operations, due to the immaturity of technology. Looking at the use of large drones in the US, Waraich et al. (2013) documents that mishaps may happen more frequently (i.e. 50–100 mishaps occur every 100,000 flight hours vs human-operated aircraft where there is one mishap per 100,000 flight hours). The mishap rate is 100 times higher in UAS remotely piloted than in manned operations. The leading causes are poor attention to human factors science, such as poor design of human machine interfaces in ground control centres (Waraich et al., 2013; Hobbes et al., 2014).

In Petritoli et al. (2017), the mean time between failures (MTBF) estimated for UAS was around 1,000 hours, approximately 100 times higher than MTBF in manned

flights. The dominant failures were in power systems, ground control system and navigation systems.

The risks of UAS operations are dependent on the operational domain, i.e. the type of operation (delivery, data collection, surveillance, inspection photography, etc.) and physical details of the drone such as weight, speed and height of operation. EASA (2016) has estimated the probability of fatality of different UAS weights and estimated probability of fatality as 1% with a UAS weight of 250 g, but 50% fatality with a weight of 600 g in case of a collision with a human when the drone drops.

Examples of undesired incidents from UAS are: collisions with personnel; interference with infrastructure (infrastructure such as airports is vulnerable and interference may lead to disruption of air traffic); actual damage to critical infastructure; damage to the drone; using the drone to spy or steal data (leading to loss of privacy, data theft and possible emotional consequences). Automated systems and UAS are vulnerable to attacks through the cyber-physical systems it consists of, such as sensors, actuators, communication links and ground control systems. As an example, an Iranian cyber warfare unit was able to land a US UAS based on a spoofing attack modifying the GPS data (Altawy et al., 2017).

There are several challenges of UAS operations in challenging climatic conditions such as low temperature, wind, winter with sleet and snow. Operational equipment may not be tested or hardened for these demanding conditions; thus, requirements, testing and certification are needed. Communication infrastructure is also demanding in the north, from 70° the quality of satellite communication is degraded. GPS spoofing may be a challenge and must be mitigated.

Lessons Learned That May Be Transferred

Automation in aviation has succeeded in establishing a high level of safety, due to systematically automating simple tasks and reducing demands on the pilot: base development on the science of human factors, building infrastructure, to control and support flights, strong focus on learning from small incidents and accidents and support from control centres that have strict control of the operational domain/operational envelope. Thus, systematic development and stepwise refinement has had a huge success in terms of safety and trust, in addition to the strong focus on keeping the human in the loop supported by sensemaking. Even in this environment of high reliability, there is a strong need to ensure compliance with human factors design standards and support for human factors assessment in aircraft testing and certification to avoid fatalities by automation as seen in the Boeing 737 Max accidents.

The reliability of drones is lower than for manned planes, and there is a need to develop improved reliability of the new technology. Systematic risk assessment is needed to mitigate the areas with the most risks. The HMI between automation and the human operator is challenging. Design must use best human factors practices to support sensemaking and ensure that the operator can intervene and take control when needed.

Autonomy in Rail

By automated metros (rail systems), we mean systems where there is no driver in the front cabin, nor accompanying staff, also called Unattended Train Operation

(UTO). UTO has been in operations since 1980. According to UITP (2013), there is 674 km of automated metros consisting of 48 lines in 32 cities. Examples of cities with UTOs are Barcelona, Copenhagen, Dubai, Kobe, Lille, Nuremberg, Paris, Singapore, Taipei, Tokyo, Toulouse and Vancouver. There is large infrastructure cost to ensure safe on and offloading of passengers and that the track is isolated from other traffic. Four distinct levels of automation are defined:

GoA1: Non-automated train operation, with a driver in the cabin.
GoA2: Automatic train operation system controls train movements, but a driver in the cabin observes and stops the train in case of a hazardous situation.
GoA3: No driver in the cabin but an operation staff on board.
GoA4: Unattended train operation, with no operation staff on board.

Safety Challenges

Wang et al. (2016) list the following as arguments for UTO: increased reliability, lower operation costs, increased capacity, energy efficiency and an impressive safety record. We have at present not found normalised accident data for UTO (incidents based on person km), and no accidents have been reported. We have found reports in newspapers about minor incidents, without any fatalities reported. Based on data and experiences so far, it seems that the UTO has exceptionally high safety. However, more systematic analysis and normalisation of all international UTO transport incidents are needed.

Even though driverless trains have an impressive safety record, experience shows that they still face some challenges related to reliability and operability. One example of this is seen in Singapore. UTOs were introduced in Singapore's Mass Rapid Transits (MRT) system in 2003. Here, the operations were monitored remotely from an operations control centre. However, in 2018, most of these trains were manned again, for improving reliability. Some of the trains experienced technical issues and failures. In these cases, a driver on board a train will immediately be able to assess the problem, and, if necessary, push another disabled train out of the way. With a driverless system, a driver had to make his way to the unmanned train, which takes time. Nevertheless, the safety record of driverless trains is impressive, maybe due to the rail track as a system. Hence, further automation of railway systems is ongoing.

Lesson Learned

As mentioned, it seems that the UTO has an exceptionally high level of safety. However, systematic analysis and normalisation of all international UTO transport incidents are needed. Thus, there is a need for systematic reporting and analysis of minor incidents/small accidents in order to support risk-based regulation and risk-based design of the technology.

A key issue related to safety is the focus on a restricted design domain and operational envelope. The environment/context of which the UTOs operates is typically underground, with few or no interaction with other traffic. Protection systems are in place at the embarkment area/platform preventing the most common incidents (people falling on tracks). There has been a focus on analysing personnel incidents when entering and leaving the UTOs and building safer infrastructure to minimise dangerous situations.

Autonomy on Road

Cities worldwide are increasingly testing and implementing autonomy as the pace of autonomous vehicle innovation picks up. Norway has long-term experiences of autonomous transport systems such as Automated Guided Vehicles (AGVs) at St. Olav Hospital and autonomous shuttle buses used from January 2018 on public roads.

Projects with autonomous vehicles (AVs): Local governments must approve self-driving pilots. In the US, in California, all companies must deliver annual self-reports on incidents with highly automated vehicles. (This is one of the reasons why Uber and many other companies moved the testing of self-driving taxis to Arizona that has adopted a more liberal attitude.) This framework condition, i.e. legislation in California, has enabled the industry to document the level of safety and identify challenges.

Related to the present development trends, there are two clear trends that are different in nature:

1. a race to develop fully AVs, i.e. self-driving cars, aiming to replace today's private cars.
2. an effort to develop fully AVs to provide mobility-as-a-service (MAAS) or robotaxis.

The aim of the private self-driving car segment is to operate more safely than human drivers are able to in real-world conditions and at high speed. Here, the self-driving cars must be able to handle all types of obstacles and interactions with other road users in all kinds of weather and traffic conditions.

The MAAS segment focusses on small shuttle buses (or robotaxis) with geofencing to establish a safe route. Many of these are unable to go around an obstacle. They stop until the obstacle has moved or been removed. They operate at low speeds between 12 and 30 km/h.

There are many projects with self-driving vehicles on public roads operating around the world. According to Philantropies (2017), at least 53 cities are currently involved in testing AVs. Legal frameworks for the regulation of pilot testing are established in Singapore, the Netherlands, Norway and the UK (KMPG, 2018). Euro NCAP has designed a set of test procedures for testing automated vehicles on SAE level 2. The US Department of Transportation has developed a framework (NHTSA, 2018) for testing automated driving systems focussing on failure behaviour, failure mitigation strategies and fail-safe mechanisms.

AGVs at St. Olav Hospital have been in operation since 2006. Today, 21 AGVs operate at a speed of approximately 2 km/h (max speed is 5 km/h) and communicate with each other, open doors and reserve elevators. The automation is quite simple as they follow a predefined path, and when there are conflicts or problems with collisions/doors/elevators, a signal is given to the operational centre, always manned by an operator who can intervene or go to the place. Manned operators in the centre are necessary to ensure continuous operations. Even in this strict operational envelope, humans are critical components in the loop. Sensemaking has been in focus, examples are that the AGVs are "speaking" to hindrances/people – saying "please move" or "this elevator is reserved".

Pilots with autonomous shuttle buses: From 2017, testing of AVs was allowed in Norway. In the SmartFeeder (2019) research project, initial data are gathered from five test sites with MAAS pilots. Each pilot tests self-driving shuttle buses carrying up to six passengers, operating at an average speed of 15 km/h, and with an operator to monitor and take over control if necessary (during the test phase). These pilots are "fixed route autonomy", where the autonomous system follows a predefined route and processes a limited amount of sensor data along the route. The motivation varies, i.e. solving a last mile problem (connecting workplaces with public transportation), testing out technology and user acceptance or property and business development. In total, the buses in the pilots have driven almost 22,000 km, with approximately 40,500 passengers in both summer and winter conditions. Initial data have been collected regarding disengagement of the system and involvement of the operator in the pilots in three categories: "obstacle emergency stop" (sensors detect something and automatically stop), "soft stop" (operator overtakes system and decelerates the vehicle) and "Manual switch" (for manually driving the vehicle). The collected data are currently being processed and cleaned for more detailed analysis, and interpretations cannot be drawn yet. However, the reliability and robustness are challenging, and demands a restricted operating envelope in addition to the need for "humans in the loop" when the unanticipated is happening.

Safety Challenges

Tesla with its autopilot has enabled automated driving at high speeds. Several severe accidents with Tesla autopilot have led Tesla to limit their autopilot functionality. These partially automated vehicle systems at SAE level 2 (SAE, 2018) always operate exclusively based on an attentive driver being able to control the vehicle. For fully automated driving (SAE level 4–5), the driver is no longer available as a backup for the technical limits and failures. Replacing human action and responsibility with automation raises questions of technical, ethical and legal risks, as well as product safety.

As far as we know from media and public accident reports there have been four fatal accidents worldwide: three with semi-automated (SAE level 2) autopilot and one with a more fully automated vehicle on public roads (SAE level 3), the Uber accident in Arizona where a Volvo refitted with Uber self-driving technology killed a pedestrian (NTSB, 2018). In all cases, the autopilot was engaged but without driver interaction or intervention with vehicle controls, highlighting the need for sensemaking and "meaningful human control".

There are few safety records (data) on SAE level 4 so far. Data from 2009 to the end of 2015 collected by Google's cars list three police reportable accidents in California while driving at 2,208,199 km (Teoh and Kidd, 2017). This is 1/3 of reportable accidents per km of human-driven passenger vehicles in the same area. In 2017, 19 of 21 reported accidents with Google-Waymo cars (level 4) were rear-ended accidents at signalised intersections. This is caused by ordinary drivers' misinterpretation of automated vehicle behaviour (as an example expecting that drivers are not halting when meeting a yellow light at an intersection.). Google-Waymo has now patented a software program allowing their vehicles to drive through yellow light. A look at accidents and incidents reported to the California Department of Motor

Vehicles (DMV) in 2019 shows that other 65 companies currently testing level 4 technology still have frequent rear-end collisions at signalised junctions. They also have trouble (and reported accidents) entering a motorway from the ramp. AVs have not yet learned the "nudging" that ordinary drivers do to see if traffic on the motorway yield and let you in.

Experience from the autonomous shuttle buses: For the pilots, it was mandatory to report incidents and accidents. No persons were injured, and only minor technical issues and malfunctions were reported. The following issues were revealed:

- Snow, heavy rainfall and fog are challenging for the sensors.
- Vegetation and light poles along the route of the bus is challenging as they interfere and disturb the sensors at times.
- The buses run along the same "track" with narrow wheels, causing significant wear and tear on the road along this track.
- Cyclists passing near the bus makes the bus stop abruptly.

These issues are related to the predefined operational envelope surrounding the vehicle, leading to abrupt stops when violated. As pointed out by Jenssen et al. (2019), AVs lack a sense of self, and software and sensors are still not designed to account for the discrepancy in the same way human drivers are able to.

When applying for testing, a mandatory risk assessment was carried out. The main risks listed were related to passenger injury as a result of an abrupt stop where passengers inside the bus are unprepared and can be harmed by falling. Risk-reducing measures are lowering the speed, installing seat belts, limiting the number of passengers and adding road signs.

AGVs at St Olav: A total of 100–130 minor incidents per year have been reported. Yearly, each AGV experiences around 15 emergency stops (Johnsen et al. 2019), where components must be changed. Reported incidents are minor crashes as a consequence of faulty navigation due to objects placed in the route, summarised in Johnsen et al. (2019). From interviews with the operators of the AGVs, the following main issues are identified:

- The AGVs ability to adapt to the surrounding infrastructure
- Keep the track of the AGVs clear of objects
- Make objects visible to the AGV: the AGVs are not able to detect all obstacles due to the sensor range
- Establish a control room with proper HMI design
- Maintain the interface to cyber physical systems: software updates has led to problems (due to poor testing and multiple vendors.)

Lessons Learned

Vehicle automation can enhance safety but also introduces new risks due to poor technical implementation and the need for rapid response from the human actor. This is especially the case with SAE automation levels 2 and 3.

The accident data collected so far with automation (AGVs and level 1–4 vehicles) indicate safety hazards of human factors and technical issues, i.e. obstacle detection

(sensors), programming (rule-based and not artificial intelligence, AI), prolonged attention (humans in the loop), HMI (Autopilot-engagement rules) and misuse. The list may become longer as more safety data are gathered and more in-depth information on accident causality of automated vehicles is established, e.g. overreliance and expectation mismatch.

Based on the experiences, there is a need to establish regulations that ensure systematic incident reporting, develop systems based on learning from incidents and invest in infrastructure to support automation, i.e. help the automation by focussing on an operational envelope that uses more data from infrastructure. The transport systems are automated but not autonomous. Autonomous systems are immature at present and must be further developed.

A SUMMARY OF MTO SAFETY ISSUES

Based on the performed reviews, the suggested key measures are listed below.

Humans: As seen from all experiences, the uncertain and complex environment for autonomous systems must ensure the need for human intervention. Autonomous transportation systems will to a varying degree need human control if failures occur or under certain operational conditions. With today's UTOs and AGVs, an operator is still needed when there is a disruption and sensors fail to detect and recognise an obstacle or determine the next actions. However, in testing and developing autonomous transportation systems with drones, AVs and vessels, we see examples of projects where the human operator is not considered from the beginning. The industries' motivation seems to be to try to automate as much as possible and assume that humans will and can monitor it. Hence, HAI and how to keep the humans in the loop is often considered a challenge to be solved late in the project after knowing the limitations of the technology and by considering the humans as the adapting back-up. Most of the projects lack early incorporation of human factors in analysis, design, testing and certification process. Thus, there are costly challenges that should have been addressed earlier by starting with technology, human limitations and possibilities, and organisational and infrastructure needs. A key issue is to define the design conditions the system should operate under by defining the operational envelope and critical scenarios (such as sensor failures). Then specify how critical scenarios can be mitigated by infrastructure support i.e. surrounding systems such as other autonomous systems nearby (cars) or control infrastructure. If human intervention is needed to handle the scenarios, sensemaking must be supported within the existing limitation of human abilities.

As aviation is the industry with the most experience with safe automated systems, the list from Endsley (2019) with design principles for improving people's ability to successfully oversee and interact with automated systems should be a very useful element, allowing for manual overrides and sufficient training to users on automation to ensure adequate understanding and appropriate levels of trust.

Technology: To date, developing autonomous or remotely controlled transportation systems (especially for AVs and MASS) appears to primarily be about a technology push rather than considering and providing sociotechnical solutions including redesign of work, capturing knowledge and addressing human factors as we and others have seen (Lutzhoft et al., 2019).

Technology in autonomous systems and their interpretation (such as through AI) are not reliable at present – thus, there is a need to address poor reliability trough improving man/technology/organisation aspects. The reliability of drones is lower than for manned planes, and we have seen how sensors and technical equipment are causing safety issues in several projects. The systems must improve for an industrial setting and for safety-critical operations, i.e. become highly reliable and resilient to bad data and have automatic self-checking behaviour and avoiding single-point failures by checking across multiple inputs. Thus, there is a need to get support from other AVs with sensors, need for developing infrastructure (such as roads and seaways with sensors), in addition to establishment of control centres for road traffic and maritime traffic that must be responsible for supporting sensemaking among the actors (i.e. automated and not automated systems). Technical barriers must be in place to a larger extent on autonomous systems to avoid and reduce the outcome of failures and component interaction accidents, which are more common as the complexity increases.

Automation transparency is important for both sharing the situation awareness and communicating the intentions towards others and for the operator in an RCC to understand the behaviour of the automation. In complex systems, a wide range of alarm issues related to diagnostics, management and assessments of multiple input data will be challenging. Hence, alarms must be unambiguous and displayed with a clear message. This requires good human factor engineering practice, such as an alarm philosophy and relevant standards.

Organisation: Experience from the projects and pilots demonstrate a need to see the technological solution in a larger sociotechnical context. Autonomous transportation systems are a system of systems. We have seen that legislation is are needed to gather data and establish the operational context. There is a need for substantial investments in infrastructure: organisational interfaces are lacking and organisational/structural issues from the operator/company/area/society are often considered the last thing to get in place. Looking at the operational context, we have seen a need to limit the operational design domain and use operational envelopes, or safety envelopes to define situations, responsibilities and system characteristics during all conditions (especially in safety-critical conditions with sensor/data failures). Regulations and guidelines have slowly been established to support autonomous transportation systems. However, few of them require systematic reporting of accidents and incidents. Experience from accidents with AVs has given valuable insight, and hence all domains should prioritise and require reporting and systematic data collection of failures, hazards and unforeseen events. Not requiring reporting and sharing of safety-critical systems is a risk in itself.

SENSEMAKING TO SUPPORT MEANINGFUL HUMAN CONTROL

Focus on the design of operational envelopes to reduce complexity and analysing the needs for cues and information to support sensemaking and meaningful human control, when needed, is a key issue. Defining operational envelopes answers the question of which functions and roles automation/autonomy should have, versus

humans, when designing a complex system. This is also an important question for certification of the autonomous transportation system.

Sensemaking and the principle of meaningful human control should be used to verify that the proper functions are allocated to the human or the automation. According to Santoni de Sio and van der Hoven (2019), two design requirements should be satisfied for an autonomous system to remain under meaningful human control:

1. A "tracing" condition, according to which the system should be designed in such a way as to grant the possibility to always trace back the outcome of its operations to at least one human along the chain of design and operation.
2. A "tracking" condition, according to which the system should be able to respond to both the relevant moral reasons of the humans designing and deploying the system and the relevant facts in the environment in which the system operates.

From a safety perspective, this can be placed in the bowtie model, where the design principle of tracking are barriers preventing a technical fault, threat or unexpected situation to lead to a dangerous situation, as a human alway has established the possibility to intervene and take over control. On the other side of the bowtie, once a hazard has emerged, the outcome can be reduced by designing after a tracing condition making it possible to trace back the operation to a human who is in the position to understand the capabilities of the system and the possible effects in the world of its use and, hence, knows how to limit the consequences of an undesired event.

CONCLUSION

We have given a summary of ongoing projects and safety issues. The main issues across the domains are technical reliability and maturity, the need for automation transparency (including awareness for the decision made by automation), the need for defining what conditions the system can operate under and assigning responsibilities to human operators and the automation. Experiences from known accidents involving a high level of automation, as in the cases of Boeing 737 MAX, Uber and Tesla, have shown overreliance on automation and poor understanding of capabilities and limitations. We need to collect and systemise data on accidents and incidents of autonomous transportation systems and design with human factor practice to support sensemaking and meaningful human control.

Design principles from meaningful human control should be used to verify if the interaction between automation and the human is safe. This can be used as an input to operational envelopes and to assist in the design of a good HAI supporting sensemaking.

ACKNOWLEDGEMENT

This chapter has been funded by the Norwegian Research Council – project 267860 SAREPTA.

REFERENCES

AAWA (2020). https://www.rolls-royce.com/media/press-releases/2016/pr-12-04-2016-aawa-project-introduces-projects-first-commercial-operators.aspx

AMOS (2020). https://www.ntnu.edu/amos/research

Autosea (2020). https://www.ntnu.edu/autosea

Autoship (2020). https://www.kongsberg.com/maritime/about-us/news-and-media/news-archive/2020/autoship-programme/

Bureau Veritas (2019). NI 641 R01 *Guidelines for Smart Shipping*.

Chinen, M. (2019). Law and Autonomous Machines. *Elgar Law, Technology and Society* (p. 109). Edward Elgar Publishing.

Cruz, B. S., & de Oliveira Dias, M. (2020). Crashed Boeing 737-MAX: Fatalities or malpractice? *GSJ* 8 (1), 2615–2624.

Cummings, M. L. (2019). *Lethal Autonomous Weapons: Meaningful human control or meaningful human certification?* IEEE Technology and Society.

Endsley, M.R. (2019). *Human Factors & Aviation Safety* Testimony to the United States House of Representatives. Hearing on Boeing 737-Max8 Crashes, December 11, 2019.

Fjørtoft, K. E., & Rødseth, Ø. J. (2020). *Using the operational envelope to make autonomous ships safer* Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio.

Hoem, Å. S. (2019). The present and future of risk assessment of MASS: a literature review. *29th European Safety and Reliability Conference*. European Safety and Reliability Association.

Hoem, Å.S., Fjørtoft, K., & Rødseth, Ø. (2019): *TransNAV 2019: Addressing the Accidental Risks of Maritime Transportation: Could Autonomous Shipping Technology Improve the Statistics?*

Hollnagel, E., Nemeth, C. P., & Dekker, S. (Eds.). (2008). *Resilience engineering Perspectives: Remaining Sensitive to the Possibility of Failure* (Vol. 1). Ashgate Publishing, Ltd.

Horowitz, M., & Scharre, P. (2015). *An Introduction to Autonomy in Weapon Systems*. Center for a New American Security (CNAS) Working Paper (CNAS: Washington, DC), p. 8

IMAT (2020). https://www.sintef.no/projectweb/imat/

INAS (2020). http://www.autonomous-ship.org/index.html#H2

Johnsen, S. O., Hoem, Å., Jenssen, G., & Moen, T. (2019). Experiences of main risks and mitigation in autonomous transport systems. *Journal of Physics: Conference Series* 1357 (1) 012012.

Kilskar, S. S., Danielsen, B. E., & Johnsen, S. O. (2020). Sensemaking in critical situations and in relation to resilience—a review. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 6(1).

KMPG (2018). Autonomous vehicles readiness index. *Klynveld Peat Marwick Goerdeler* (KPMG) International.

Lutzhoft, M., Hynnekleiv, A., Earthy, J. V., & Petersen, E. S. (2019). Human-centred maritime autonomy-An ethnography of the future. *Journal of Physics: Conference Series* 1357 (1), 012032.

MUNIN (2020). http://www.unmanned-ship.org/munin/

NFAS (2020). http://nfas.autonomous-ship.org/index.html

NHTSA (2018). *A Framework for Automated Driving System Testable Cases and Scenarios*. DOT HS 812 623. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13882-automateddrivingsystems_092618_v1a_tag.pdf

NTSB (2017). National Transportation Safety Board 2017. Collision between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida, May 7, 2016. Highway Accident Report NTSB/HAR-17/02. Washington, DC.

NTSB (2018). *National Transportation Safety Board 2018*. Preliminary Report: Highway HWY18MH010.

Porathe, T. (2019). Interaction between Manned and Autonomous Ships: Automation Transparency. *Proceedings of the 1st International Conference on Maritime Autonomous Surface Ships*.

Porathe, T., Hoem, Å., Rødseth, Ø. J., Fjørtoft, K., & Johnsen, S.O. (2018). At least as Safe as Manned Shipping? Autonomous Shipping, Safety and "Human Error". *Proceedings of ESREL 2018*, June 17–21, 2018, Trondheim, Norway.

Ramos, M. A., Utne, I. B., Vinnem, J. E., & Mosleh, A. (2018). Accounting for Human Failure in Autonomous Ship Operations. Safety and Reliability–Safe Societies in a Changing World. *Proceedings of ESREL 2018*, June 17–21, 2018, Trondheim, Norway.

Relling, T., Lützhöft, M., Ostnes, R., & Hildre, H. P. (2018). A Human Perspective on Maritime Autonomy. *International Conference on Augmented Cognition* (pp. 350–362). Springer, Cham.

Rødseth, Ø. J. (2018). Defining Ship Autonomy by Characteristic Factors, *Proceedings of ICMASS 2019*, Busan, Korea, ISSN 2387–4287.

SAE International (2018). Standard, SAE J3016_201806. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Revised

Santoni de Sio, F., & Van den Hoven, J. (2018). *Meaningful human control over autonomous systems: a philosophical account*. *Frontiers in Robotics and AI* 5, 15.

SmartFeeder (2019). https://www.sintef.no/prosjekter/smart-feeder/ (in Norwegian)

Teoh, E. R., & Kidd, D. G. (2017). Rage against the machine? Google's self-driving cars versus human drivers. *Journal of Safety Research* 63, 57–60.

Thieme, C. A., Guo, C., Utne, I. B., & Haugen, S. (2019, October). Preliminary Hazard Analysis of a Small Harbour Passenger Ferry–Results, Challenges and Further Work. *Journal of Physics: Conference Series* 1357 (1), 012024).

UITP (2013). *Observatory of Automated Metros World Atlas Report*. International Association of Public Transport (UITP), Brussels

Wang, Y., Zhang, M., Ma, J., & Zhou, X. (2016). Survey on driverless train operation for urban rail transit systems. *Urban Rail Transit* 2, 106–113. https://doi.org/10.1007/s40864-016-0047-8

Yara Birkeland (2020). https://www.kongsberg.com/maritime/support/themes/autonomous-ship-project-key-facts-about-yara-birkeland/

Zeabuz (2020). https://zeabuz.com/