By what hubris? – the readiness of the human operator to take over when the automation fails or hands over control

Jerry Wachtel[1*]

[1] The Veridian Group, Inc. 567 Panoramic Way, Berkeley, CA, USA
[*]jerry@veridiangroup.com

Abstract: As Level 2 automated vehicles become pervasive in the traffic stream and as Levels 3 and 4 vehicles become increasingly common, automation failures and sudden handoffs due to coding errors, unanticipated events, or hacking will also increase. Despite some encouraging findings we argue that a non-trivial percent of drivers will be ill-equipped to handle such situations. We demonstrate that, in three highly technological industries with better prepared operators, better controlled working environments, and more rigorously designed and tested equipment, accidents and near misses (incidents) still often occur during automation failures and handoffs, as well as due to the operators' misunderstanding of the automation or the state of the equipment. We express our opinion that specialized driver training and/or "chatty" on-board interfaces may be potential solutions to this problem, and that there is little or no evidence that either of these methods is in use or contemplated in the field. Finally, we propose a thought experiment to test our hypothesis about the viability of these two approaches.

1. Introduction

The U.S. based Society of Automotive Engineers (SAE International) has identified six levels of vehicle automation, supplanting the previous listing by the U.S National Highway Traffic Safety Administration (NHTSA) [1]:

Level 0 - Zero autonomy – the driver performs all driving tasks.

Level 1 – Driver Assistance – the vehicle is controlled by the driver, but some driving assist features may be included in the vehicle design.

Level 2 – Partial Automation – vehicle has combined automated functions, like acceleration and steering, but the driver must remain engaged with the driving task and monitor the environment at all times.

Level 3 – Conditional Automation – driver is a necessity, but is not required to monitor the environment. The driver must be ready to take control of the vehicle at all times with notice.

Level 4 – High Automation – the vehicle is capable of performing all driving functions under certain conditions. The driver may have the option to control the vehicle.

Level 5 – Full Automation – the vehicle is capable of performing all driving functions under all conditions. The driver may have the option to control the vehicle.

Despite ongoing trials in several countries, fully automated vehicles are not likely to become commonplace on our roadways for many years to come. NHTSA estimates that the "highway autopilot" with "fully automated safety features" will become widely available from 2025 [2]. Others are far more pessimistic [3, 4].

If and when full automation becomes commonplace, it is widely agreed that it will bring about substantial benefits to society: in increased fuel economy, reduced air pollution, travel efficiencies, and, most of all crash reduction and injury prevention. Prior to that time, however, Level 2 automation has become increasingly common, and Level 3 automation is beginning to be introduced in a limited but growing number of high end production vehicles. Since we have years to go before Level 5 may be achieved on a widespread basis, human factors experts and vehicle designers are concentrating their attention on Levels 3 and 4. This paper argues that we introduce vehicles with such increasing levels of automation with considerable hubris, based on results from other industries and growing experience with such vehicles "on the streets."

By presenting exemplar accidents and incidents that have occurred with automation in other industries, and then comparing operations and operators in those industries to the automotive environment, we hope to point out why we believe that we are engaged in hubris, and then propose a thought experiment in an effort to address the major concerns that we see.

2. Automation Failures and Hand-offs in Other Industries

2.1 Aerospace – The Apollo 10 Anomaly

In April of 1969, the U.S. launched its final rehearsal space mission for the ultimate goal of landing the first man on the moon, which was to take place three months later. On this last rehearsal flight, identified as Apollo 10, the moon landing vehicle (called the Lunar Module, or LM), was to separate from the Command Module (CM) which remained in orbit some 60 miles above the lunar surface. The LM was then to descend to 10 miles above the surface, perform certain mission related objectives, and then fly back to complete a "rendezvous and docking" with the CM. By completing its activity, the LM would complete every step of the actual landing except the final descent and touchdown on the lunar surface. But upon ascent to rendezvous, while testing the Abort Guidance System (AGS) something went wrong. The mission Commander, who was flying the LM, complained that, when he put the Rendezvous Radar switch into the "Automatic" mode, the LM began to gyrate wildly. He quickly put the switch into the "Off" position to gather his wits. When he put it back into "Automatic" again, the spacecraft performed exactly the same way, and he was very confused and quite angry. It was only via instruction from flight controllers on the ground in Houston, Texas and Bethpage, New York, that he was able to put the switch into the "Attitude Hold" mode and fly the LM manually to achieve radar contact with the CM and, ultimately, to achieve a successful docking. The net result was the need to fly an additional lunar orbit, a very angry astronaut, a contentious flight debriefing, and a forced delay of the next flight, the first manned lunar landing, while the LM's manufacturer (Grumman Aerospace Corporation) undertook a very public and painful analysis of the "failure" as ordered by NASA. This analysis was all the more painful because both NASA and Grumman knew that there was no problem – the spacecraft had performed exactly as it was supposed to; the telemetry data proved it. But in those days astronauts were considered national heroes who could do no wrong. And although engineering and human factors staff at both organizations knew that the Commander had erred, that he had misunderstood how the "Automatic" function worked, and, as a result, placed the vehicle into an unintended flight mode, no one would call him on it, and so the Grumman team spent two months investigating a non-event at its customer's direction. In the end, the switch was fitted with a guard "to prevent inadvertent actuation." Some folks were mollified; most were not. But the program went on, and the manned lunar landing was successfully performed during Apollo 11.

The following is from the NASA Apollo 10 Mission Report [5]:

"… lunar module attitudes deviated from expected during the staging maneuver. Telemetry data indicated the automatic mode was engaged twice for short periods prior to and at staging. Since the automatic mode had been used previously to point the lunar module's Z-axis at the command module, the guidance system returned the vehicle to that attitude. While considerable deviation in attitude was experienced temporarily, no adverse effects on the rendezvous resulted." (p. 4-3).

In Section 15.2 of the report, anomalies related to the Lunar Module are discussed. Anomaly 15.2.14 addressed "attitude anomalies at staging." "Large attitude excursions occurred prior to and during staging. Body rates of 19 deg/sec in pitch and greater than 25 deg/sec in roll and yaw were recorded. Smaller attitude excursions occurred approximately 40 seconds prior to staging. The mode switching, telemetry, and associated attitude commands indicated that the abort guidance mode changed from ATT HOLD to AUTO coincident with the vehicle gyrations. … it is considered highly remote that switch malfunctions could have caused the anomalies at staging. … It is … concluded that the anomaly was caused by the inadvertent cycling of the abort guidance mode control switch, followed immediately by an incorrect output of the yaw rate gyro. … the abort guidance mode control switch was transferred to the AUTO position, resulting in high vehicle rates during the staging sequence."

2.2 Nuclear Power Industry – the Chernobyl Catastrophe

On April 25-26,1986, Unit 4 of the Chernobyl Nuclear Power Plant near Kiev, Russia, was being powered down for routine maintenance. While this process (which takes many hours) was underway, the operating crew initiated an experiment which had been previously attempted unsuccessfully. This test involved simulating a "station blackout" (loss of all offsite power), during which safety systems were intentionally switched off to test whether the plant's turbines, while spinning down to idle speed, could provide intermediate power to the backup diesel generators which were to provide power to the plant (onsite power) during the blackout. As stated above, the test had been tried unsuccessfully at least three times in the past, but it could only be performed during a planned power outage which only occurred for maintenance or fuel replacement every several months.

Despite their robust training and preparation, and following their detailed procedures, the crew was not aware of two flaws in the design of the RBMK reactor, and this contributed directly to the accident. The first was that this reactor design was unstable at low power levels; the second was that, for the first few seconds of control rod insertion (a procedure used to stop a nuclear reaction), reactor power actually increased rather than reduced as desired. There is also evidence, as recorded by a centralized (remote) control system, that an emergency shutdown of the reactor was initiated when the "EPS-5 button was pressed – this fully inserted all control rods, some of which had been withdrawn earlier" [6]. This action was wrong and proved to be the immediate trigger for the subsequent initial explosion.

Over a period of nine hours, the reactor became unstable and the crew "lost control" of it. The reactor overheated, melting the nuclear fuel and causing a series of steam explosions that tore off and lifted the 2,000-ton metal plate over the rector, blew the roof off the building, and spewed radioactivity for hundreds of miles, causing radioactive particles to be carried by prevailing winds into Western Russia and Eastern Europe.

Two deaths were recorded in the facility, 134 first responders were hospitalized, of whom 28 died of acute radiation poisoning, and 14 more died of radiation induced cancers. In addition, 15 childhood thyroid cancer deaths were recorded. Russia immediately evacuated the nearest town of Pripyat, where most of the plant's employees and their families lived. That city has been permanently abandoned and its occupants resettled. A concrete sarcophagus has been erected over the ruined facility. This, the worst disaster to confront the nuclear industry (until the 2011 meltdown at the Fukushima Daiichi nuclear plant in Japan caused by an earthquake and resultant tsunami), was caused by a highly trained crew failing to understand the behaviour of automated systems within the plant and failing to respond appropriately when these systems began to become unstable. That a "safety-related" switch was also erroneously pressed, immediately triggering the initial explosions which ultimately led to the reactor core meltdown, is further evidence of the workers' misunderstanding of the consequences of their actions during takeover from an automated system.

2.3 Aviation Industry – The Crash of Asiana Airlines Flight 214

Asiana Airlines flight 214 was a transpacific flight from Incheon International Airport near Seoul, South Korea to San Francisco International Airport. It crashed during the final approach to landing on July 6, 2013. It was the first crash of a Boeing 777 aircraft involving fatalities since that aircraft was entered into service in 1995.

The flight was cleared for a visual approach to the runway at 11:21 am, and again at 11:27. The weather was fine. There was light wind, no precipitation, and no reports of wind shear. Visibility was 10 miles – the maximum that the system could report.

The aircraft crashed into the seawall short of the runway at 11:28 am. Both engines, the tail section, and the main landing gear separated from the fuselage upon impact. After skidding along the runway, the aircraft came to rest some 2,400 feet from the initial point of impact.

The three flight crew members had extensive flying experience. The pilot in command (who also served as a check/instructor captain, had over 12,000 hours of flying experience, of which over 3,000 were in a Boeing 777 aircraft. (12,000 hours at a driving speed of 62 mph (100 km/hour) would equate to driving 740,000 miles (1.2 million km) The captain receiving his training had nearly 10,000 hours of flight experience, of which 43 were in a 777 over nine flights.

The final report of the National Transportation Safety Board (NTSB) was issued on June 24, 2014 [7]. The Board determined that the probable cause(s) of the accident were: "the flight crew's mismanagement of the airplane's descent during the visual approach, the (pilot's) unintended deactivation of automatic airspeed control, (and) the flight crew's inadequate monitoring of airspeed… " Contributing factors included: "the complexities of the autothrottle and autopilot flight director systems that were inadequately described in Boeing's documentation and Asiana's pilot training, which increased the likelihood of mode error; (and) the flight crew's nonstandard communication and coordination regarding the use of the autothrottle and autopilot flight director systems."

We have highlighted the Asiana crash because it is recent and has been in the news, and because it is a representative example of crashes (and near misses) that are the focus of this paper – the operators' failure or inability to understand the automation to a sufficient degree to take over when the automation fails or needs to hand off control. But Asiana is just one of many recent aviation examples that represent such a condition. In a recent report, Mumaw [8] has compiled brief descriptions of 42 aviation accidents and events relating to "autoflight" use and misuse. While some of these incidents date to the 1970s, the vast majority have occurred within the last 20 years, when this technology became more prevalent. Some of the event categories bear a strong resemblance to concerns about autonomous vehicles: The autopilot (or autothrottle) is off or failed and the pilot thought it was engaged; the autopilot takes an action that the pilot is not aware of; the autopilot reverts to another mode; the pilot does not understand the mode's behaviour.

2.4 The Similarities Between These Events

What are the similarities between these three events, occurring in three different industries and separated by four decades? One event, what we might call an incident, resulted in mission delays and (ultimately) considerable embarrassment. Another, what we would term an accident, resulted in the loss of two lives, the injuries of many, and hundreds of millions of dollars in a lost aircraft and legal claims arising from the event. The Chernobyl event, widely described as a catastrophe, killed several people immediately, more over the decades that followed, led to the permanent abandonment of a small city, the construction of a concrete sarcophagus around the doomed property, and the pollution of huge swaths of previously productive farmland in several countries.

The underlying factor behind these three events is the failure by the operators to understand how the automated system worked, and their inability to take over operational control of the system when the automation needed a hand-off or showed signs of failing.

2.5 The Operational Environment in These Three Industries Compared to Automated Vehicles

In our three selected industries:

- The equipment being operated is all of a specific type, (e.g. Airbus 330 or Boeing 777). The operator is “type-rated” and operates only the specific system for which he or she has been trained…
  - o But the automobile may be any of dozens of brands and hundreds of models, and other vehicles on the road may be 20 or more years old and may well be poorly maintained.

- The equipment being operated is maintained rigorously…
  - o But, although some U.S. states have minimal vehicle maintenance requirements and periodic vehicle inspections, many, including the largest, have none.

- The time scale of unfolding events demanding attention may be minutes or hours…
  - o But drivers have at most a few seconds to address an impending crash.

- There are comprehensive operating manuals that cover both normal and abnormal operations – manuals that must be read and understood in order to perform the required operations…
  - o But even the once ubiquitous owner’s manual is no longer made available to drivers; it has been replaced by online documentation that may or may not be reviewed. And there is no requirement that the operator possess any familiarity with vehicle operating procedures before taking the wheel.

- The software in aviation, aerospace, and nuclear power is typically quite stable over time, and when changes are made, operator retraining is performed prior to the update being placed into service…
  - o In automobiles, software updates may occur whenever the manufacturer deems it appropriate (an approach followed, for example, by Tesla), and there is little if any concomitant operator training, thus adding to the likelihood of some unexpected outcome or loss of system reliability.

- Operators are trained to avoid inattention to their tasks and distractions are typically prohibited. Crews of two or more personnel operate at all times, such that one member can compensate for another who may be distracted or inattentive.
  - o Automobiles are typically driven by a solo driver, who may be distracted by in-vehicle infotainment or devices (such as mobile phones) brought into the vehicle. Manufacturers paint a picture of the future driver relaxing with a magazine or television while the autonomous vehicle is in complete control.

2.6 The Capability and Preparation of the Operators in These Industries Compared to Those of Vehicle Operators

In our three chosen industries, operators:

- Are highly trained, for both normal and abnormal operating conditions...
  - But automobile drivers, at least in the U.S., receive perfunctory training at best, and none for emergencies
- Are rigorously tested and licensed…
  - But the driver's licensing process in the U.S. does not measure critical driving skills; and the license may be valid for five years or longer without any ongoing testing
    .
- Follow specific procedures that cover both normal and off-normal operations…
  - But automobile drivers follow no procedures while driving, save for the "rules of the road."
- Are medically examined regularly, and must be medically fit to maintain licensure…
  - But most drivers in the U.S. are given a standard eye test that measures only static visual acuity and must meet little or no continuing medical standards.
- Must demonstrate proficiency in a provisional capacity at the hands of a senior instructor before being permitted to operate…
  - But the provisional ("Graduated") license is generally overseen by parents, not experts, and it relates more to time behind the wheel than it does proficiency. There is typically no required proficiency demonstration for unusual or emergency events.
- Undergo periodic retraining and retesting…
  - But for drivers in most of the U.S., no retraining or retesting is required, except (in some States) for drivers over a certain minimum age.
- May not work if they are under the influence of drugs or alcohol...
  - But in the U.S., the BAC limit for driving is 0.08 percent; and there is no specified limit (or test) for drugs. Little random testing is done, and no regular testing.

(Note that, in the U.S., the operating environment and operator readiness are considerably more rigorous for interstate truck and bus drivers than they are for automobile drivers).

3. The Capability and Readiness of Drivers to Assume Control

Several authors have addressed some of the anticipated difficulties with human takeover of failed or compromised vehicle automation but have generally done so in the abstract. The present paper asks the question: by what hubris do we continue to design vehicles with advanced automation without accounting for the manner in which the human will interact with such automation when it fails or hands-over control, when extensive data from other industries (particularly aviation) highlights the often-flawed manner in which humans interact with technology in those industries, and therefore calls into question our assumptions for safe operations in the highway environment?

While it has been argued [9, 10] that drivers should have a deep understanding of how automated systems work in order to successfully respond when they fail, this goal seems all but unattainable in the automotive world when it has been shown to fail in other industries where training is rigorous, in depth, and continuous. While it is true that nuclear power plant operators as well as pilots and astronauts are thoroughly and repeatedly trained to have such underlying knowledge of the systems they operate, we do not see how such deep learning can be imparted to automobile drivers – given the time and resources required, the lack of a legal framework to require such training, and the competitive nature of the automobile industry in which manufacturers are loath to share information about their technical systems.

Stanton, writing in [9] describes the “utopian vision of the motor vehicle” that has an “onboard auto-driver, similar to the autopilot in aircraft (to) take over the driving tasks, allowing the human driver to work, rest or play.” He opines that “the Catch-22 of vehicle automation is that, while car owners are stripped of the need to perform driving tasks, they are still required to monitor their auto-driver and take manual control if the situation demands. However, when vehicles become fully autonomous, even the most observant human driver’s attention will begin to wane. Their mind will begin to wander, and they may start to mentally switch off from the job of driving.”. As Stanton and others paint this “utopian vision,” they typically include the image of the driver being able to engage in other activities or, simply, rest. These “utopian” ideals, which always include such distractions, exacerbate the conflict between a proposed need for deeper understanding of system operation and loss of focus on the driving task, should takeover from automation become necessary. While this is most commonly addressed in discussions about fully autonomous vehicles, it is of particular concern with Level 3 and Level 4 systems.

Stanton’s simulator and test-track research has shown that drivers of automated vehicles are generally less effective in emergencies than drivers of manual vehicles, and he has “repeatedly witnessed the failure of drivers to intervene when systems fail whereas almost all drivers of manual vehicles recover in the same situation.”

As a result of his research, Stanton has suggested that automation must have graduated, gradual hand-over if it is to successfully support human drivers. And he proposes that the interface between the driver and the vehicle automation be in the form of a “chatty co-pilot, not a silent auto-pilot.”

Nunes, Reimer and Coughlin [10] strike a similar tone. They believe that one approach to this problem is to educate consumers about how the automated system works, and to alert them to safety concerns that may arise. Yet, they point out, “self-driving cars are underpinned by sophisticated technologies that are hard to explain or understand.” (p. 170). They believe that “developers are designing such products to be easy to use. … However, users are then less able to anticipate how the underlying systems work, or to recognize problems and fix them.” (p. 170)

Setting a rather different tone than many other writers, these authors believe that some form of human intervention will always be required, regardless of the degree of automation. The irony of this statement comes about from the same authors’ admission that governments worldwide are freeing developers of automated vehicles from having to meet current safety requirements such as providing a steering wheel, rear view mirror, and manual braking control.

Other ironies exist. If we accept the premise that autonomous vehicles will always require some degree of user intervention, then individuals with cognitive impairments or age related cognitive decline may find the operation of such vehicles challenging. Yet these are cohorts that are expected to be among the greatest beneficiaries of automated vehicles.

Further, existing legislation in the U.S. makes no mention of either competency requirements or proficiency testing for users, and, without such standards, these authors worry, the risk of incidents might increase.

The report ends with a call to policymakers to recognize that “driverless does not, and should not, mean without a human operator;” and that automation (essentially) changes the work that people must perform – it does not eliminate it. They further posit that vehicle operators should be required to demonstrate competence – “that proficiency standards are necessary for users of autonomous vehicles and that competency should be tested by licensing authorities and should supplement existing driving permits.” (p. 171). They further advocate mandatory regular checks on user competency “so that proficiency is kept up as cognitive abilities change, and technology evolves.” (p. 171) This is a laudable and appropriate position, but, as discussed herein, likely impossible to achieve.

In her seminal chapter, “Ironies of Automation,” Bainbridge [11] could be writing for those responsible for autonomous vehicles. She describes, for example, two ironies stemming from “the designer’s view … that the operator is unreliable and inefficient, so should be eliminated from the system.” (p. 272). The first irony is that “designer errors can be a major source of operating problems,” just as we have seen with, for example, the problematic algorithm that led to the false positive situational interpretation that resulted in a pedestrian death in a crash with an Uber vehicle in Tempe, Arizona [12]. The second irony is that “the designer who tries to eliminate the operator still leaves the operator to do the tasks which the designer cannot think how to automate.” Compare this expressed irony to the Cunningham and Regan [13] and Wolmar [4] examples of autonomous vehicle failures under conditions of snow, dust, or even rain covered roads, hand-signalling by police officers, or roadside construction zone detours and sudden lane changes and drops. Bainbridge’s prescient writing reminds us that skills deteriorate when they are not used, and so an erstwhile experienced operator may become an inexperienced one when suddenly having to take over for a failed automated process that has functioned properly for an extended period. She argues that, “when manual takeover is needed there is likely to be something wrong with the process, and the operator needs to be more rather than less skilled to handle it.” (p. 272). Both Cunningham and Regan [13] and Nunes, Reimer and Coughlin [10] suggest that, in order to properly be prepared to take over in the event of automation hand-off or failure, the operator of an autonomous vehicle needs to have a deep understanding of system operation. Perhaps the “safety driver” who was “unable to prevent” the pedestrian fatality in Temple, Arizona would have been more successful had he or she possessed such deep knowledge, sufficient to timely override the faulty decision-making algorithm within the Uber vehicle’s software. Here, too, Bainbridge has offered cogent arguments some 30 years before the fact, and summarizes with the rather pessimistic view that the “current generation of automated systems” which are monitored by “former manual operators” are riding on the learned skill sets of these operators, and that future generations may not possess such skills, a view that could well apply to tomorrow’s safety drivers. Promised distractions from the driving task will further exacerbate this issue.

Eriksson and Stanton [14] state: “When the driver is assumed to resume control of a vehicle when its operational limits are reached, a critical weakness in the system is exposed. As the driver (has) been out of the control loop for an extended period of time, they may be a victim of some of the ironies of automation, where situation awareness is reduced.” Under such circumstances, they posit, the driver must receive support and guidance necessary to re-enter the control loop – and they propose the paradigm of the “chatty co-driver.” In their view, this facet of automation would provide continuous feedback via specialized user interfaces following the convention of the Gricean Maxims of successful conversation [15].

4. Two Approaches to Driver Preparation

It does not seem likely that, in the future, prior to the introduction of Level 5 vehicles into the traffic stream, either the time scale of motor vehicle operations or the physical roadway spacing in which such vehicles operate will change, except for an increase in the density of both, nor that the competition between vehicle manufacturers will permit designs or implementation of automated systems in vehicles to be harmonized. Therefore, the best hope for reducing the potential for errors when automation fails or requires a handoff lies with the human operator. And since we are not likely to see, at least in the U.S., greater rigor in the medical fitness arena or in the testing phase of the driver licensure process, it seems undeniable that improvements will have to come in the realm of driver training and preparation for dealing with automation, or in the constant feedback provided by an interface to equip the operator with current knowledge of system status and function. Either of these two approaches would mark a major step forward, although neither is likely to receive Government support or enforcement.

Although it has been shown that, "even brief training in how to respond to AV failure seems promising [13], In the U.S., at least, it can be argued that driver training has not advanced in recent years – if anything, such preparation to drive has been declining over time, with the exception of certain States' Graduated Driver Licensing (GDL) programs. Some authors have suggested that a new era of driver training is necessary, with potential vehicle purchasers required to receive training in vehicle showrooms as part of the new car purchasing experience, and we agree that specialized training and rehearsal of a driver's interaction with vehicle automation would be useful if we are to close the gap, to even a small degree, between vehicle drivers and those who operate nuclear power plants, aircraft or space vehicles. Others, including Eriksson and Stanton [14], recommend the "chatty co-driver" approach, and this novel intervention also seems to have potential. There is, however, no existing model of such a system in commercial use, and the closest approximation would appear to be the currently available on-line owner's manual. Such manuals are not, of course, real time information systems, and they require the operator to seek them out and investigate them thoroughly for them to be at all effective.

In short, two theories have emerged that purport to address a means to fill the gap when a distracted or inattentive driver is confronted with a (potentially) sudden need to re-enter the control loop and take manual control of the vehicle in the event of automation failure or hand-off. These two approaches involve specialized training in the workings and failure modes of the automation; and a continuously informative user interface to keep the driver abreast of the status and functioning of the automation at all times. Each is intended to fulfil the goal of preparing the driver to take over control at a moment's notice when it becomes necessary if the automation can no longer manage the vehicle's movements.

There are, of course, potentially serious disadvantages to each approach. In the first case, training to a level presumably necessary to handle such automation failures or hand-offs is nearly impossible given the size of the driving population, the uniqueness of each manufacturer's automation implementation, and the logistics of requiring the purchasers of automated vehicles to participate in such training. Further, training to a level necessary to respond to any and all failures or hand-offs (especially when many may not be known) as is the case with pilots, astronauts, and nuclear power plant operators (who *still* exhibit occasionally fatal misunderstandings of the automation), is an unreasonable and unreachable expectation given the nature of the driving environment and driver availability to participate in such training. In addition, recurrent training, to refresh skills or keep pace with changes in automation, routine in these other industries, is less feasible still. In the case of the "chatty co-driver," such interfaces would have to be designed for every implementation of vehicle automation, and system designers would be tasked with designing such a supportive interface for hand-off and failure modes that might not be fully understood. On the implementation side, a near-constant source of voice communication might provide exactly the type of in-vehicle information that drivers of automated vehicles are hoping to escape – seeking rest and relaxation (read inattention and distraction) while the vehicle drives itself. Thus, there is the risk that drivers will turn off (if possible) the interface or learn to ignore it, thus defeating its very purpose.

Nonetheless, given the constraints of the operating environment and the overall lack of preparation of vehicle operators to take over from an automated system, these two approaches seem to offer promise to improve the likelihood of success in such takeovers.

A failure to begin to evaluate interventions such as these would be abrogating our responsibility to maximize road safety in Level

3 and, especially Level 4 vehicles. It is with hubris that we continue to move forward with automation technologies while failing to prepare present and future vehicle operators to interact with those technologies, particularly when they require human intervention; and such automated systems are likely to require such intervention for many years to come.

Accordingly, we have proposed a thought experiment to examine the feasibility of the two interventions discussed above.

5. A Thought Experiment

Other than those authors who seem to think that the movement toward fully autonomous vehicles will provide a utopian, highly functional and risk-free driving environment, others have pointed out that driver complacency, distraction and inattention, coupled with a lack of understanding of the inner-workings of the automated system, will result in a dangerous driving environment for years to come. Although the data set is small and the results, therefore, not significant, crashes per million vehicle miles are far higher in autonomous vehicles on the streets than they are for the overall vehicle population [16], and the number of handoffs of the automation to a "safety driver" (called "disengagements") are, not surprisingly, quite high. *Recode* has reported on 2017 data showing that Uber disengagements occurred nearly once per driven mile, and that "critical" disengagements (to avoid hitting a person or causing more than $5,000 in property damage) occurred, on average, once per 125 miles driven [17]. It has been suggested that a form of specialized operator training, or an in-vehicle interactive assistant could enable a reduction in otherwise foreseen driver failures to timely respond to automation failures or hand-offs.

We, therefore, propose a thought experiment to examine the viability of these two possible interventions.

5.1. A Training Protocol

A training protocol, likely an interactive, computerized series of lessons based on existing online operators' manuals, would be developed. For testing purposes, this protocol would be limited to a specific, challenging subset of possible automation failures or handoffs. In order to be acceptable to the automotive industry and the public alike, a pilot test of the effectiveness of the protocol would likely have to be conducted in automotive dealerships with volunteer participants and/or as part of the vehicle purchasing process. A reward would be provided for participation, perhaps in the form of dealership merchandise. Participants would be encouraged to bring to the session their own choice of entertainment or relaxation (e.g. music, reading materials, computer games, etc.). A 20-30-minute session conducted in a part-task interactive driving simulator in the showroom would first familiarize the participants with the selected subset of automated features and the failure and recovery modes for these features, provide an opportunity for any questions that the participant may have about the training, conduct the actual training, and then test its effectiveness on simulator driving scenarios. Scenarios that are functionally equivalent to those used in the training session would be used to test the appropriateness and timeliness of the participant's responses. Ideally, a follow-up session would test retention of the information after several weeks. Participants would be queried regarding their opinion of, and satisfaction with, the training model. A careful review of failures would need to be kept to advise on possible revisions to the protocol.

5.2. A Smart Assistant

The intervention of a "smart assistant" or "chatty co-driver" would also be introduced through part-task simulation, and would follow the same protocol discussed above but, since this system is meant to operate on the road in real time, upfront training would be limited to an introductory familiarization session on how the system functions, and how it should be used. After this introduction, any participants' questions would be addressed.

The same scenarios and automation failures/hand-offs as in the training model would be presented, with the (previously developed) smart assistant providing continuous information and feedback to the participant/driver. The same equivalent form simulator scenarios as used in the training protocol would be used here, again to test the appropriateness and timeliness of the participant's response to failures and hand-offs. Again, a follow-up session would test retention of the information after several weeks. And again, participants would be queried regarding their opinion of, and satisfaction with, the smart assistant system. Finally, as in the training protocol, a review of failures would be critical for designing any necessary revisions to the smart assistant.

It is suggested that this series of trials would shed light on the functionality, viability, and consumer (and manufacturer and regulator) acceptance of the training approach vs. the smart assistant.

6. Conclusions

We have described three different incidents that have occurred in three different industries, in each case where the operators were highly trained, rigorously tested, and medically fit. We pointed out the vast differences between the operating environment and operator readiness in these industries compared to that in the highway setting. We then explored different theories of the "ironies of automation," and looked at different approaches to addressing the safety implications of these ironies, particularly for Level 3 and 4 automation. We proposed a thought experiment to evaluate two such approaches – in-depth training into system operation and failure, and an on-board "chatty assistant" to keep the driver continuously informed about automation state. We express our concern that, given the experiences with automation in highly regulated and controlled industries and the current lack of such regulations and control in the automotive field, it is with considerable hubris that we continue to advance vehicle automation with full knowledge that driver takeover will be required for many years to come but without any real commitment to driver preparation for such takeover.

7. References

[1] 'Hyatt, K., Paukert, C.: 'Self-driving cars: A level-by-level explainer of autonomous vehicles', https://www.cnet.com/roadshow/news/self-driving-car-guide-autonomous-explanation, accessed 15 May 2018

[2] 'Automated vehicles for safety', https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety, accessed 15 May 2018

[3] Godsmark, P. 'The definitive guide to the levels of automation for driverless cars', https://driverless.wonderhowto.com/news/definitive-guide-levels-automation-for-driverless-cars-0176009/, accessed 15 May 2018

[4] Wolmar, C.: Driverless cars: 'On a road to nowhere' (London Publishing Partnership, 2018)

[5] National Aeronautics and Space Administration, 'Apollo 10 Mission Report' (Houston: Manned Spacecraft Center, 1969), pp. 1-323

[6] 'The Chernobyl Gallery', http://chernobylgallery.com/chernobyl-disaster/cause/, accessed 15 May 2018

[7] 'Board meeting: Crash of Asiana flight 214 accident report summary' (National Transportation Safety Board, June 24, 2014, https://www.ntsb.gov/news/events/Pages/2014_Asiana_BMG-Abstract.aspx, accessed 15 May 2018

[8] Mumaw, R.J. **'**Addressing mode confusion using an interpreter display' Contractor Technical Report, NASA Contract # NNL16AA15C. San Jose State University Research Foundation: Moffett Field, CA. DOI 10.13140/RG.2.2.27980.92801, 2018

[9] Poulin, C., Stanton, N.A., Cebon, D.: 'Response to: Autonomous vehicles', Ingenia online, 2015, Issue 62, unpaginated,

http://www.ingenia.org.uk/Ingenia/Articles/943, accessed 17 May 2018

[10] Nunes, A., Reimer, B., Coughlin, J.F.: ‘People must retain control of autonomous vehicles’, Nature, 2018, 556, pp. 169-171

[11] Bainbridge, L. ‘Ironies of automation’, in Rasmussen, J., Duncan, K., Leplat, J. (Eds.): ‘New Technology and Human Error’ (Wiley, 1987) pp. 271-283

[12] Efrati, A. ‘Uber finds deadly accident likely caused by software set to ignore objects on road’, 7 May, 2018, https://www.theinformation.com/articles/uber-finds-deadly-accident-likely-caused-by-software-set-to-ignore-objects-on-road, accessed 14 May 2018

[13] Cunningham, M.L., Regan, M.A. ‘Driver distraction and inattention in the realm of automated driving’, IET Intell. Transp. Syst., 2017, pp. 1-7

[14] Eriksson, A., Stanton, N.A. ‘The chatty co-driver: A linguistics approach to human-automation-interaction’, Human Factors in Organizational Design and Management – XI, Nordic Ergonomics Society Annual Conference – 46, Undated, unpaginated

[15] Grice, H.: ‘Logic and conversation’, in Cole, P. and Morgan, J. L. (Eds.): ‘Speech Acts’ (New York, Academic Press), 1975), pp. 41-58

[16] Els, P. ‘How much testing will prove automated cars are safe?’, Automotive IQ, 2018, Unpaginated, https://www.automotive-iq.com/autonomous-drive/articles/how-much-testing-will-prove-automated-cars-are-safe, accessed 15 May 2018

[17] Bhuiyan, J. ‘Uber’s autonomous cars drove 20,354 miles and had to be taken over at every mile, according to documents’, Recode, 2017, https://www.recode.net/2017/3/16/14938116/uber-travis-kalanick-self-driving-internal-metrics-slow-progress, accessed 16 May 2018