Shared control versus traded control in driving: a debate around automation pitfalls

J. C. F. de Winter, S. M. Petermeijer & D. A. Abbink

To cite this article: J. C. F. de Winter, S. M. Petermeijer & D. A. Abbink (2022): Shared control versus traded control in driving: a debate around automation pitfalls, Ergonomics, DOI: 10.1080/00140139.2022.2153175

To link to this article: https://doi.org/10.1080/00140139.2022.2153175

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 15 Dec 2022.

Submit your article to this journal

Article views: 203

View related articles

View Crossmark data
1. Introduction

Automation, defined as ‘the execution by a machine agent (usually a computer) of a function that was previously carried out by a human’ (Parasuraman and Riley 1997, 231), was initially used for relatively simple routines. Examples of successful implementation of such automation can be found in households (e.g. dishwashers) and factories (e.g. robotic arms, assembly line), where the task is relatively constrained. Automation is now becoming increasingly viable in unstructured ‘open’ environments, such as the driving domain. Here, full automation is possible under some conditions, but not all. For example, in the current state of technology, computer vision algorithms are unable to detect and understand the social cues of vulnerable road users (Camara et al. 2021; Rudenko et al. 2020; Saleh et al. 2017). Although there have been continual advancements in sensors and artificial intelligence, automation in unstructured environments currently still requires some form of human supervision, interaction, or correction.

Given that human involvement is necessary, the question is how to design human-automation interaction, and in particular whether driving functions should be traded between human and automation, or shared. The topic of sharing versus trading was first addressed by Sheridan and Verplank (1978), who explained: ‘In the case where both computer and human are working on the same task at the same time,
we call this SHARING control. When they work on the same task at different times this is TRADING control (4–6) (Figure 1). Thus, in traded control, the human operator delegates tasks to the automation, supervises the automation, and occasionally resumes manual control. On the other hand, in shared control, human and automation perform the same task congruently (Abbink et al. 2018; Sheridan and Verplank 1978).

The term shared control tends to be used inappropriately in the literature, even by human factors experts (Tabone et al. 2021), to describe any human-automation interaction in which the automation is imperfect and the human has a vital role. However, as noted by Norman (quoted in Tabone et al. 2021), many of such interactions are, in fact, traded control, in which the human has to monitor while the automation executes a control task. In an attempt to define shared control more sharply, Abbink et al. (2018) explained that shared control refers to a congruent human-automation interaction in a perception-action cycle to perform a task that the human or the automation can, under the right conditions, also perform alone. They clarified that the use of warning and decision support systems does not qualify as shared control. These systems provide unilateral suggestions on what action the human should (not) take, while the human subsequently has to make the decision and execute the action. Thus, although sensory and perceptual task elements are shared between human and machine, there is no congruent interaction. Stability and control augmentation systems do not belong to the category of shared control either, as such systems are meant to perform tasks that humans cannot do by themselves. For example, the control of various military aircraft relies on rapid computer-based ‘inner loop’ control, as such aircraft have marginal or even negative aerodynamic stability (Avanzini and Minisci 2011). In these cases, although humans and automation perform a task at the same time, it is not the same task that is shared; in fact, if the computer fails, the aircraft will be unable to fly.

Shared control can come in a variety of forms. Two main types of shared control can be distinguished (Abbink et al. 2012): shared control involving a physically coupled interaction between input device and vehicle or robot (called haptic shared control; Benloucif et al. 2019; Ghasemi et al. 2019; Griffiths and Gillespie 2005; Losey et al. 2018; Mars et al. 2014) and shared control involving a physically decoupled interaction (also called: blending, indirect, or input-mixing shared control; Crandall and Goodrich 2002; Dragan and Srinivasa 2013; Storms and Tilbury 2014). We refer to review papers for additional consideration and detail (Ghasemi et al. 2019; Marcano et al. 2020; Wang et al. 2020). Shared control methods involving multiple humans and/or multiple robots (Crandall et al. 2017; Gao et al. 2005; Musić and Hirche 2017; Shang et al. 2017; Tso et al. 1999), or methods where human and automation jointly arrive at a plan, decision, or strategy have been conceived as well (e.g. Abbink et al. 2018; Kaber and Endsley 2004; McCourt et al. 2016; Pacaux-Lemoine and Itoh 2015). It is further noted that shared control, like traded control, does not need to refer to the entire task but can also be applied to separate control inputs. For example, a driver may use haptic shared control on the steering wheel, but without car-following assistance. In this example, lateral control is shared, while longitudinal control is performed manually. If the same driver uses adaptive cruise control (ACC), then lateral control is shared, and longitudinal control is traded. If the same driver uses a haptic gas pedal (e.g. Mulder et al. 2008), then lateral and longitudinal control are both shared.

Traded control may also come in variants that differ in their activation and deactivation triggers. For example, a human may enable the automation system via a button press, but in time-critical conditions, the automation may activate automatically and override the human (Lu et al. 2016). The latter solution is seen in advanced emergency braking (AEB), for example. Parasuraman et al. (1992) report several other activation triggers for traded control, such as task phase (e.g. an automated vehicle may change mode depending on geospecific information) and physiological signals (e.g. an automated vehicle may switch back to manual if it
determines that the driver is distracted). Each of these examples is a manifestation of traded control.

Since the seminal work of Sheridan and Verplank (1978), a large number of human factors researchers have examined specific variants of traded control and shared control. However, only a few articles have empirically evaluated sharing and trading side-by-side (Metcalf et al. 2010; Van Dintel et al. 2021) or have provided a theoretical outline on sharing versus trading (Inagaki 2003). De Winter and Dodou (2011) provided a critical reflection on shared control in driving, noting that it yields only meagre performance benefits compared to manual control. Abbink et al. (2012, 2018) explained that the aim of haptic shared control is not to improve performance, but to keep humans in the loop and to let them catch automation errors when the automation (suddenly) reaches the limits of its capabilities. Although a large number of empirical papers have been published on shared control and traded control, a rigorous reflection on the pros and cons of these methods of human-automation interaction seems lacking.

1.1. Aim

The present article offers critical reflections on the arguments for and against shared or traded control, which originated from debate sessions the authors organised during human factors lectures over the past 10 years. This write-up of the debate is meant to stimulate the exchange of ideas, and to formulate new hypotheses for research, an approach sometimes used in academic outlets (e.g. Wilde et al. 2002, for a debate on a road safety theory with a ‘for’ vs. ‘against’ structure).

In a debate, it is important to clearly define the positions of both parties. That is why it was decided not to focus the debate on shared and traded control in general, but on the specific domain of car driving. So far, the introduction of automation in driving has been evaluated mostly through the lens of traded control. To illustrate, Zhang et al. (2019) documented as many as 129 experiments that measured how quickly drivers take over control from an SAE Level 3 ‘traded control’ automated driving system, while Jansen et al. (2022) surveyed 189 studies on the design of take-over warnings. To encourage sharp debate, it was decided to focus on one specific type of shared control, namely haptic shared control. In this type of shared control, the driver receives guiding torques at the control interfaces, but there remains a directly coupled relationship between input (e.g. steering wheel angle) and output (e.g. wheel angle). Basic haptic shared control solutions in the form of torque feedback for hands-on-wheel lane-keeping assistance are currently available on the automotive market (Reagan et al. 2018).

This article first surveys known and lesser-known pitfalls of human-automation interaction in general. These pitfalls have mostly been studied in the context of traded control, but as this debate paper will elucidate, some of the pitfalls may also apply to shared control. Although the pitfalls of automation, sometimes called ironies of automation, have already been recognised for many decades (Bainbridge 1983; and see Parasuraman and Riley 1997, for automation misuse and disuse specifically), a concise yet comprehensive overview of the origins, definitions, and operationalisations of the pitfalls was still lacking in the literature.

Next, we focus on driving and present the position of one of the authors that haptic shared control is the remedy for these pitfalls and the recommended way for introducing automation on the roads. This is followed by a rebuttal that argues that haptic shared control is not the way forward and explains why traded control is the better alternative. We end with a reflection towards a consensus by outlining the main factors that determine whether to share or trade control.

2. De winter: general pitfalls of automation

Although automation offers many benefits, it comes with several pitfalls, as a result of which the benefits of automation are not realised to their full extent. When examining the literature, six pitfalls can be identified. These pitfalls all relate to the traded-control solution for human-automation interaction. In traded control, the human is not in direct control of the process, but most of the time supervises the automation. That is, the human intermittently monitors and adjusts the automation, which itself controls the process (Sheridan 1992). This brings the human out of the direct control loop, resulting in problems when the human needs to be in control again after the boundaries of the operational design domain are exceeded.

2.1. Pitfall 1: loss of situation awareness and mode awareness

When supervising automation, the human may sample limited information from the task environment and automation-status displays, as a result of which the human may lose situation awareness, mode awareness, or both.

First, loss of situation awareness occurs when the human fails to detect and perceive relevant elements of the environment, does not comprehend the
meaning of these elements, and does not anticipate how the environment will evolve (Endsley 1995; see Figure 2 for an illustration of the meaning of situation awareness).

Loss of situation awareness can be detected in virtual environments through so-called freeze-probe methods, such as the Situation Awareness Global Assessment Technique (SAGAT; Endsley 1988), whereas in real environments, real-time probes such as the Situation-Present Assessment Method (SPAM; Durso et al. 1998) and eye-tracking (De Winter et al. 2019) can be used. In certain applications, basic human-monitoring systems are already available. For example, driver monitoring systems that use cameras to detect whether the driver is keeping attention to the road can be found in several car models, which can then warn the driver to redirect attention to the main task (e.g. BMW 2019; and see Stein et al. 2019, for a system developed for train drivers).

Although automation is known for being able to cause a loss of situation awareness, the literature contains several examples where automated driving improves situation awareness, something that can occur if the driver is motivated or incentivized to remain attentive to the task environment (for a review of experimental data in automated driving, see De Winter et al. 2014; for an interview study of Tesla Autopilot users, see Nordhoff et al. 2022). A likely explanation is that because the automation is in control, the human is offloaded from necessary manual tasks and therefore may have more spare visual and mental capacity.

Second, mode awareness may occur when humans lose awareness of the mode the automation is in, where the term mode refers to the automation’s functionality (control laws, settings) that is currently active. Humans may then commit mode errors, where ‘the system behaves differently depending on the state it is in, and the user’s action is inappropriate for the current state’ (Norman 1980, 32). In the same vein, Potter et al. (1990) defined a mode error as ‘acting in a way appropriate to one mode when the device is actually in another mode’ (393). Mode errors are more likely when the automation features a large number of modes and when mode changes can be triggered automatically based on environmental triggers (Sarter and Woods 1995).

Mode errors have been implicated in aviation accidents (Silva and Hansman 2015), oil spill accidents (Seastreak Wall Street; National Transportation Safety Board, 2014)2, and industrial accidents (Three Mile Island; Rogovin and Frampton 1980), and also start to be of concern in daily tasks in which automation finds its entrance (Andre and Degani 1997), such as driving (Wilson et al. 2020). Current automated driving
systems consist of various subsystems, such as lane-keeping assistance, AEB, and traffic light detection, and drivers may have difficulty understanding which of these subsystems are active (mode on vs. off, or their current functionality) (Banks et al. 2018; Dönmez Özkan et al. 2021; Feldhüter et al. 2018).

2.2. Pitfall 2: deskilling

Known under various terms, including skill erosion (Wall et al. 1980), skill loss (Wiener and Curry 1980), skill degradation (Birkemeier 1968), and skill atrophy (Sheridan et al. 1983), deskilling refers to the hypothesis that the human operator may, over time, lose manual control skills because of automation use. That is, because of not rehearsing manual skills, operators become less proficient relative to those who continue to perform the task manually. Deskilling is a frequently mentioned topic in the human factors literature (Billings 1997; Endsley and Kiris 1995; Ferris et al. 2010; Lee and Seppelt 2009).

Experiments on the retention of skills in tracking and procedural tasks show that performance declines with the duration of the no-practice interval (Ammons et al. 1958). Other research suggests that retention of performance in tracking tasks is generally high (Fleishman and Parker 1962) and that performance on discrete procedural tasks is forgotten more readily than performance in tracking tasks (see Adams 1987, for a review). Figure 3 shows participants’ performance on a tracking task during a practice phase and a retention phase 1–2 years later. In this lab experiment by Ammons et al. (1958), participants had to keep a model plane, which was affected by various disturbances, ‘on target’, i.e. level and pointed straight ahead. Deskilling can be recognised from the beginning of the retention interval, that is, performance in the first 15 min after the long break was lower compared to the end of the practice phase.

In the area of human-automation interaction, experimental evidence for deskilling is scarce, or as noted by Wood (2004): ‘Little research exists to provide a structured basis for determination of whether crews of highly automated aircraft might lose their manual flying skills’ (v). The reason may be that proving whether deskilling has occurred requires a longitudinal study with an experimental group that agrees not to use the automation for an extended period and a control group that continues to use the automation, something that is difficult to accomplish.

So far, the evidence of deskilling is correlational (Casner et al. 2014; Ebbatson et al. 2010). For example,
a flight simulator study by Haslbeck and Hoermann (2016) found that short-haul pilots had better manual control skills than long-haul flight pilots, presumably because the latter group completed fewer landings per year. However, it is difficult to draw definitive conclusions because the two groups flew different aircraft (i.e. different weight classes). It is also difficult to determine whether a skill deficit is caused by the small amount of initial practice in manual flying or whether it is caused by forgetting how to fly manually due to the no-practice interval (cf. Figure 3).

At the same time, anecdotal evidence of deskilling is abundant. We can refer here to matters of daily life where it is said that people have become less proficient in mental arithmetic due to the introduction of computers and calculators (e.g. Penny 2022). An interview study by Nordhoff et al. (2022) revealed that some users of Tesla’s Autopilot felt uncomfortable or somewhat scared when having to drive a manual car after being used to automated driving. In aviation, it has been reported that airline pilots have little stick time and that simulator training is essential to maintain manual skills (Doyle 2009; Hanusch 2017). The Federal Aviation Administration (2017) issued a Safety Alert for Operators, asking air carriers to ‘permit and encourage manual flight operations … when conditions permit, including at least periodically, the entire departure and arrival phases, and potentially the entire flight, if/when practicable and permissible’.

2.3. Pitfall 3: unbalanced mental workload

When supervising automation, human operators may (have the impression that they) have little to do because the computer controls the task environment. The human operator should set up and monitor the automation, but most of the time does not have to provide any inputs. This situation may lead to low workload, where workload can be defined as ‘the cost of accomplishing task requirements for the human element of man-machine systems’ (Hart and Wickens 1990, 258).

Although low mental workload is not a problem when the automation works as intended, it may lead to problems when the automation acts on the wrong information, makes wrong decisions, or fails altogether. As noted by Fitts (1951, 6): ‘if the primary task of the human becomes that of monitoring, maintaining, and calibrating automatic machines, some men will need to be capable of making intelligent decisions and taking quick action in cases of machine breakdown or in unforeseen emergencies.’ The phrase ‘99% boredom and 1% sudden terror’ (Bibby et al. 1975, 665) captures the notion that workload can increase strongly when the human has to reclaim control. The human operator may have to comprehend warning signals on the human-machine interface or diagnose why the automation malfunctions. Krebsbach and Musliner (1999) noted that ‘minor incidents may cause dozens of alarms to trigger, requiring the operator to perform anywhere from a single action to dozens, or even hundreds, of compensatory actions over an extended period of time’. When the human operator is stressed or mentally overburdened, accident probability is elevated, especially when the human needs to regain situation awareness simultaneously.

In non-automated tasks, humans can usually cope well with high task demands, something that Norman and Bobrow (1975) called the ‘principle of graceful degradation … when human processes become overloaded, there often appears to be a smooth degradation on task performance rather than a calamitous failure’ (44–45). Similarly, Fitts (1951) used the example of a telephone switchboard, with switchboard operators managing to connect most calls even in busy times while ‘automatic dial telephone systems are known to have broken down completely under overload conditions’ (6). This problem of automation has also been called automation brittleness, i.e. automation operates well for the range of conditions for which it was designed but requires human intervention for situations not covered by its programming (Cummings et al. 2010). When automation fails abruptly, the human is confronted with a situation that needs immediate resolution. The term clumsy automation is used for automation that has been designed in such a way that it reinforces issues of unbalanced mental workload rather than resolves it (Cook and Woods 1996; Wiener 1989).

Hart and Wickens (1990) explained that workload ‘may be reflected in the depletion of attentional, cognitive, or response resources, inability to accomplish additional activities, emotional stress, fatigue, or performance decrements’ (258). Accordingly, mental workload can be measured using questionnaires, such as the widely used NASA Task Load Index (Hart and Staveland 1988). The use of secondary tasks to measure spare mental capacity (Brown and Poulton 1961; Williges and Wierwille 1979) and eye-tracking and other physiological measurements are common as well (Cain 2007; Kramer 1991; Marquart et al. 2015). Under conditions of workload and stress, common physiological responses are visual tunnelling (Rantanen and Goldberg 1999), increased heart rate (Roscoe 1992), and dilated pupils (Kahneman and Beatty 1966). Psychophysiological
signals can act as a dependent variable in human-sub-
ject experiments, or in combination with performance
measures to assess workload/performance (dis)associa-
tions (Hancock 1996). Alternatively, psychophysiological
signals can serve as a trigger in traded-control automa-
sion systems (Cabrall et al. 2018).

2.4. Pitfall 4: behavioural adaptation

Behavioural adaptation refers to the fact that human
operators may be inclined to use new technology,
including robotic or automation systems, in such a
manner that the benefits of this technology are not
realised to their expected extent. The Organisation for
Economic Co-operation and Development (1990)
defined behavioural adaptations in the context of driv-
ing as ‘those behaviours which may occur following the
introduction of changes to the road-vehicle user system
and which were not intended by the initiators of the
change’ (115).

Risk compensation/homeostasis (cf. Wilde 1982),
but also workload/difficulty homeostasis (Fuller 2005),
control-activity homeostasis (Melman et al. 2018), a
satisfaction of personal needs (Organisation for
Economic Co-operation and Development 1990), or a
production/protection trade-off (Reason 1997) are
among the proposed mechanisms behind behavioural
adaptation. The essence of behavioural adaptation is
that automation, which helps perform tasks, results in
a compensatory mechanism by the human operator.

Behavioural adaptation has been observed in man-
ual control and human-supervised automation. When
ships were fitted with radar in the mid-twentieth
century, this protective feature resulted in the vessels
travelling faster through the seaways (Schmidt 1958;
see also Reason 1997). Similarly, when a simulated
vehicle was equipped with ACC, drivers drove this
vehicle at greater speeds compared to manual driv-
ing (Hoedemaeker and Brookhuis 1998). Thus, the
automation feature was used in such a way that the
safety benefits were not realised to their expected
extent, or, in the words of Elvik (2004), the engineer-
ing effect (e.g. a radar providing better visibility
compared to no radar) was offset by a behavioural
effect (sailing the ship faster with radar than with-
out), see Figure 4.

Behavioural adaptation has been described as a
universal phenomenon of systems (Wilde 2013). Also
at a macroscopic level, protective features may lead
organisations to take more risks (Reason 1997).
Behavioural adaptation can also be seen in other
forms of human-machine interaction. For example, in
military robotics, there are concerns that remoteness
from actual hazards may cause operators to engage in
war activity more easily (Lin et al. 2008).

2.5. Pitfall 5: misuse

Misuse refers to the unjustified use of automation by
its user. The term is often used in conjunction with
the terms overtrust (i.e. trust that is too high given
the automation’s actual reliability), overreliance, and
complacency (i.e. insufficient scepticism and insuffi-
cient counterchecking of automation functioning).
Misuse also relates to the term automation bias, that
is, the user’s reliance on automation to preserve cog-
nitive effort (Parasuraman and Manzey 2010).

Automation misuse may be exacerbated by a poor
human-machine interface, for example, when the
automation modes depicted on the interface are insuf-
ficiently salient or when the automation’s behaviour is
insufficiently transparent. Billings (1991) explained:
‘automation … should never be permitted to fail
silently’ (85). Furthermore, highly reliable automation
may give the impression that the automation is per-
fectly reliable, as a result of which the operator may
enter a state of overtrust (Figure 5). For example,
Radlmayr et al. (2018) showed that time to collision
(TTC) in a take-over scenario was reduced when the
automation reliability was 100% compared to 50%-reli-
able automation and manual driving, i.e. participants
braked later because they may have assumed that
automation would handle the situation for them.
According to Parasuraman and Riley (1997), low oper-
ator confidence and low skill are other purported

![Figure 4](image-url)

**Figure 4.** Illustration of behavioural adaptation, where an automation feature is less effective than it could be, because of an adverse behavioural effect that mitigates the engineering effect (based on Elvik 2004).
causes of misuse, as unskilled operators may more easily delegate responsibility to automation.

Misuse of automation has often been reported as a cause of accidents. One of the first fatal accidents involving a partially automated vehicle, for example, has been attributed to overreliance on automation (National Transportation Safety Board 2017). In this case, the car crashed into the side of a truck, while the driver was reported to have ignored various hands-on-wheel alerts.

Misuse can also manifest itself when interacting with decision aids. An error of commission is said to occur when the human follows the advice of a decision aid even though other available indicators suggest that this advice is invalid (Skitka et al. 1999). Bahner et al. (2008) found that operators who made more errors of commission while supervising a simulated life support system of a spacecraft showed signs of complacency, i.e. they were less likely to countercheck the raw sensor data. Oppositely, errors of omission occur when the human operator did not take appropriate action because the decision aid failed to provide advice (Skitka et al. 1999). Lin et al. (2017) reviewed 158 errors of omission and commission associated with the use of GPS devices while driving, 28% of which involved a fatality.

### 2.6. Pitfall 6: disuse

Disuse is the opposite of misuse and refers to the failure to use automation when automation use would, in fact, be appropriate. There are various causes of disuse. One factor is the human’s general negative predisposition to automation or dislike of ‘bells and whistles’ in the cockpit. The literature reports substantial individual differences in automation (dis)use, with some accidents caused by operators (creatively) disabling the automation system (Parasuraman and Riley 1997).

A second factor is annoyance or distrust that arises due to the need to switch off the automation and false alarms (Bliss 2003; Muir and Moray 1996; Sorkin 1988). Parasuraman et al. (1997) explained that false alarms are essentially unavoidable given omnipresent sensor noise and the imperfect predictability of the task environment: If the warning threshold is liberal, the operator has sufficient time to respond, but at the cost of false alarms. On the contrary, if the warning threshold is strict, the likelihood of false alarms is low, but the operator may have too little time to intervene. Hence, imperfect automation implies that the human operator should decide when not to use or ignore the automation. As a result, some operators may develop distrust in automation (Figure 5) and never reap its benefits.

Also with today’s assisted/automated driving, such as ACC and lane-keeping assist, drivers show disuse for reasons related to personal predispositions (e.g. ‘I don’t need it’, ‘I am able to stay in the lane myself’) and low perceived reliability (‘The system does not work well’, ‘The system annoys me’, ‘The systems distract me and I want to concentrate only on driving’) (Stiegemeier et al. 2022).

### 2.7. Automation pitfalls summary

In summary, automation induces a number of pitfalls. Pitfalls 1–3 concern a change in human knowledge or mental/physiological state, while Pitfalls 4, 5, and 6 concern a change in the way the human operator uses automation.

Apart from these six pitfalls, which describe the effects on human operators, researchers have been concerned with the broader implications of automation. The term abuse of automation (Parasuraman and Riley 1997) describes the implementation of automation by managers without due regard for its implications for human well-being and safety. Others have lamented that automation may cause a loss of jobs or job dissatisfaction (Bhargava et al. 2021; McClure 2018), issues going back to the Luddite movement (Darvall 1934; Hobsbawm 1952). These broader factors are beyond the scope of the present work.

### 3. Abbink: a solution to the pitfalls: haptic shared control

I fully agree with the above pitfalls of automation, and argue that they are very much present in the automotive domain. It should be noted that these pitfalls are not the result of automation itself, but rather the
consequence of the design choice to use traded control for driver-automation interaction. I argue that traded control should be avoided and that shared control is the solution to the discussed pitfalls.

### 3.1. In an open world, driver input will remain needed

For many decades, society has been promised fully self-driving vehicles. As early as 1956, America’s Independent Electric Light and Power Companies published an ad depicting self-driving vehicles in which families were comfortably transporting on the highway. In the meantime, many vehicles have been demonstrated to operate ‘driverlessly’ under specific conditions, with the latest examples, such as Waymo, operating in the less congested parts of some major cities. However, we are still a long way from having driverless vehicles commonly available, operating without human input always and everywhere.

So far, commercially available vehicles are not self-driving, but rather automated, which means that the driver must remain attentive at all times. Higher levels of automation that permit the driver to engage in non-driving tasks are mostly still in a research phase. These higher levels of automation would have to be able to provide enough time (typically discussed as being 10–30 s) to allow the driver to get back in the loop. Note that even if it becomes technologically feasible to ensure lead times that are sufficiently large for drivers to engage in multitasking while being able to take over control (beyond slow-moving traffic), such solutions are likely to suffer from deskilling (Pitfall 2) and behavioural adaptation (Pitfall 4). In other words, in the present state of automation technology, there is still an important role for the driver, and safe traded-control solutions hardly exist.

Real-world evaluations of commercially available traded-control automation systems confirm that these systems are not immune to unsafe incidents. The first broad examination of crashes involving automated driving technologies on US roads showed that almost 400 crashes were reported by automakers between July 2021 and May 2022 (National Highway Traffic Safety Administration 2022a). Kim et al. (2022) showed that participants experienced safety-critical events with traded-control automation such as ACC. The same study also showed that drivers felt comfortable engaging in secondary tasks (Pitfall 5), something they clearly should not, given the safety-critical events mentioned above. Participants also reported discomforting situations (automation disengagements, insufficient braking or acceleration), as well as more dangerous events such as unnecessary braking without any lead vehicle, or even silent automation failures (Kim et al. 2022). These phenomena may result in annoyance and disuse (Pitfall 6).

Three phenomena outside of the academic literature further illustrate the need for human inputs in automated driving. First, no car manufacturer assumes responsibility for accidents when their automation mode is being used, and only a few car manufacturers seem to commit themselves to Level 3 or higher automation. To leave the driver legally responsible for situations in which automation fails to work can be seen as moral scapegoating (Elish 2019) and points to so-called responsibility gaps (Calvert et al. 2020) and a lack of meaningful human control (Siebert et al. 2021). A recent study (Beckers et al. 2022) illustrates such moral scapegoating in the general public. When asking participants to attribute blame for a crash resulting from a sudden automation failure, which the driver could in principle prevent when not suffering from lack of situational awareness (Pitfall 1), the respondents would blame the driver rather than the manufacturer or legislator. Second, ironically, during the outbreak of COVID-19, the major developers of ‘driverless cars’ had to shut down their operations due to the inaccessibility of test drivers (Vengattil 2020). In fact, it can be argued that the only reason why ‘driverless cars’ are able to drive around safely outside designated areas is that test drivers catch all automation failures. Third, a cursory glance on YouTube (e.g. ‘autopilot’ failures) reveals ample examples where supposedly driverless vehicles appear unable to perform basic operations, such as interacting with pedestrians, handling intersections, late merging, or even staying in the lane. Automated vehicles can also be misled by ambiguous perceptual information (e.g. false-positive traffic light detections), and misused by pedestrians (e.g. Millard-Ball 2018).

Of course, high-end self-driving cars are being tested on public roads, but there is a big difference between demonstrating that automation can work versus demonstrating it can work always and everywhere. Remember that already in 1947, a fully automated plane demonstrated the ability to perform take-off, flying, and landing (Boxer 1948; Howard 1973). Today, we still need two pilots in the cockpit to guarantee high levels of safety. Similarly, in the automotive domain, the challenge is not only to demonstrate autonomous capabilities on wide roads under uncongested and sunny conditions but to make the automation work always and everywhere in mixed
traffic. As long as drivers are responsible for always taking over, we have to deal with driver-automation interaction and the associated pitfalls of automation.

### 3.2. Why traded control works well in aviation but not in driving

The traded control approach has been widely used in commercial aviation. However, this has only been possible because pilots are selected, extensively trained, and paired to operate the aircraft, because there are strict procedures, checklists, and assistance through air traffic control, and because time margins are generally much larger than in the driving domain. Yet even in aviation, the six pitfalls of automation are a concern.

The pitfalls of automation are more severe in the automotive domain than in the aviation domain. The requirements for automation algorithms (including underlying computer vision) are more daunting since the driving environment is less structured and more dynamic, and generally with much smaller temporal safety margins. Engineers are inclined to think that collecting more data will allow them to train smarter neural networks, resulting in more intelligent automated vehicles that will soon not require human input. This has been the narrative for the last two decades, often citing Moore’s law. Yet the actual exponential growth that takes place concerns the billions of dollars spent on creating supposedly driverless vehicles (Metz 2021). Today, there are no driverless vehicles that function always and everywhere; and how many times did drivers intervene to correct a dangerous situation? We still need human input (see also Noy et al. 2018), and traded control is not the right design approach for that.

Automation, such as AEB and automated evasive steering (AES), may indeed save lives and prevent safety-critical situations. However, such automation systems do not replace the driver always and everywhere, but constitute ‘envelope protection’ only for specific cases that are beyond human control bandwidth. AEB, for example, engages only at the last possible moment (when time-to-collision drops below, e.g. 0.8 s; Federal Ministry of Transport and Digital Infrastructure 2015; United Nations 2013), and the pitfalls of automation are therefore much less of a concern: No driver presses the gas pedal down and trusts the AEB system to maintain safety for them. Also, the AEB controllers are not useful for nominal car-following (see the definition of shared control in the Introduction, which stated that shared control concerns tasks either the human or the automation can perform). Therefore, such systems should not be confused with the current debate about trading or sharing control over lateral and longitudinal tasks, which either the automation or the driver could do.

Inappropriate human-automation interaction costs lives, as was dramatically made clear by fatal accidents involving Tesla and Uber, where ‘the automatic system misperceived a situation that a human driver would be very unlikely to misperceive’ (Noy et al. 2018, 70). Investigation reports on these crashes refer to the driver’s distraction and overreliance on automation as key causes (Pitfalls 1 & 5; National Transportation Safety Board 2019a, 2019b, 2020).

How many more crashes will occur for higher penetration rates of vehicles that apply traded-control automation? And what should guide our design solutions? Should we attribute these crashes to human error, or should we see them as the result of inappropriate human-automation interaction design? Consistent with the systems view on human error (Read et al. 2021), I argue the latter: These tragic accidents occur because traded control is used where it should not be, exactly because it produces the well-known associated pitfalls of automation. In fact, researchers have been talking about these pitfalls for many decades, such as Bainbridge (1983) in her work ‘Ironies of automation’: Why expect this to be different in the automotive domain?

In short, we should not apply the traded control paradigm to the driving domain, even if it is successful in aviation. The six pitfalls of automation will certainly be more pernicious in driving compared to aviation: automation limitations will be encountered more often and with much less time to respond, and drivers will largely remain non-professionals. So, if the problem is human-automation interaction design, what is the solution?

### 3.3. Haptic shared control as a solution

In agreement with many colleagues, advocate that when a driver is still ultimately responsible while interacting with automated systems, the driver and automation should not trade control but should share control (Abbink et al. 2012). In the automotive domain, the shared control approach has generated much research in design and evaluation (Flemisch et al. 2003; for reviews, see Abbink et al. 2018; Marcano et al. 2020).

Note that shared control and traded control are interaction design choices for exactly the same automation system (see Figure 1). As Grill Pratt of Toyota put it: ‘The technology that will eventually make self-driving cars safe can be implemented in human-driven cars’ (Ross 2018). If the driver is coupled to the automation via haptic shared control, the driver is
continuously informed and guided, leading to a more balanced workload (reducing Pitfall 3) but also allowing the driver to feel the automation limitations, thus resulting in better mode and situation awareness (Pitfall 1) and better trust calibration (Pitfalls 4, 5 & 6). This continuous engagement in the driving task can also be expected to mitigate the erosion of skills (reducing Pitfall 2). Accordingly, a shared control approach to human-automation interaction can be expected to be more resilient than a human-alone or automation-alone approach. Of course, this means that the driver cannot let go of the vehicle controls and engage in secondary tasks, i.e. the promise of fully autonomous vehicles. But to be clear, neither should drivers do this with traded control!

The driver must understand the automation, including its boundaries and system failures, and the automation must also understand the driver. In currently available traded-control automation, this understanding is achieved via warnings and other displays that aim to keep the driver’s eyes on the road or request the driver to intervene when needed. Furthermore, when a driver in the supervisory role prefers another action, the driver will have to adjust setpoints (e.g. setting different target headways in ACC), while the automation disengages as soon as the driver exerts a small control action. This non-continuous form of communication (i.e. warnings, changing of automation parameters, automation on/off) is characteristic of all traded-control automation (Sheridan 2011).

Haptic shared control, on the other hand, involves continuous force feedback on a control interface (e.g. the steering wheel) without decoupling the mechanical connection between the steering wheel and front wheels. It allows drivers to develop a feel for the driver-automation interaction, which may help in building up appropriate mental models of the automation capabilities and behaviours. An appropriate metaphor for this is the horse metaphor (Flemisch et al. 2003). In horse riding, the rider and horse have a dynamic relationship: They communicate via forces and tactile cues to understand each other’s limitations and to learn about each other’s behaviour over time.

In our lab, which started in 2002, we developed a haptic gas pedal as an alternative to the traded control approach of ACC. Using the functionality of the same automation system behind ACC, we used its inputs to translate them into forces on the gas pedal. This approach reduces control effort, increases safety margins to the lead vehicle, and allows drivers to respond to automation failures more quickly. Together with colleagues in the shared control community, this concept was extended to lateral control, exploring different design options, including adjustments of the steering wheel stiffness (e.g. Abbink and Mulder 2009; Saito et al. 2021). With haptic shared control for lane-keeping (commercially available as active lane-keeping or lane-keeping assist), drivers continuously feel the automation’s intentions to steer, which they can follow or correct, without prompting the automation to switch off. When drivers correct the force feedback, they

**Figure 6.** Example of the benefits of shared control on a road with mild curves (left panel) and upon automation failure (right panel). The left panel (adapted from Mulder et al. 2012) shows that haptic shared control on the steering wheel results in better driving performance (i.e. smaller maximal lateral excursions, averaged across participants) than manual control and automatic (= haptic shared control with hands-off) control. Note that no comparison with traded control is shown. The right panel (from Flemisch et al. 2008) depicts behaviour right after a silent automation failure that reverts the driver back to manual control just before a curve.
provide context-specific cues to the automation that can be used to update the reference trajectory (Scholtens et al. 2018). In this way, one enables a bidirectional channel of communication (Abbink et al. 2012), as well as a channel for reciprocal learning, which we have coined symbiotic driving (Abbink et al. 2018).

Does the concept of haptic shared control work? Many driving simulator studies have provided evidence that it does. For example, Mulder et al. (2012) found that haptic shared control in a steering task improves lane-keeping performance compared to manual steering (Figure 6, left panel). Note that the forces are relatively small, and when releasing the steering wheel, a relatively poor lane-keeping performance will result, although the vehicle still stays on the road see Figure 6, left panel. In short, when using haptic shared control, the joint system (i.e. driver and automation together) yields better performance than manual driving (driver alone) and automated driving (hands-off driving), i.e. the whole is greater than the sum of its parts. But these safety and comfort benefits in basic driving environments are not the main benefit of shared control. It is because the driver stays within the loop that automation pitfalls are mitigated. An example of this is illustrated in the right panel of Figure 6, where in the case of a silent automation failure right before a curve, drivers with shared control (green line) easily recover and stay on the road, while drivers with traded control (red line) get back in the loop too late.

In conclusion, haptic shared control is a design philosophy that impels drivers to stay in the loop by providing small safety and comfort benefits while mitigating automation pitfalls that occur when the automation’s limits are exceeded. Continuous engagement of the driver with the automation allows drivers to build a mental model of the automation, inherently augmenting mode awareness and situation awareness (Pitfall 1), avoiding deskilling (Pitfall 2), maintaining appropriate workload (Pitfall 3), and discouraging misuse (Pitfall 5). Behavioural adaptation (Pitfall 4) does still occur with shared control but can be anticipated and resolved in the interaction design. Additionally, the continuous communication avoids the issues with threshold settings that plague binary trading or alerts (reducing disuse, Pitfall 6) and engages the fast neuromuscular reflexes that allow quick responses also used in walking or biking. Moreover, because the forces constitute bidirectional communication, the automation can learn from corrections and adapt trajectories to reduce conflicts, which may prevent misuse (Pitfall 5) and disuse (Pitfall 6).

Note that the interaction paradigm of haptic shared control, like that of traded control, can be linked to any automation system that generates target steering inputs. Once the automation is reliable enough for automotive companies to take responsibility when the automation is active, traded control is the best solution. But as long as drivers are held responsible when automation fails and are required to be able to intervene at any time, shared control should be used.

4. De Winter: use traded control, not shared control

Abbink makes a caricature of the capabilities of automated driving and suggests that this technology is so dangerous that only a few car manufacturers seem to commit themselves to it. He fails to mention that companies such as Waymo currently deploy vehicles that do not even have a steering wheel, something regulators are already adapting to (National Highway Traffic Safety Administration 2022b), while Tesla’s beta-testing program demonstrates increasingly compelling automated driving in city environments. In the same vein, in Europe, Mercedes-Benz recently claimed to be the first to deploy a hands-off SAE Level-3 automated driving system on public roads, where under certain conditions, the driver can focus on other activities such as work or reading the news on the media display (Mercedes-Benz 2021).

Abbink further suggests that future driving should be like riding a horse, i.e. instead of removing the human from the loop, he wants to bring an animal into the loop. The horse metaphor, which was proposed by Flemisch et al. (2003) and which gained broader popularity through Norman’s (2007) book The Design of Future Things provides no suggestions about how the horse metaphor should be accomplished. It should also be noted that horse riding is difficult and dangerous, or as noted by Norman: smooth, graceful interaction between horse and rider requires considerable skill (25). In Camargo et al. (2018), various statistics are reported that indicate that horse riding has a high likelihood of injury and mortality, possibly the highest of all recreational activities. This is surely not what the authors of the horse metaphor had in mind!

Some of my arguments below resemble those from Sheridan’s (1995) paper: Human Centred Automation: Oxymoron or Common Sense? In this work, Sheridan argued that keeping the human in the loop is not necessarily a good idea, as humans may provide slow and unreliable input. It is exactly for this reason that...
operators of high-speed trains and nuclear power plants are excluded from the control loop in critical situations, ‘hands-off’ (Sheridan 1995, 824; see also Carvalho et al. 2008). Sheridan also hinted that human-centred automation is an oxymoron: ‘to the degree that a system is human-centred is it precisely to that degree NOT automatic?’ (823). Similarly, I indicate below that haptic shared control does not lead to a driver-automation partnership that is ‘greater than the sum of its parts’. Rather, it provides negligible benefits compared to manual driving while inheriting some of the same automation pitfalls that were discussed for traded control, and also giving rise to confusion and conflicts.

4.1. Shared control versus traded control in routine driving

The research so far on haptic shared control has been performed in car-following, lane-keeping, and other tracking tasks (see Xing et al. 2020, for a review), where it has revealed modest performance benefits. For example, a study in a driving simulator found that the mean absolute lateral position on a narrow road was 0.087 m for manual steering and 0.074 m for haptic shared control on the steering wheel (Melman et al. 2017). That is, participants, on average, stayed closer to the lane centre, indicating better driving performance. On a wider road, the mean absolute lateral position was found to be 0.23 m for manual steering versus 0.14 and 0.16 m for different types of haptic shared control (Petermeijer et al. 2015). Similarly, in car-following, Mulder et al. (2008) found a standard deviation of time headway of 0.23 m for shared control on the accelerator pedal versus 0.20 m for manual car-following. De Winter et al. (2008), on the other hand, reported slightly poorer car-following performance for haptic shared control (standard deviations of time headway of 0.49–0.55 s) than for manual car-following (0.48 s).

Although modest performance improvements can be expected, haptic shared control comes at a price. First, the evidence so far indicates that haptic shared control suffers from behavioural adaptation (Pitfall 4). A simulator study found that drivers, on average, drove faster (113.3 km/h) with shared steering control than without (105.7 km/h) (Melman et al. 2017). That same study also found that behavioural adaptation could be prevented by gradually decreasing the feedback torques when the vehicle was driving at a speed greater than a threshold speed. However, if behavioural adaptation (Pitfall 4) is indeed deemed to be a concern, then traded control in the form of intelligent speed adaptation (ISA) (Carsten and Tate 2005) could be enforced to make speeding impossible in the first place.

Furthermore, although deskilling is a known pitfall of traded control (Pitfall 2), haptic shared control may also result in deskilling. In the field of motor learning, there is a rich empirical base on the guidance hypothesis, summarised by Marchal-Crespo et al. (2010): ‘A number of studies have confirmed this hypothesis finding that physically guiding movements does not aid in motor learning and may in fact hamper it’ (205). A proponent of haptic shared control could argue that deskilling in haptic shared control is less serious than the deskilling that occurs with traded control. However, this statement is entirely speculative. In fact, the long-term effects of driving with haptic shared control may well be worse compared to traded control, because the driver makes transitions between manual driving and a car with varying force-feel characteristics.

Another concern is that of conflicts. The latest-technology vehicles on the market can perform automated lane changes (Lambert 2022), but how lane changes would work with haptic shared control is rather contentious. Suppose the algorithm plans to overtake another vehicle but the driver intends to stay in the lane, or conversely, the driver wants to change lanes while the algorithm thinks that staying within the lane is more appropriate. This scenario will lead to counter-torques on the wheel and possible annoyance (Pitfall 6). One may argue that conflicts can be prevented through clever driver modelling, such as an intent-inference module that predicts what action the human driver wants to take (Benloucif et al. 2019; Doshi and Trivedi 2011). However, the prediction of human intentions can never be perfect, which means that occasional conflicts are inevitable. An early study already recognised that ‘the counterforce on the gas pedal … was assessed by drivers as uncomfortable’ (Van der Hulst et al. 1996, 179). More recent research on drivers’ trust in warning and automation systems of commercially available cars found that lane-keeping assist (which applies torque to the steering wheel, equivalent to haptic shared control) produced relatively low trust ratings, with discomfort being one of the explanations (Kidd et al. 2017). It goes almost without saying that if drivers disengage the shared-control system (Pitfall 6), they cannot reap any benefit from automation intelligence, and the driving task becomes equivalent to manual driving. Traded-control automation, such as ACC, as well as warning systems
generally, received higher trust ratings than lane-keeping assist (Kidd et al. 2017).

In summary, haptic shared control suffers from automation pitfalls and yields only modest performance improvements compared to manual driving. Ironically, these performance improvements have been found in tasks for which automation is known to surpass humans. ACC, for example, can keep a near-constant headway, whereas automated lane-keeping ensures that the car stays neatly in the centre of the lane (Figure 7, left panel; also note that Figure 6, left panel, demonstrates that high guidance, i.e. hardly overridable haptic shared control outperforms human drivers). If performance improvements in lane-keeping and car-following were truly the goal, there would be no need to close the control loop through a human.

4.2. Shared control versus traded control in non-routine driving and accident scenarios

It was shown above that haptic shared control offers no meaningful benefits compared to automation when there is no need to take over control. What then are the advantages of haptic shared control? The proclaimed advantages are that haptic shared control helps prevent loss of situation awareness and overreliance (Pitfalls 1 & 5) and therefore prevents crashes.

First, it should be noted that haptic shared control may well cause a loss of mode awareness (Pitfall 1). The haptic forces may at times be below perceptual thresholds, for example, when the driver follows the target path closely, when having a low grip on the steering wheel, or when releasing the accelerator pedal, such as during coasting to a traffic light. Accordingly, the driver cannot just rely on the perceived forces to determine whether the system is engaged, risking mode confusion. Current haptic shared control systems, such as Volvo’s Pilot Assist, require status lights as well as warnings to ensure that the driver is keeping the hands on the steering wheel, which contributes to increased workload relative to manual driving (Solís-Marcos et al. 2018). Thus, haptic shared control currently on the market is not as continuous and reflexive as Abbink suggests it to be, but in fact prone to mode confusion (Pitfall 1) and workload issues (Pitfall 3).

It is true that loss of situation awareness (Pitfall 1) and overreliance (Pitfall 5) are critical pitfalls of traded-control automation: Fatal automation-related crashes with inattentive drivers have received extensive news coverage (e.g. Uber’s fatal crash with a pedestrian in Tempe, Arizona, in 2018), while within academia, there is a wealth of driving-simulator studies that have shown that some drivers are sometimes unable to reclaim control in time (De Winter et al. 2021; Zhang et al. 2019). Frequently cited is a study by Flemisch et al. (2008), which showed that participants driving with haptic shared control kept their vehicle in the lane while all participants driving with automation drove their vehicle off the road after a simulated sensor failure just before a curve (as also shown in Figure 6, right panel).
But how likely is it that drivers are misinformed about system limitations and unaware of their responsibilities, combined with a dangerous cocktail of a complete sensor outing at the most unfortunate moment that can be imaged (i.e. right before a curve, such as shown in Figure 6, right panel)? These situations are rare, and if they occur, receive formal scrutiny (e.g. National Transportation Safety Board 2017, 2019a) or in-depth studies with oncoming traffic (Mercedes-Benz 2021). Soon, more cars will be able to drive themselves on designated highway sections (Mercedes-Benz 2021). Soon, more cars will be able to drive themselves on highways, including changing lanes, without needing driver input. Will Abbink continue to argue: ‘that’s just low-level control, not true traded-control automation’?

Worldwide, 1.35 million fatal accidents happen each year (World Health Organization 2019), primarily because of human errors in attention, perception, and control, as well as violations such as speeding and running red lights (Stanton and Salmon 2009). It is for good reason that regulatory bodies enforce traded-control automation, in particular AEB (European Commission 2020). Continuous non-emergency shared-control-type technologies, such as lane-keeping assist systems, are not

In summary, traded-control automation is not as dangerous as Abbink suggests it to be. Serious automation-driving-related accidents are rare, and in the fatal accidents that have occurred, automation pitfalls were not the only cause.

Although it is undeniable that traded-control automation is one of the contributors to crashes, the exclusive focus on the dangers of automation is unproductive (De Winter 2019). The risks of traded-control automation should be compared with the number of lives saved relative to manual driving or haptic shared control. The power of traded-control automation was demonstrated by Bhardwaj, Lu, et al. (2020) for AES. In their simulator study, the automated steering was programmed to always avoid a collision with a pedestrian. To test the impact of automation errors, it was also programmed to cause a collision with oncoming traffic, called adversarial automation, or to not respond at all, called idle automation. The experiment showed that haptic shared control caused a substantial number of collisions, even when the automation worked as intended, because it still required valid driver input. At the same time, haptic shared control also resulted in collisions when coupled to idle or adversarial automation, because it required drivers to steer (against the guidance) (see Figure 7, right panel, for results for intended and adversarial automation). Granted, traded control resulted in more collisions during idle or adversarial automation, but in reality, how often would idle or adversarial automation happen compared to correct automation actions?

Suppose that the automation algorithms function reliably in 99% of the emergency cases and adversarially in 1% of the emergency cases. Based on Figure 7, the expected number of collisions for 1000 emergency steering cases would be 10, 55, and 228 for automation, high-guidance haptic shared control, and low-guidance haptic shared control, respectively. Even in an extreme case of automation that is only 75% reliable and crashes itself in 25% of the emergency cases, automation would still cause fewer crashes than haptic shared control. These numbers echo Sheridan’s (1995) statements that human-centred automation is an oxymoron; keeping the human in the loop preserves, or even causes extra, human errors.

Abbink argued above that AES and AEB should be excluded from this debate because these systems surpass human ability and do not suffer from the pitfalls of automation. Abbink’s argument constitutes a debating fallacy known as No true Scotsman (Flew 1977, 47), that is, protecting one’s argument (shared control is better than traded control) from a counterexample (AEB is superior to shared control) by trying to remove the counterexample from the debate. Abbink’s attempt to exclude AEB from the debate is reminiscent of the Al effect (Kaplan and Haenlein 2020), which says that humans tend to discount automation progress post hoc. For example, chess was previously considered the epitome of intelligence, but after a computer first defeated the world champion, critics argued: ‘that’s just computation; it’s not true intelligence’. In the same way, traded-control automation continues to advance, and driving has become increasingly automated over the past half-century. Cruise control initially took over speed control, while full-range ACC is now able to bring the car to a full stop. More recent technology permits drivers to engage in non-driving tasks on designated highway sections (Mercedes-Benz 2021). Soon, more cars will be able to drive themselves on highways, including changing lanes, without needing driver input. Will Abbink continue to argue: ‘that’s just low-level control, not true traded-control automation’?
among the evidence-based mandatory technologies, and why would they be? Haptic shared control hardly outperforms manual driving, as shown above.

Abbink also argues that while traded-control automation is successful in aviation, it should not be applied to driving. He points out that traded-control automation is suitable in aviation because the flight task is highly structured, with large time margins that allow trained pilots to resume control when needed. The actual reason the flight task is highly automated is necessity (it seems unsafe and highly impractical to let long-haul pilots fly manually along a precise flight path in a vast three-dimensional airspace) and economy (e.g. fuel efficiency gains can be achieved through automation). In driving, there are still major safety gains to be achieved through automation. In fact, driving is an excellent candidate for automation: the driving task is so dangerous, and causes of death are often so banal (e.g. distraction, running a red light) that it is surprising that manual driving is still allowed today.

In conclusion, shared control is an oxymoron. By keeping the human in the loop, one remains vulnerable to driver error and inherits some of the pitfalls of automation while also introducing nuisances in the form of conflicting torques on the steering wheel. Given the large number of human-induced traffic accidents, it seems a moral imperative to delegate control to automation where feasible. The pitfalls of automation should be studied with the help of vehicle data and remedied via technological advancements, just as these pitfalls are studied and remedied in aviation. Consequently, the operational design domain of traded-control automation keeps expanding, and road safety keeps improving. Although Bainbridge's (1983) Ironies of Automation suggest a doomsday scenario, the reality is that traded-control automation is already upon us.

5. Petermeijer: some nuance in the debate

Abbink and De Winter make strong cases for their respective solutions to the pitfalls of automation in the driving domain. Although both agree that a fully autonomous vehicle is a desirable end goal, they strongly disagree on the best approach for imperfect automation. Abbink advocates haptic shared control, arguing that this provides comfort benefits in routine driving situations while mitigating automation pitfalls when the automation capabilities are exceeded. De Winter counters that such an approach is an oxymoron and that shared control retains human error while still suffering from automation pitfalls. How do these two seemingly rational scientists come to such different conclusions, and who is right? In order to answer this question, we need to look at their arguments a little more closely.

5.1. Convenient blind spots in the argumentation

Abbink and De Winter overlook important aspects in the attempt to make their case. In particular, Abbink does not mention that imperfect automation (due to sensor and algorithm limitations) will result in nuisances not only for traded control but also for shared control. This was made clear during a recent study conducted on a test track in which shared and traded control were compared for the same underlying control algorithm (Petermeijer et al. 2022). In this study, the car was guided along a target path closely past a parked car. During traded control, the response of drivers of the automated vehicle was to take over, which turned off the automation and put them in manual control. In the haptic shared control condition, on the other hand, the automation could indeed be corrected without switching it off, but this did result in drivers experiencing counter-torques on the steering wheel. Both situations can be expected to be annoying and could result in disuse (Pitfall 6), although for the experimental conditions studied, the shared control solution was preferred by drivers over traded control.

Moreover, Abbink argued that drivers stay engaged in the driving task through the physical interaction of haptic shared control, mitigating the impact of Pitfalls 1 and 3. Although this is plausible, there is currently little evidence to support this claim. It is possible that prolonged use of a haptic shared control system results in a driver who is physically engaged (hands-on) but mentally distracted (mind-off).

De Winter presents the argument that nearly all accidents are due to human error. However, it is a fallacy to use this statistic to argue in favour of automation. By current law, drivers are responsible for the control of a vehicle, so as long as the vehicle operates as designed, it cannot be blamed for accidents. Moreover, reducing the argument of automated driving to mere statistics ignores any ethical or societal debate regarding automation-induced crashes. According to survey research, self-driving vehicles will only be tolerated if they are considerably safer than human-driven vehicles (Liu et al. 2019).

Abbink and De Winter both present valid points in favour of their arguments but can also be accused of
cherry-picking certain results to support their case. I argue that it is not a question of which approach is better in general, but a question of when traded control and shared control can be used appropriately.

5.2. Towards a resolution

Abbink and De Winter make certain assumptions that provide indications on when trading or sharing should be used. For example, the assumptions of Abbink and De Winter about environmental complexity appear to be very different. De Winter appears to suggest that automation can function in most types of environments (or that it is only a matter of time until they do) and that automation drives in a way that is overall safer than manual driving. Abbink takes the stance that automation progress will be slow and, for the foreseeable future, will have a hard time dealing with situations that are moderately or highly complex. Abbink therefore believes that automation requires human correction in the form of haptic shared control.

If the environment is simple (or the automation is capable, i.e. accurate perception and prediction of the environment), then take-over requests can be provided early, i.e. with long time budgets. Early take-over studies (Damböck et al. 2012; Gold et al. 2013) already clarified that if time budgets are short, driver mistakes are common, which argues in favour of haptic shared control. Conversely, when time budgets are long (i.e. the driver has plenty of time to make a transition to manual), driver mistakes will be uncommon, which argues in favour of traded control. The difficulty, however, is that time budgets are not a priori known in scenarios such as urban driving. In urban driving, therefore, shared control, or traded control where the human keeps the hands near the steering wheel at all times, will be required.

It seems sensible to conclude that traded control is preferred in relatively structured environments and when proper safeguards are in place (e.g. the vehicle can bring itself into safety). In these conditions, it is unlikely that drivers will have to take control in safety-critical conditions, and the adverse consequences of the automation pitfalls will be small. Haptic shared control, on the other hand, is preferred when system limits are ill-defined and when the environment is more complex and dynamic. In these cases, the in-the-loop driver can act quickly upon automation failure.

These observations are in line with Parasuraman et al. (2000), who suggested that ‘automation can be applied across a continuum of levels from low to high, i.e. from fully manual to fully automatic’ (286). These authors explained that higher levels of automation (cf. traded control instead of shared control) become justified (1) when ‘human performance consequences’ (i.e. the automation pitfalls) are expected to be small, (2) when automation is reliable, and (3) when the cost of decision outcomes is low. Parasuraman et al. further pointed to a relationship between automation reliability and human performance consequences, where, for example, complacency (Pitfall 5) is more likely when the automation is highly (but not perfectly) reliable. This echoes the ‘lumberjack hypothesis’ of Wickens et al. (2020), which says that the better the automation performs in routine circumstances, the more severe the negative consequences when automation happens to fail or commit errors (‘the higher the trees are, the harder they fall’).

A representation of hypothesised efficacy of traded control, shared control, and manual control is shown in Figure 8. It clarifies that if environmental complexity is moderate, shared control is more appropriate than traded control because drivers can correct automation errors as they occur. The range of conditions in which shared control can be safely used is wider than that of traded control.
traded control, but in complex environments, such as in city centres with mixed traffic, even shared control will have to be disengaged.

Figure 8 further illustrates that traded control is only feasible in relatively simple environments. If the environment becomes complex, the driver may have to suddenly reclaim control, which may be dangerous considering the above-mentioned pitfalls. The latest traded-control solutions implemented thus far, such as Mercedes-Benz’s Level-3 automation, for example, operates on specific motorway sections up to 60 km/h to ensure that the automation functions reliably and take-over requests can be provided 10 s in advance (Mercedes-Benz 2021). Similarly, the mandatory use of AEB can be explained because environmental complexity is low (i.e. it is relatively easy to establish that a crash is about to occur and that the best action is to brake strongly). Furthermore, the cost of acting is low as well (in fact, not braking autonomously guarantees an adverse outcome).

The illustration shown in Figure 8 is, of course, still hypothetical. The shapes of the curves depend on automation functionality (that may improve over time). It can be anticipated that, as technology progresses, automation will be able to perform an increasing number of subtasks, and the shape of the traded-control curve will advance to the right. Shared control will also benefit from technological progress, although when systems can safely hand over control (such as envisioned in SAE L4 automation), the need for shared control will become smaller (but note the above-mentioned lumberjack effect; Wickens et al. 2020).

Future research should empirically evaluate shared versus traded control using a wide range of measures, ranging from objective performance measures to more qualitative outcomes. Think-aloud methods, for example, may clarify why drivers disable the automation (Kolekar et al. 2021; Pitfall 6) and whether they develop a proper understanding of system functioning (Pitfall 1). Furthermore, it is recommended that any driver-automation interaction approach (sharing or trading) should not just be evaluated in environments of low complexity (where automation is expected to surpass humans) but also in complex environments where the performance envelope may be exceeded. A key aspect of honest evaluation is that the automation algorithms that underlie each interaction approach (i.e. trading and sharing) should be identical (see Figure 1). Furthermore, data from real-life driving is needed to move beyond current arguments that are mainly based on the results of simulator studies.

6. Conclusion

The question of sharing versus trading is a fundamental one, first brought up by Sheridan and Verplank (1978). The current debate focussed on driving, a dynamic and safety-critical task in a relatively unstructured environment governed by formal and informal rules (Björklund and Åberg 2005). Consequently, driving can be seen as a task representative of a broader set of human-robot applications.

This article provided a series of arguments as to why shared control may be a powerful remedy to the pitfalls of automation. More specifically, shared control was argued to allow the driver to benefit from automation functionalities while remaining involved in the driving task. At the same time, this article provided a series of counterarguments, stating that drivers should not be kept in the loop because machines surpass humans in vital ways. Additionally, while tasks can be successfully shared between driver and automation, there is a risk of misalignment between driver and automation.

This work ended with the statement that it is a matter of environmental complexity. A thought experiment may prove the point: if the driving environment were very simple (or the automation perfectly reliable and capable), then the pitfalls of automation would not be an issue, and there would be no need for shared control. Conversely, if the complexity of the environment causes the automation to fail at unexpected moments, then the automation pitfalls are a concern and shared control is required to ensure that drivers complement the automation.

There are still several unresolved issues that need to be addressed in future research. In particular, there may be a qualitative or legal barrier when it comes to permitting the driver to perform non-driving tasks while driving, with some arguing that manufacturers should ‘skip Level 3’ (e.g. Ayre 2017; Weigl et al. 2021). Then again, Mercedes-Benz recently claimed to be the first company to have received certification for Level 3 automated driving (Mercedes-Benz 2021).

Some may argue that shared control and traded control are not mutually exclusive and should be combined. For example, haptic lane-keeping assistance (shared control) may be combined with automated evasive steering, thus benefiting from both philosophies. However, this type of unification may need further research regarding biomechanical safety and arbitration (i.e. the algorithmic negotiation of control actions of the driver and automation, see e.g. Marcano et al. 2020). Yet other types of philosophies to consider are maneuver-based control (Detjen et al. 2020;
Hakuli et al. 2011) and envelope protection, as well as if/how the driver should be able to override such systems—topics currently of interest in aviation automation (e.g. Van BaeLEN et al. 2020). In an attempt to integrate the different control philosophies that may exist, Flemisch et al. (2014) proposed a framework of ‘cooperative control’, which includes wireless communication, shared control, traded control (task delegation), and arbitration as core concepts (see MillOT and Lemoine 1998, for a similar framework). Whether and how to integrate different control philosophies is a worthwhile topic of further research. We hope this article provides a way forward for researchers and practitioners to critically reflect on the arguments for and against sharing and trading control for the automotive domain and between human and automation systems in general.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

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

Notes

1. In a preceding lecture, De Winter had explained the six pitfalls of automation. On the day of the debate, Abbink gave a 20-min presentation in which he explained that traditional (traded-control) automation is a problem in driving, and why haptic shared control is a suitable solution to the pitfalls of automation. This was followed by a 20-min presentation in which De Winter explained that traded control, not shared control, is the best solution. After a 15-min break, Abbink and De Winter engaged in a debate (25 min), which left the students in the room (who occasionally brought in arguments as well) somewhat surprised about the fact that two academics could disagree so much. Next, Petermeijer added nuance by means of his presentation (15 min), followed by closing remarks (5 min). The current paper is written using the same chronological order as the lectures, i.e. (1) review of pitfalls, followed by (2) Abbink’s perspective, followed by (3) De Winter’s perspective, and (4) a nuance and conclusion.

2. A mode error was also one of the causes of the Torrey Canyon oil spill in 1967. However, the term mode error did not seem to exist at the time; the incident was simply described as the master noticing that the lever was in the wrong position (e.g. Oudet 1970).

3. As with the term mode error, deskilling was recognised before the introduction of the term in the human factors literature. As early as 1951, Fitts anticipated that ‘tasks can be set up so that human operators eventually become deficient in certain important skills which are infrequently used. As an illustration, a pilot who relies too much on the auto-pilot may lose some of his skill in manual control, or one who routinely uses automatic landing equipment may lose his skill in making manual landings’ (10).

4. This phrase is used often; as early as 1954, it was written that ‘flying is made up of hours of boredom broken by occasional moments of stark terror’ (Buck, 1954, 4). Note that captain Buck cites incidences of engine failure and poor visibility, but the greatest source of danger he referred to was complacency, such as ‘not bothering to know procedures well enough’ (5), see also Pitfall 5. Hancock (2021) proposed the more ludicrous ‘months of monotony and milliseconds of mayhem’ (63).

5. The results shown in Bhardwaj, Lu, et al. (2020) concern a decoupled steering wheel. Another option is to use hands-on automated emergency steering, an approach that risks conflicting torques on the steering wheel; Bhardwaj, Ghasemi, et al. (2020) reported that ‘participants were oftentimes unwilling to let go of the steering wheel and give away the driving authority to automation’ (157).

6. Some authors reject this way of looking at driver error by arguing that a systems approach is more appropriate (e.g. Read et al. 2021). However, as pointed out by Reason (1999), an overly systemic perspective does not offer explanatory, predictive, or remedial value. Further, note that some have claimed that driving is ‘remarkably safe’ (e.g. Owens et al. 1993, 367) and that ‘drivers, in general, do a reasonable job of controlling their vehicles’ (Hancock and Stackhouse 1992, 494) because the number of fatal accidents per kilometer driven or per interaction is very low. However, this argument is not sustainable when considering that the lifetime probability of dying in a car accident in the USA is 1 in 101 (National Safety Council 2020), i.e. a substantial portion of the population dies by driver aberrations (e.g. distractions, misperceptions, violations) that could have been prevented by automation.

ORCID

J. C. F. de Winter http://orcid.org/0000-0002-1281-8200
S. M. Petermeijer http://orcid.org/0000-0002-8961-2884
D. A. Abbink http://orcid.org/0000-0001-7778-0090

References


