TO DELEGATE OR NOT TO DELEGATE: A REVIEW OF CONTROL FRAMEWORKS FOR AUTONOMOUS CARS

Dale Richards, Alex Stedmon
(dale.richards@coventry.ac.uk; alex.stedmon@coventry.ac.uk)
Centre for Mobility and Transport
Coventry University, Priory Street, Coventry CV1 5FB UK
(Tel: +44 02477 658946, 657390)

There have been significant advances in technology and automated systems that will eventually see the use of autonomous cars as commonplace on our roads. Various systems are already available that provide the driver with different levels of decision support. This paper highlights the key human factors issues associated with the interaction between the user and an autonomous system, including assistive decision support and the delegation of authority to the automobile. The level of support offered to the driver can range from traditional automated assistance, to system generated guidance that offers advice for the driver to act upon, and even more direct action that is initiated by the system itself without driver intervention. In many of these instances the role of the driver is slowly moving towards a supervisory role within a complex system rather than one of direct control of the vehicle. Different paradigms of interaction are considered and focus is placed on the partnership that takes place between the driver and the vehicle. Drawing on the wealth of knowledge that exists within the aviation domain and research literature that examines technology partnerships within the cockpit, this paper considers important factors that will assist the automotive community to understand the underlying issues of the human and their interaction within complex systems.

1 Introduction

With increasingly congested road networks the existing road infrastructure is unsufficient at meeting the growing and future demands that will be placed on it. Alongside this is a strong desire to improve efficiency and safety. At the centre of accident causality, human error remains a primary concern and advances in autonomous systems are hailed as the harbinger of a technology that can potentially reduce road fatalities in the future.

In the scope of this paper, the term autonomous system will be defined as
the quality of a technology that is able to perceive information from the environment and its ability to act upon it without human intervention.

With the advent of autonomous systems, what better way to reduce human error than by removing the human driver? The impetus behind an initiative such as this is directly related to the advances in technology that can assist in the management of the traffic infrastructure such as intelligent transport systems (ITS) or in-vehicle driver assistance systems such as advanced driver assistance systems (ADAS).

Several states in the United States (i.e. Nevada, Florida, Michigan and California) have reflected this growing appetite by passing legislation that allows the introduction of autonomous vehicles onto public highways. If we look across the current range of autonomous cars (Google, Toyota, Nissan, BMW, to name but a few) we can see they are all actively researching the integration of autonomous decision-making technologies. Although there are differences across these manufacturers in terms of their approach to integrating autonomous systems, they all have one thing in common: the driver who is ultimately responsible for the vehicle.

With the onset of smaller and cheaper sensors we have seen a migration of such technology transfer from other domains into the automotive community. For example, the development of Light Radar (LiDAR) was initially designed for uses in analyzing meteorological conditions (specifically cloud density). Modern LiDAR systems have been used in unmanned ground vehicles for detecting obstacles whilst navigating. Perhaps the best-known use of this within the automotive domain is the Google ‘Chauffeur’ car with its recognisable spinning LiDAR sensor mounted on the roof. At the moment this technology is expensive but there are already initiatives to produce a more affordable and mainstream version of this technology that could be integrated into other cars.

LiDAR is but one of the many different sensor technologies available that could be integrated within an intelligent automotive system. Within current ADAS functions, ultrasound technology is predominantly used for parking and proximity/separation such as adaptive cruise control (ACC), collision
warning systems (CWS) and driver awareness functions such as blindspot and intersection warning. A number of possible applications that sensors may be integrated into the vehicle are shown in Figure 1.

![Diagram showing possible applications of automotive sensors](image)

**Figure 1: Some available automotive sensor applications**

With these technologies employed to assist the driver, if we assume that ADAS functions such as intelligent collision warning/avoidance are integrated into the wider traffic network, how might these forms of automation actually support drivers?

There would appear to be two key ways in which the autonomous system could interact with the user. For example, an autonomous car will be able to respond to an event or situation that is perceived by the system as a potential threat (using on-board sensors) and either advise the driver on the appropriate action to take and place authority on the driver to respond; or the car will be authorised to take action on behalf of the driver in order to avoid an accident. Both cases highlight the need for a framework of delegating authority between the user and the system so that future solutions are developed with a common user-centred perspective.

2 Automation and human performance
The implication of incorporating an element of autonomy within a system predicates the delegation of authority, by the user, to the system. That is, the user who traditionally is seen as being in control of the system and ‘in-the-loop’ (Weiner and Curry, 1980) accepts that the system is performing certain functions either without their full knowledge (e.g. a ‘blackbox’ scenario) or whilst they adopt a supervisory role. However, this can lead to ‘out-of-loop’ situations where the operator is not fully-engaged in the task and may have diverted their attention to other activities but then be faced with taking back control at short notice and without fully understanding the current situation.

A certain degree of transparency must exist, which Norman (1990) argues, is the operator’s ability to understand the automated systems and ‘see through’ the system’s processes. Thus, the lower the transparency, the more removed the human is from the information processing which might have serious implications for situation awareness (SA).

There are many theories of automation that suggest that the human should always have the final say in any decision involving safety (Billings, 1997; Woods, 1989; Stanton, Harris and Starr, 2015). Such a stance represents a user-centred approach to automation, whereby the human always has authority over the decision-making elements within the system. However, delegation of control authority has been outlined in theories of adaptive automation (Parasuraman, et al., 1992; Inagaki, 2003) whereby the system is authorised to make certain decisions on behalf of the human. An existing example of this is the demonstration of automotive collision avoidance braking systems (Coelingh, Eidehall and Bengtsson, 2010; Isermann, Mannale and Schimitt, 2010).

The application of automation can be viewed in most domains as an attempt to reduce the workload burden of the operator whilst also offering a higher level of safety and efficiency. This is particularly valid in the aerospace domain, where over the last thirty years we have witnessed a revolution in automated flightdecks (Harris, 2011; Stanton, Harris and Starr, 2015). Of course, while there is a great deal of literature citing the benefits of increasing automation, there is evidence that highlights potential drawbacks. What we can conclude from the literature is that by increasing the level of
automation in an attempt to mitigate instances of human error, it may not eliminate it altogether. In fact what we are confronted with is a different type of human error borne out of the ironies of automation (Bainbridge, 1983). Again, we can look at examples in aerospace where incidents of automation bias (Mosier, et al., 1998) and automation surprise (Sarter, Woods and Billings, 1997) have been regarded as a confounding factor in many accidents. As a consequence, it has been argued that automation should take on tasks for the pilot rather than instead of the pilot and support, rather than take over from the pilot (Stedmon and Selcon, 1996).

For example, the tragic flight of Air France 447 in 2009 is testament to how a highly skilled flight crew can suddenly lose SA when a system is under automatic control. While cases such as these are rare, we are compelled to learn from them in order to assure that the same mistakes are not made again.

It is important to compare those piloting aircraft (who are generally highly trained and monitored, working in a sector that is closely regulated, and with technologies maintained to a high standard), operating aircraft worth millions of pounds and owned by an aviation company (a party who measures the pilot’s actions in the interests of profit and safety and who themselves will be under international scrutiny) with those members of the public operating their own vehicles with differing degrees of training, responsibility and levels of maintenance for their own cars. For example, young drivers in the fast moving, congested arterial roads during rush hour, who are using vehicles close to the end of their life cycles operate in a different context to those piloting aircraft.

The importance of providing the user with a reasonable understanding of what the system is doing (and why) is essential, especially in instances where a system failure or change in situation demands occurs with little notice for the user to engage with rectifying the situation. Much like humans,
systems can fail and are fallable (Reason, 2013). Therefore it is important that we do not stand in awe of such advanced systems but rather try to optimise the relationship in a safe and effective manner.

3 Frameworks for delegating control authority

Autonomous cars are sometimes, rather misleadingly, referred to as ‘driverless’ vehicles. It is not about taking control from the driver, but allowing them to delegate authority to the system in a manner that they understand and feel comfortable with. To facilitate the interaction between the user and the system a framework is required that defines the delegation of authority under a variety of different circumstances.

The traditional model for defining the levels of automation was put forward by Sheridan and Verplank (1978), and later revised by Parasuraman, Sheridan and Wickens (2000). This framework (presented in Table 1) provides ten levels of automation distributed between the user and the system. These range from the system making all decisions on behalf of the user (Level 10) to the human making all decisions (Level 1).

<table>
<thead>
<tr>
<th>Level of Automation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Fully autonomous: the automation system decides everything; acts autonomously, yet collaborating with other automation systems; ignores the human</td>
</tr>
<tr>
<td>9</td>
<td>The automation system informs the human only if the system decides to</td>
</tr>
<tr>
<td>8</td>
<td>The automation system informs the human, only if asked</td>
</tr>
<tr>
<td>7</td>
<td>The automation system executes autonomously and then necessarily informs the human supervisor</td>
</tr>
<tr>
<td>6</td>
<td>The automation system allows the human supervisor a restricted time to veto before automatic execution</td>
</tr>
<tr>
<td>5</td>
<td>The automation system executes a suggestion if the human supervisor approves</td>
</tr>
<tr>
<td>4</td>
<td>The automation system suggests one decision action alternative</td>
</tr>
<tr>
<td>3</td>
<td>The automation system narrows the decision choice selection down to a few</td>
</tr>
<tr>
<td>2</td>
<td>The automation system offers a complete set of decision/action alternatives</td>
</tr>
<tr>
<td>1</td>
<td>The computer offers no assistance, the human must make all</td>
</tr>
</tbody>
</table>
decisions and take all actions

It is possible to view this scale as a progressive range of delegation of control from the user to the system. There are various iterations of delegated authority between these two extremes and it thus provides us with a useful understanding of the type of interaction required.

Within the aerospace domain there is a variation of this, whereby a pilot may delegate authority to the aircraft to perform some preordained tasks. This is referred to as the Pilot Authorisation and Control of Tasks (PACT) framework (presented in Table 2). Bonner, et al. (2000) outline the different levels of delegated authority that can exist between a user (e.g. the pilot) and a system that may be either highly automated or autonomous.

Table 2: The PACT Framework (Bonner, et al., 2000)

<table>
<thead>
<tr>
<th>Level of Automation</th>
<th>Mode</th>
<th>Operational Relationship</th>
<th>Computer Autonomy</th>
<th>Pilot Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Automatic</td>
<td>Automatic</td>
<td>Full</td>
<td>Interrupt</td>
</tr>
<tr>
<td>4</td>
<td>Assisted</td>
<td>Direct support</td>
<td>Advised action unless revoked</td>
<td>Revoking action</td>
</tr>
<tr>
<td>3</td>
<td>Assisted</td>
<td>In support</td>
<td>Advice and, if authorised, action</td>
<td>Acceptance of advice and authorising action</td>
</tr>
<tr>
<td>2</td>
<td>Assisted</td>
<td>Advisory</td>
<td>Advice</td>
<td>Acceptance of advice</td>
</tr>
<tr>
<td>1</td>
<td>Assisted</td>
<td>At call</td>
<td>Advice only if required</td>
<td>Full</td>
</tr>
<tr>
<td>0</td>
<td>Commanded</td>
<td>Under command</td>
<td>None</td>
<td>Full</td>
</tr>
</tbody>
</table>

The PACT framework offers three basic modes of automation: fully automatic, assisted, and human command. Balanced against this are operational relationship, computer autonomy and pilot authority factors that provide a rich understanding of how different levels of autonomy can be
assigned to different tasks (ranging from routine processes to safety critical events).

Within the automotive sector there has been a similar push to address the levels of autonomy for driver-vehicle interaction. In the United States, the National Highway Traffic Safety Administration (NHTSA), a Government Agency concerned with producing and enforcing regulatory standards for the highways, has defined several levels of autonomous driving (presented in Table 3).

<table>
<thead>
<tr>
<th>Level of Automation</th>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Full self-driving automation</td>
<td>Vehicle able to perform all safety-critical driving functions and monitor external conditions</td>
</tr>
<tr>
<td>3</td>
<td>Limited self-driving automation</td>
<td>The driver is able to cede all safety-critical functions to the vehicle in some instances</td>
</tr>
<tr>
<td>2</td>
<td>Combined function automation</td>
<td>At least two primary control systems are automated in order to assist the driver</td>
</tr>
<tr>
<td>1</td>
<td>Function-specific automation</td>
<td>One or more specific primary control systems utilise automation</td>
</tr>
<tr>
<td>0</td>
<td>No automation</td>
<td>Driver in control</td>
</tr>
</tbody>
</table>

Using this classification we can clearly see that the majority of vehicle automation currently in development/operation (such as the Google system) may be viewed as adopting a function that is close to Level 3.

It is important to develop a better understanding of how a driver interacts with an intelligent vehicle. This must allow for different levels of autonomy to generate the flexibility for a driver to delegate different levels of control to the system at different times. However, this is a complex problem (illustrated in Table 4) even if we try to combine the elements of the previous models.

<table>
<thead>
<tr>
<th>Level of Automation</th>
<th>PACT Framework</th>
<th>NHTSA classification</th>
<th>Description incorporating</th>
</tr>
</thead>
</table>

Table 3: NHTSA classification of vehicle automation

Table 4: Combined elements of vehicle automation
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Fully autonomous; automatic; full self-driving automation</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>4</td>
<td>The automation system informs the human only if the system decides to; automatic; full self-driving automation</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>4</td>
<td>The automation system informs the human, only if asked; assisted; limited self-driving automation</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>3</td>
<td>The automation system executes autonomously and then necessarily informs the human supervisor; assisted; limited self-driving automation</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>3</td>
<td>The automation system allows the human supervisor a restricted time to veto before automatic execution; assisted; combined function automation</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>The automation system executes a suggestion if the human supervisor approves; assisted; combined function automation</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>The automation system suggests one decision action alternative; assisted; function-specific automation</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>The automation system narrows the decision choice selection down to a few; assisted; function-specific automation</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>The automation system offers a complete set of decision/action alternatives; assisted; function-specific automation</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>The computer offers no assistance, the human must make all decisions and take all actions; commanded; no automation</td>
</tr>
</tbody>
</table>

By comparing the elements from the different frameworks it is possible to identify areas of common understanding (indicated by the different zones of
shading in each column) and also to highlight where different interpretations of autonomy exist by those who might refer to one framework or another (indicated by different zones of shading across the rows). This representation of the complexity of the models only goes so far in helping us understand the scope of the problem. A key limitation is that it does not encompass the dynamic nature of automated processes where different levels of authority are required throughout a single driving experience.

There may be instances that dictate the driver having full control of the vehicle (simply to allow the individual to choose when they want to drive) or opportunities for the vehicle to be controlled by the autonomous system. This would either be seen as a benefit in the reduction of frustration or workload of the driver, or perhaps the potential for the autonomous system to act as the supervisor of the driver (e.g. a training and/or safety feature).

4 **Control delegation in autonomous vehicles**

Building on the established knowledge we propose a model (presented in Figure 2) that highlights the relationship between the user and the vehicle in terms of control and the delegation of authority.

---

**Figure 2:** Model of control delegation between the driver and the car
It is possible to categorise manual (Driver Authority), semi-autonomous (Adaptive Assistance), and fully autonomous (Car Authority) modes. The shift in terms of control is seen as the balance of interaction between the driver and the car and the dynamic changes based on what level of control (direct/indirect) is delegated from the user to the system. However, it is important to remember that in all instances the driver will be responsible for the safe operation of the car, regardless of the level of assistance that is engaged.

It can be argued that with increasing levels of automation or decision support available to the user, it is equally important to provide the user with a better understanding of what the system is doing (Bainbridge, 1983; Norman, 1990). The active monitoring of a highly automated system is cognitively demanding (Tsang and Johnson, 1989) and requires a high degree of vigilance on behalf of the user (Molloy and Parasuraman, 1996). In order to reduce the likelihood of human error it is important that the individual attains a sufficient level of SA pertaining to their situation and the context in which the system they are interacting with operates (Endsley, 1995; Endsley and Jones, 2012). Mental workload has also been cited as having a detrimental effect on human performance and safety (Tsang and Vidulich, 2006). The potential for a lack of vigilance has been linked to a number of accidents (Warm, Parasuraman and Matthews, 2008). Humans are poor at monitoring systems due to susceptibility of cognitive processing to switch off and miss stimuli where perceptual thresholds are low (Kantowitz and Sorkin, 1987).

In these circumstances, especially if the autonomous system operates in a way that is difficult for the user to understand (i.e. they are ‘out-of-the-loop’ or the system is not transparent but rather ‘opaque’), their mental model of the system is compromised. This is particularly important in terms of evaluating when a mode error is made in automated systems (Lankenau, 2001) and
also in terms of subtle changes to the control delegation that the user may not even be aware of.

Furthermore, in situations where the user is more of a supervisor of the automated vehicle rather than an active driver of the car, situations of mental underload can materialise. This is a particularly problematic as it is very difficult to identify when someone is suffering from underload (Lavie, 2010). If mental workload is reduced and SA is maintained then the issue of monitoring the system suddenly becomes a critical aspect in using the system (Saxby, et al., 2013; Young and Stanton, 2002).

To compensate for mental underload, the introduction of an interactive cognitive task can help to raise the effort required for the user to engage with the task both in terms of physiological measures of arousal and subjective assessment of alertness (Gershon, et al., 2011). By increasing cognitive effort, in terms of a secondary task, it is possible to maintain a level of attention that facilitates a degree of functional vigilance.

Traditionally adaptive decision support systems have been used to provide assistance to users who need to make timely (and sometimes) safety-critical decisions whilst under great task demand or mental overload. For example, if we consider an adaptive automation system on the flightdeck the pilot might welcome a decision support system that would monitor user physiological indices for symptoms of mental overload. Similarly, an adaptive system could also monitor for signs of mental underload and provide cognitive cues (akin to an interactive cognitive task) in order to maintain levels of vigilance and alertness.

A further aspect of a reliance on automation is that the reliability of such systems will degrade over time just as current mechanical ones do. The design of the failure track is presumably part of the process for the systems engineers and it is important to consider the autonomy lifecycles so that systems are future proofed and potentially incorporate principles of graceful degradation so that the entire system is not vulnerable to complete failure.
5 Discussion

With the advent of autonomous vehicles we are seeing a shift in the traditional role of the driver, but this does not diminish the driver’s responsibility; it merely changes how the driver interacts with the system and how vehicles are controlled. The majority of current use cases for autonomous cars place the user in the traditional setting of being in the driving seat in front of a steering wheel although the vehicle may well be operating without direct driver input (in essence ‘hands free’). However, that is not to say that the driver requires less opportunity to interact with the vehicle; in some instances we could argue that the driver requires more information of what is happening so that they maintain SA. As soon as the driver delegates control authority to the vehicle then this is more than a simple task shift, and a more complex interaction of trust, reliability and safety. In autonomous mode the driver no longer requires the traditional control interface with the vehicle. The placement of hands on the steering wheel and feet situated above pedals seems superfluous to the act of delegation. Indeed, when the vehicle is within autonomous mode the steering wheel and pedals act as means by which the driver may take control back from the system much like the way in which ADAS currently operates. However, there will still be a requirement for the driver to be supplied with appropriate cues for effectively monitoring and supervising the autonomous system. Some results have already suggested that users are willing to accept certain levels of delegated authority when it comes to safety. For example, Itoh, Horikome and Inagaki (2013) found drivers approved of a semi-autonomous collision avoidance system that would present the driver with an auditory tone before performing a safety manoeuvre.

Within the aviation domain the pilot and avionics systems interact to form a working team and just as a conventional team of humans operate, modern cockpits are characterised by trust in the system, functionality of team members, communication within the team, and where authority should be invested in the team (Taylor and Selcon, 1990; Stedmon and Selcon, 1996).
Taking examples from the highly automated flight deck there have been many instances of human error routed within failures of vigilance and a lack of SA. There are a number of psychological phenomena that have been cited as occurring in automated systems. These range from mode confusion, automation bias to automation surprise.

Providing an increased level of support to the user by introducing automation and decision support has obvious benefits in terms of reducing cognitive load and reducing some elements of human error. However, Kantowitz and Sorkin (1987) observed that increasing levels of automation can leave the user human acting as a simple supervisor of the automated process. Automation, in itself, may also requiring specific training and place new/additional demands on the user that direct task involvement may not.

The technology that will facilitate the introduction of the autonomous car has entered a phase of demonstration, with the Technology Readiness Levels (TRL) getting closer to market introduction. What is less mature is the associated understanding of how drivers will adopt to this new style of driving. We often view these systems as being intelligent and in some cases out-performing human abilities with little regard for the implicit nature of the sharing of the primary task and objective that in essence represents a shared goal between human and system (Baxter and Richards, 2010). On the occasion that the human is happy to delegate control to the system, thought is needed as to how to keep the user ‘in-the-loop’ in terms of maintaining SA. Good SA is essential not just for monitoring the system in terms of ensuring it is safe, but more so for predicate events that suddenly occur when there is a system failure or the system recommends or hands control back to the user. In such instances human trust in the system may very well lead to a dangerous degree of complacency (Bainbridge, 1983). As illustrated in other domains this is all too common and can lead to tragic consequences. This is why, for the foreseeable future, a driver of an autonomous car will be legally required to be paying attention to the road at all times (as is legally required in some of the US States that have already passed legislation).

In order to facilitate discussion and research around these points, the Model of Control Delegation provides a framework for understanding aspects of the
dynamic and complex interactions that need to take place between the user and the system in order for seamless and effective autonomous driving scenarios to be developed on our roads.

6 Conclusion

The use of an autonomous car is not about taking control away from the driver, but allowing the driver to delegate authority to the system. This changes the nature of the driving role with the driver adopting a more supervisory approach to monitoring an intelligent system. In order for this interaction to be effective it is important to design the system that allows the user to understand not only what the system is currently doing (and plans to do) but also what the system is not able to do. This builds a partnership of trust between the user and the system that recognises not just human limitations but combines these with systemic limitations in order to determine a user-centred socio-technical system for autonomous driving.

Acknowledgements

Thanks to Professor Don Harris (Coventry University) for useful comments on a later draft of this paper.

References


Norman (1990)


