ARTICLE IN PRESS

Safety Science xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Safety Science

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

Driver error or designer error: Using the Perceptual Cycle Model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016

Victoria A. Banks*, Katherine L. Plant, Neville A. Stanton

Transportation Research Group, University of Southampton, UK

ARTICLE INFO

Keywords:
Driving automation
Human error
Mental models
Perceptual cycle model
Schema theory

ABSTRACT

"Human error" is often implicated as a causal factor in accident investigation yet very little is done to understand 'why' such errors occur in the first place. This paper uses the principles of Schema Theory and the Perceptual Cycle Model (PCM) to further explore the circumstances surrounding the fatal Tesla crash in May 2016 in which the driver was fatally injured using team-PCM representations. The preliminary National Highway Traffic Safety Administration accident investigation concluded that the driver of the Tesla Model S was at fault. However, the analysis presented in this paper argues that rather than "driver error", the underlying cause of this tragic incident could be in fact more akin to a "designer error" implicating the design of the Autopilot feature itself. This is in line with the National Transportation Safety Boards more recent announcement that suggests systems design may have contributed to the crash. It would therefore appear that the drivers expectation of system functionality may not have matched the real life capabilities of the system. This is likely to be a product of inappropriate mental models relating to system function.

1. Introduction

The Society of Automotive Engineers (SAE) taxonomy of automation (SAE J3016, 2016) is a widely accepted industrial standard and defines the allocation of system function between the driver and automated subsystems. It ranges from Level 0 (Driver only) to Level 5 (Full automation). Level 2 (Partial automation) systems first entered the commercial marketplace in 2015 in the form of Mercedes Distronic Plus (Mercedes, 2016), Volvo's Intellisafe Autopilot (Volvo Cars, 2016) and most famously Tesla's Autopilot (Tesla Motors, 2016). These systems use a combined function approach that automate both longitudinal and lateral aspects of control as well as automating aspects of driver decision-making (Stanton et al., 1997; Eriksson and Stanton, 2017). Tesla's Autopilot specifically automates both longitudinal and lateral control, as well as being capable of performing automated lane change manoeuvres if requested by the driver. The impact of automated subsystems on driver behaviour has been extensively researched since the 1970s (e.g. Sherdian, 1970) but it is only in recent years that on-road trials have been conducted (e.g. Banks and Stanton, 2015; 2016, Endsley, 2017; Eriksson and Stanton, 2017). These have provided some worrisome findings fuelling concerns relating to driver trust (Walker et al., 2016), complacency (e.g. Parasuraman et al., 1993; Lee and See, 2004), their ability to resume control (Stanton et al., 1997) and to perform an extended vigilance task associated with 'partial autonomy' and beyond (e.g. Molloy and Parasuraman, 1996; Stanton, 2015). Despite the undeniable benefits of vehicle automation to improve road safety, deliver mobility to all, and reduce the number of accidents occurring on the road (Stanton and Marsden, 1996; Stanton and Salmon, 2009), human-automation interaction is often overlooked throughout the design process in the pursuit of growing functional capabilities (Schaefer et al., 2016) putting the safety of drivers and other road users at risk.

The first fatal accident involving a Tesla Model S, being operated in Autopilot mode, occurred on 7th May 2016. The vehicle collided with a tractor trailer that was crossing an intersection on a highway west of Williston, Florida causing fatal harm to the driver of the Tesla. The driver of the tractor trailer was unharmed. The National Highway Traffic and Safety Administration (NHTSA) commissioned the Office of Defects Investigation to conduct a full investigation of the circumstances surrounding this incident. The findings of this investigation were published in January 2017. Data extracted from the Tesla vehicle in question revealed that the vehicle was being operated in Autopilot mode, that the Autonomous Emergency Brake system had not provided any warnings or attempt to initiate an automated braking manoeuvre and finally, that the driver had made no attempt to override the Autopilot feature by performing evasive action. Overall, the NHTSA (2017) report did not identify any design defects that could have caused the collision to occur. Instead, the preliminary report concluded that "human error" was the primary cause of the incident and speculated

* Corresponding author.
E-mail address: v.banks@soton.ac.uk (V.A. Banks).

https://doi.org/10.1016/j.ssci.2017.12.023
Received 14 August 2017; Received in revised form 1 December 2017; Accepted 19 December 2017

that the driver must have been distracted from the driving tasks for an 'extended period'.

Whilst human error can be predicted using Human Error Identification (HEI) techniques such as the Systematic Human Error Reduction and Prediction Approach (SHERPA; Embrey, 1986) and the Human Error Template (HET; Stanton et al., 2009a,b), these techniques seek to predict and classify errors that may occur within complex systems rather than providing explanation. It is essential to understand why actions or assessments made by operators make sense at the time within the context of local rationality (Dekker, 2011). Local rationality refers to the operator goals, current knowledge and focus of attention (Reason, 1990; Dekker, 2011). Rather than "human error" being a terminology used to explain an underlying 'cause' of a failure, Dekker (2006) argues that "human error" should in fact be the starting point of an investigation and therefore "demands an explanation" (p.68). This is so that we can further understand how peoples decision-making and responses made sense to them at the time (Plant and Stanton, 2012). In order to do this, we must, to some extent, rely upon models and theories of human behaviour (Dekker, 2006). The Perceptual Cycle Model (PCM; Neisser, 1976) has previously been used to explain human error within a variety of high profile incidents including the Kegworth plane crash (Plant and Stanton, 2012), the Ladbroke Grove rail crash (Stanton and Walker, 2011), the Kerang rail crash (Salmon et al., 2013) and the Stockwell shooting (Jenkins et al., 2011). The PCM heavily emphasises the role of schemata arguing that human thought is closely coupled with their interaction within the world and therefore capable of exploring the principle of local rationality. This is something that current HEI techniques fail to address. As such, the authors use the principles of Schema Theory and PCM to explain the circumstances surround the recent Tesla 2016 crash within the driving automation domain. This novel approach paves way for new appraisals relating to accident causation.

1.1. Schema theory

Schema theory dates back to the early 1900s (e.g. Head, 1920; Piaget, 1926; Bartlett, 1932) and describes how individuals form mental templates of past experiences that can be used to influence their behaviour within the subsequent world. Bartlett (1932) introduced the concept of 'schema' and described them as active organisations of past reactions and experiences that could be combined with information in the world to produce behaviour. Similarly, Neisser (1976) describes a schema as an organised mental pattern of thoughts and/or behaviour that can help organise our knowledge and understanding of the world. Neisser suggests that embedded schemata belong in a hierarchical structure, a viewpoint echoed by Plant and Stanton (2012) whom suggest that our knowledge should be considered as networks of information that become activated through our experience of the world. When an individual carries out a task, schemata both affects and directs the way in which they interact with and perceive the information available to them in the world as well as influencing the way in which this information is stored for future reference (Mandler, 1984). This means that schemata can allow individuals to orientate themselves towards incoming stimuli and adapt their responses to it accordingly based upon previous experience (Bartlett, 1932). If the schema is appropriate to the situation, appropriate behavioural responses are produced (Stanton et al., 2009a,b). The initial triggering of a schema is a bottom-up (BU) process. This occurs when situations within the environment initiate the triggering of schemata that are based upon past experiences, expectations or interactions within the world. The process then becomes top-down (TD). Notably, BU and TD processes can occur simultaneously. Norman (1981) argues that if 'triggers' within the world are wrongly interpreted, maladaptive behaviour may occur. These culminate in slips of action or lapses in attention. Two types of schemata are proposed; genotype and phenotype (Bartlett, 1932; Neisser, 1976). Genotype schema reflect the residual structure of the mind that can go on to direct behaviour within the world. Genotype schema therefore act as the underlying template for our action responses. These templates have the possibility for continued development, but a key determinant of their development is interaction within the environment (Plant and Stanton, 2013a). In contrast, the phenotype schema reflects 'in-the-moment' behaviour and is exhibited through our action within the world (Stanton et al., 2009a,b). According to Norman (1981), there are three basic genotype schema-related errors that can be used to account for the majority of errors: the activation of the wrong schemata, failure to activate appropriate schemata and a faulty triggering of active schemata. All of these error types have been found to occur within the road vehicle environment (Stanton and Salmon, 2009).

1.2. Perceptual Cycle Model

The Perceptual Cycle Model (PCM; Neisser, 1976) is based upon the idea that a reciprocal, cyclical relationship exists between an operator and the environment in which they are situated. The PCM heavily emphasises the role of schemata arguing that human thought is closely coupled with their interaction within the world. This interaction can trigger existing schemata based previous experience and interaction that can lead to anticipation over certain types of information (TD processing). Previous experience directs subsequent behaviour that attempts to interpret information available to them within the environment (BU processing). Notably, environmental experience can modify and update cognitive schemata which in turn can influence future interaction within the environment hence the reciprocal, cyclical, nature of the model (see Fig. 1).

The PCM has been used as a means to explore systemic decision-making processes in the form of retrospective accident analysis (e.g. Stanton and Walker, 2011; Plant and Stanton, 2012). This makes the concept of construct validity particularly important (Annett, 2002). This is because in instances whereby first-hand accounts are not available, ergonomics theories must be used to propose valid explanations of behaviour post event (Salmon et al., 2013). Whilst the discussion surrounding reliability and validity is typically concerned with method selection, Plant and Stanton (2015) argue that reliability and validity are also important for in relation to theory. Whilst methods can be reliable without being valid, they cannot be valid without being reliable (Stanton and Young, 1999). This is also true for Ergonomics theories such as PCM. Plant and Stanton (2015) recognised that the validity of PCM had not been explored and so utilised the PCM to study aeronautical decision-making. They concluded that the PCM does indeed hold both construct validity and test re-test reliability and can be used for accident analysis with confidence.

2. Schematic analysis of the Fatal Tesla Crash

The role of the driver within automated driving systems has continued to be a contentious research area (e.g. Banks et al., 2018a). Whilst the Society of Automotive Engineers (SAE) have gone some way in standardising the definitions relating to different levels of autonomy, the remaining responsibilities of the driver have been left open to interpretation by regulators, manufacturers and drivers alike (Banks et al., 2018a). As the level of automation increases within the driving task, the driver and automated subsystems must coordinate their behaviour in order to ensure safe and normal driving practices (Banks et al., 2014). Taking a systems view, the driver and automated subsystems become analogous to 'agents'. 'Agents' in this sense refer to both human and non-human entities that can receive, hold and share information with others in order to achieve a common goal (Stanton et al., 2006; Salmon et al., 2008). Stanton et al. (2006) recognise that different agents may view their environment in a different way to other agents. The concept of Distributed Situation Awareness (DSA; Stanton et al., 2006; Stanton, 2016) explains how individual situation awareness may be compatible with the awareness of other agents involved