Design and Assessment of External Displays on Autonomous Vehicles for Pedestrian Safety

by

Zehao Qin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of

Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2019

AUTHOR'S DECLARATION

This thesis consists of material all of which I authored. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The most vulnerable group in road agents is pedestrians. In the Netherlands, from 2005 to 2009, the average lethality rate for vulnerable road users was 14 per 100 serious road injuries (SWOV, 2012). Prior to autonomous vehicles (AV), conventional vehicles had a human driver that could communicate with pedestrians through signals such as eye gaze, head movements, and hand and arm gestures. With the introduction of AVs, pedestrians can no longer rely on such communication signals. In the future, when all of the control and responsibilities of the human driver gradually transfer to the autonomous driving program, the vehicle's intent communication to pedestrians must evolve as well. The aim of this proposed research was to investigate the efficacy of different external human-machine-interface in communicating autonomous vehicle intent to pedestrians in crossing situations where negotiation between the AV and the pedestrian is required (i.e. jaywalking).

With SAE level 4 high automation enabled vehicles, what impact do external human-machine interfaces have on pedestrians' crossing behaviour? what impact do external human-machine interfaces have on pedestrians' general perception of AVs?

Three novel design concepts were created to fill the gap of the lack of visual experimentation with displaying the key mental model factors - external speedometer display of the vehicle, speed change indicator (i.e. decelerating/ accelerating), and gap estimation count down timer. The experiment was a within-subject design with 29 levels. The stimulus was structured into a 14 (design cases) x 2 (coloured vs. non-coloured) factorial design. A combination of iconography, text, anthropomorphic features and colour were compared and measured in perceived safety, urgency, usefulness, understandability, emotion comfort, as well as the influence on crossing decisions. A 100-person online study was conducted to understand the impact of external visual displays with high automation (SAE level 4) vehicles on pedestrians' crossing behaviours. The novel concepts open a new discussion for the perception of warning designs where the new visual concepts (i.e. explicitly displaying and varying the symbolism of speed) had strong performance across all measures.

Acknowledgments

First and foremost, I would like to extend my deepest gratitude towards Professor Catherine M. Burns for her support and sagacity and for always being the north star during this journey. Her composure and kindness helped me navigate the toughest of challenges of this thesis. Her commitment to my development both as a student and a researcher is truly appreciated. I am extremely lucky to have had the highest calibre of supervision.

I would like to thank Professor Shi Cao and Professor Siby Samuel for their role as members of this thesis' reading committee and their role as mentors for me over the past two years. Thank you for your insights and your teachings. The knowledge you passed on from your classes is embedded in every thought of this thesis.

Thank you to the friends at the Advanced Interface Design Lab. Thank you to Dr. Yeti Li, Murat Dikmen, Thana Hussein, Fan He, Yahui Wang, Umair Rehman, Ramtin Lotfabadi, Adam Linton, Reicilis Casares Li, Rodrigo Arcuri, and Harjot Singh Parmar. I appreciated every second of the time I spent learning from some of the most diligent, intelligent, and ambitious peers I've ever had the luck of meeting.

Thank you to Ming Ho Yip and Ding Yang for their unwavering support and optimism.

Thank you to the friends I've met along my journey in System Design Engineering.

Thank you to the University of Waterloo.

Dedication

This is for Hui Liu and Yi Qin.

Table of Contents

List of Figures

List of Tables

List of Symbols/ Abbreviations/Acronyms

ACC	Adaptive Cruise Control
ADAS	Advanced Driving Assistance Systems
AEB	Automatic Emergency Braking
ANOVA	Analysis of Variance of Variance
AR	Augmented Reality
AV	Automated Vehicles
AVIP	Autonomous Vehicles' Interaction with Pedestrian
BASt	German Federal Highway Research Institute
COA	Course of Action Development
DDT	Dynamic Driving Task
eHMI	External Human-Machine Interfaces
GPS	Global Positioning System
Hex	Hexadecimal
HIT	Human Intelligence Tasks
KPH	Kilometres Per Hour
LED	Light Emitting Diodes
Lenm	Ministry of Infrastructure and the Environment

LKA	Lane Keeping Assistance
MTurk	Amazon Mechanical Turk
MPH	Miles Per Hour
NHTSA	United States National Highway Traffic Safety Administration
OLED	Organic Light-Emitting Diodes
SAE	Society of Automotive Engineers
SAM	Self-Assessment Manikin
SDV	Self-Driving Vehicles
SWOV	Netherland's Institute for Road Safety
UNECE	United Nations Economic Commission for Europe
VR	Virtual Reality

1. Introduction

The minimum weight of a Tesla Model 3 is 1672 kg; at the residential one-way street speed limit of 40 km/h, or 11.11 m/s, the vehicle generates kinetic energy of around 103.2 kJ. The braking force of the strongest bone in the body, the femur, is approximately 4000 N. Without stopping, that's enough force to break every bone in the human body.

The most vulnerable group in road agents is pedestrians. In Netherland, from 2005 to 2009, the average lethality rate for vulnerable road user is 14 per 100 serious road injuries. (SWOV, 2012) In 2014, there were 4884 pedestrian fatalities in pedestrian-vehicle accidents in the US - a pedestrian was killed every 2 hours and injured every 8 minutes. (NHTSA, 2014) Prior to autonomous vehicles (AV), conventional vehicles had a human driver that could communicate with pedestrians through signals such as eye gaze, head movements, and hand and arm gestures. With the introduction of AVs, pedestrians can no longer rely on such communication signals. In the future, when all of the control and responsibilities of the human driver gradually transfers the autonomous driving program, the vehicle's intent communication to pedestrians must evolve as well. Efforts should be applied both on the implicit vehicle decisions as well as the explicit vehicle communication of such decisions to human agents involved, both internal and eternal of the vehicle. The research community has extensively explored critical out-of-loop issues with increased automation in vehicles such as mental overload, mental underload, over-trust, loss of situation awareness, and skill degradation of the driver. However, these issues are all focused on the human agents (i.e. driver and passenger) inside the vehicle. (Bellet et al. 2003) There is significantly less research effort focused on the human agents outside the vehicle, whose safety concerns should be addressed with equal levels of scrutiny as the former user group.

There have been a variety of design suggestions made. Designs such as a LED strip indicator, zebra crossing light projection, and various visual interface designs have been suggested. To date, no regulation has been made anywhere in the world that mandates AVs to adapt to any visual protocols in the scope of pedestrian communication. The pedestrian AV communication problem still remains in the early design research stages; hence, an iterative design cycle should be considered in order to rapidly understand the effectiveness of the designs and improve on promising concepts. However, it's difficult to compare findings of prior studies on the effectiveness of pedestrian AV interaction designs as most are limited in their design choices and effectiveness criteria. Designs are tested in one of the three following methods: a real-life prototype, a virtual reality (VR) simulated design, or a linear 2D visual design. Real-life prototype and VR simulated design are inherently time-exhaustive and require heavy efforts in creating stimuli conditions. The characteristic cost for such studies makes them suitable for testing and evaluation in the final phases of designing displays for an autonomous vehicle to pedestrian communication. Linear 2D visual designs could be effective in this early exploratory design stage to test and understand the effectiveness of different design elements.

1.1. Overview

The current research investigated the efficacy of different external human-machine-interface in communicating autonomous vehicle intent to pedestrians in crossing situations where negotiation between the AV and the pedestrian is required (i.e. jaywalking). With SAE level 4 high automation enabled vehicles, what impact do external human-machine interfaces have on pedestrians' crossing behaviour? What impact do external human-machine interfaces have on pedestrians' general perception of AVs?

Section 2. Background discusses the background necessary to understand the problem space that's scoped for this research: definition of automation, level and types of automation, levels of automation in autonomous or automated vehicles, human factor's point of view on potential problems AVs face, human-machine ethics of AV design, public acceptance of AVs, and automated vehicle and pedestrian interaction.

Previous studies found that (Li et al., 2018) when evaluating different external displays, pedestrians more commonly relied on their own mental model of their perceived vehicle speed and gap estimation instead of relying on the information being displayed. This study attempts to bridge the mental gap between the perceived vehicle kinematics and gap estimation to the actual information by exploring visual design concepts to explicit indicate vehicle speed and gap distance. Furthermore, studies (Lagstrom et al., 2015; Habibovic et al., 2016) also found that there were significant differences when pedestrians interacted with AVs instead of conventional vehicles. The hypothesis for this study will attempt to reaffirm that external displays can enhance the perceived safety and comfort levels of pedestrians when interacting with AVs.

Section 3. Design discusses the underlying design rationale and design research that was used to create the designs. Three novel design concepts were created to fill the gap of the lack of visual experimentation with displaying the key mental model factors - external speedometer display of the vehicle, speed change indicator (i.e. decelerating/ accelerating), and gap estimation count down timer.

Section 4. Experiment discusses the experiment design. The experiment was a within-subject design with 29 levels. There were 28 different stimuli and one baseline scenario with no designs included. The stimulus was structured into a 14 (design cases) x 2 (coloured vs. noncolored) factorial design. A combination of iconography, text, anthropomorphic features and colour were compared and investigated and measured in perceived safety, urgency, usefulness, understandability, emotion comfort, as well as the influence on crossing decisions. A 100-person online study was conducted to understand the impact of external visual displays with high automation (SAE level 4) vehicles on pedestrians' crossing behaviours. The novel concepts open a new discussion for the perception of warning designs where the new visual concepts (i.e. explicitly displaying and varying the symbolism of speed) had strong performance across all measures.

Section 5. Results break down the quantitative and qualitative analysis of the current mixed-method study. Section. 6. Discussion and Section 7. Limitation and Future Research discusses ideas and improvement that's inherent to the result of the current research. Section. 8. Conclusion synthesizes the findings of the current research for further research.

2. Background

2.1. Automation and Automated Vehicle

2.1.1. Definition of Automation

In a complex system, a human operator is required to process a large amount of information from dynamic input sources and make effective and safe decisions to meet system output requirements. As Moore's law unfolds on pace, the rapid advancement in technology development in software and hardware allows for many, if not all, aspects of the complex systems to become automated. However, automation does not equivocate to technological progress. For example, using a cryptocurrency instead of cash to purchase a product does not constitute as automation.

Automation is defined as a hardware or software device or system that performs a task that could be performed manually by a human operator. (Parasuraman & Riley, 1997) Automation replaces functions that humans do not want to perform or cannot perform as well as automation in terms of accuracy and reliability. In today's world, automation is pervasive in virtually every single industry as it has the potential to exponentially increase task performance and process efficiency while lowering economic cost. The following are a few prominent examples of automation in multiple fields. In naval defence, within a frigate command center, the radar and sonar operators' task of target identification (i.e. friendly, hostile, unknown, etc.) is automated in an attempt for reducing crew workload. In financial trading, automated trading algorithms process a large amount of financial data that was formally gathered and analyzed by human analysts. In aviation, features such as autopilot and flight envelope protection systems aid the pilots in the cockpit in their aviation monitoring and control tasks.

The same benefit crosses over into the vehicle domain as well. Automated vehicles (AV) have been shown to provide many benefits to road efficiency and safety. It has been postulated that an increase in automated vehicles on the road will decrease congestion, lower road accident rates, improve vehicle usage rate, improve traffic efficiency, and reduce energy consumption. (Anderson et al, 2014; Kuehn et al, 2009; Alkim et al, 2007; Bement et al. 1998; Stanton & Marsden, 1996;) As a result, there are a highly competitive research and development arms race in both the consumer software industry (i.e. Google, Waymo, Uber, Lyft) and original equipment manufacturers (OEM; i.e. Tesla, GM, Daimler Mercedes-Benz, Toyota) towards the development of increasingly advanced vehicle automation systems.

Many vehicle automated systems exist in today's commercial vehicles; technologies such as lane-keeping assistance (LKA), adaptive cruise control (ACC), automatic overtaking system, blind-spot monitor, automatic emergency braking (AEB), and proximity warnings are some of the most well-known automated systems. Stanton and Young (2004) break down vehicle automation in two groups: a system that *supports* the driver, and a system that *replaces* the driver. Visual enhancers such as blind-spot monitors, power-assisted steering, collision warning, and parking aids are classified under systems that support the driver. With driver support systems, there is no fundamental change in the driving tasks and the automation system acts as a performance enhancement aid to the human driver. Systems such as ACC, LKA, Global Positioning System (GPS), replace tasks that are traditionally performed by the human operator and change the tasks themselves fundamentally. For example, GPS automatically calculates a highly accurate geographical location of the vehicle and the driver through satellite triangulation, relinquishing the driver of the task of finding navigational cues and keeping track of their position themselves. The navigational task, with the assistance of GPS, transforms into a supervisory monitoring task instead of an information acquisition task. The system that replaces driving tasks is able to perform both longitudinal (i.e. acceleration and braking, speed control) and latitudinal (i.e. steering) tasks.

2.1.2. Level and Type of Automation

When discussing automation, two key concepts involved are the *level* and *type* of automation. The taxonomy of the levels of automation, shown in Table 1. Levels of Automation, Sheridan and Verplank, 1978, was introduced by Sheridan and Verplank (1978) and remains a seminal concept to date. There are ten levels; starting with full manual control by the human operator at level 1 and ending with full automation with no human input or supervisory control at level 10.

From a human factors point of view, the most interesting interactions and the hardest design problems occur in the middle range of the taxonomy from level 4 to level 7. These levels are where humans are interacting with partial automation, akin to communicating with a new team member. This fundamentally changes the nature of tasks instead of assisting in the original intended manual tasks. Function allocation becomes a new design problem for automation designers.

Specifically, the level of automation can be interpreted as the degree to which automation seeks confirmation and action from the human operator. At low levels, human operator owns the control of the task with the automation providing aid in enhancing the human operator's performance. At the higher levels, the full control of the task slowly shifts towards the automation and with the human operator gradually retired from the control loop.

Parasuraman, Wickens, and Sheridan (2000) extended the concept of four different types of automation, including information acquisition, information processing, decision making, and action implementation, for a model construct for the automated system designer. This concept of automation type has extended to a similar five-level model outlined by Kennedy and McCauley (2007), which further divides the levels of automation (see Table 2. Types of Automation, Kennedy and McCauley, 2007).

Table 1. Levels of Automation, Sheridan and Verplank, 1978

High	10	The computer decides everything and acts autonomously, ignoring the human.
	9	Informs the human only if it, the computer decides to
	8	Informs the human only if asked, or
	7	Executes automatically, then necessarily informs the human, and
	6	Allows the human restricted time to veto before automatic execution, or
	5	Executes that suggestion if the human approves, or
	4	Suggests one alternative
	3	Narrows the selection down to a few, or
	2	The computer offers a complete set of decision/ action alternatives, or
Low	1	The computer offers no assistance: the human must take all decisions and actions

Table 2. Types of Automation, Kennedy and McCauley, 2007

Type	Description
Information Acquisition	Acquisition and registration of multiple sources of information. Positioning and orienting of sensory receptors, sensory processing, initial preprocessing of data before full processing, and selective attention
Information Analysis	Conscious perception and manipulation of processed and retrieved information in working memory. Also includes cognitive operations (rehearsal, integration, and inference) occurring before the point of decision-making
Course of Action Development (COA)	Generating (1) the decisions that need to be made, followed by (2) formulating options or task strategies for achieving goals

Decision Selection	Selection of a particular option, COA, or strategy to carry out. Decision(s) is reached based on the Analysis stage (cognitive processing), the COA Development stage, and expertise (human or software)
Action Implementation	Consistent with the decision selection(s), carrying out the chosen option, COA, or strategy, whether through control actions at an interface or other means

In the domain of vehicle automation, three international authorities, including the German Federal Highway Research Institute (BASt; Gasser & Westhoff, 2012), the Society of Automotive Engineers (SAE, 2014), and the United States National Highway Traffic Safety Administration (NHTSA, 2013), have constructed their own classifications of levels of automated driving systems. The range of the automation range from no automation to full automation or fully automated vehicles. The most commonly used definition of the taxonomy used in vehicle automation is from the Society of Automotive Engineers (SAE), which NHTSA and BASt draw parallels with as well.

There exist six levels of automation, as shown in Table 3., that are specified by three criteria: dynamic driving tasks, dynamic driving task fallback user, and system capabilities or operational design domain. Dynamic driving task (DDT) breaks down into three primary driving tasks (Lu et al., 2016), longitudinal vehicle control tasks (starting, accelerating, and stopping), latitudinal vehicle control tasks (steering, lane change, curve driving), and environment and vehicle monitoring task. The first two driving tasks of longitudinal and latitudinal control, under the SAE taxonomy, are grouped together under sustained vehicle control tasks.

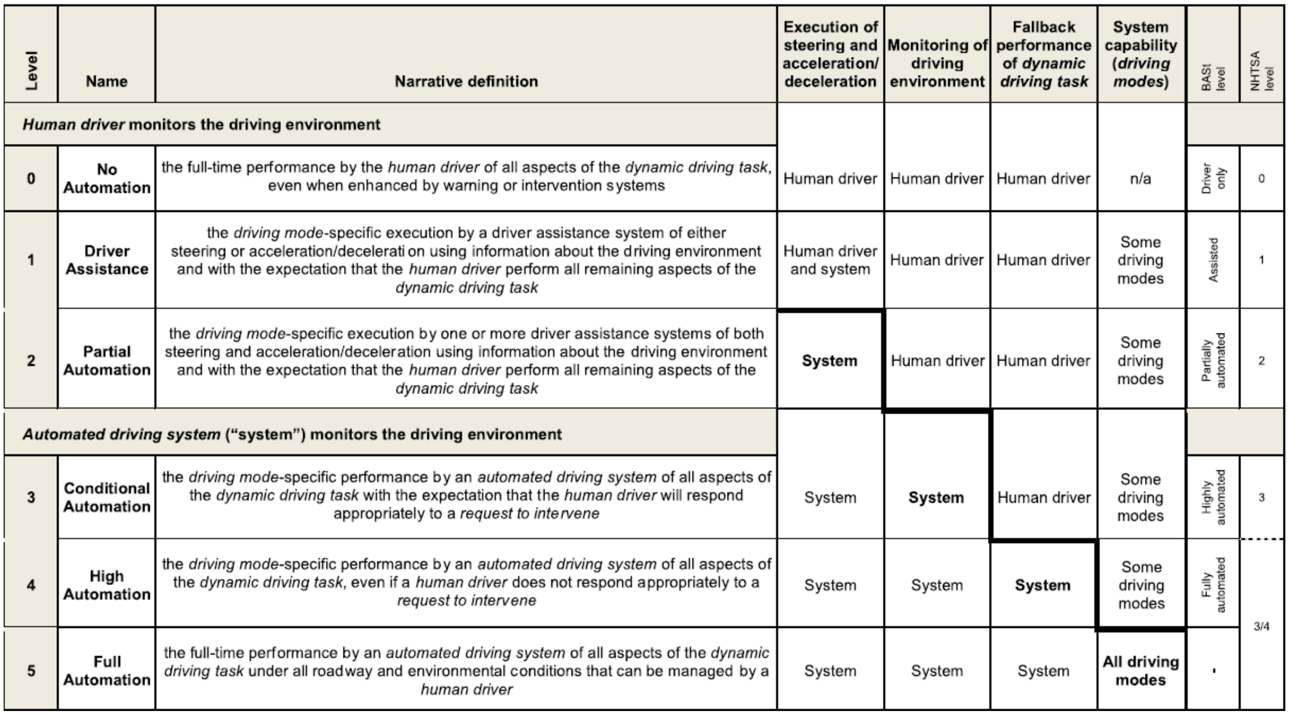
DDT fallback is defined by which primary actor, whether the human or the automated vehicle, is responsible for emergency backup to make a minimal risk decision when the vehicle does not perform its intended course of action, such as in the case of system failures. Operational design domain (ODD) is defined as the specific conditions, (i.e. high speeds, low speeds, all speeds) that the vehicle is intended to perform under. For example, adaptive cruise control can vary drastically based on the ODD of the design criteria.

The major differentiating line between human and system included this taxonomy is the agent responsible for monitoring the driving environment. The levels, in the order of increasing automation, are defined as follows: level 0 – no automation, level 1 – driver assistance, level 2 – partial automation, level 3 – conditional automation, level 4 – high automation, and level 5 – full automation. From level 0 to level 2 of driving automation, the monitoring task is delegated entirely to the human operator. From level 3 (conditional automation) to level 5(full automation), the monitoring task is delegated instead to the automated driving system. It's important to note that until full automation is achieved, the human driver still is required to maintain their driving capabilities and more importantly, be kept in the loop of the driving system. As predicted by previous research, the most interesting and hardest design problems for automation are the cooperation and function allocation between the human operator and the automation. Evident in the fallback column of Table 3., the human driver remains the responsible fallback actor up until level 4. Only at level 5 of full automation is the human driver removed from all control and monitoring tasks.

For example, Adaptive Cruise Control (ACC) would be considered as level 1 driver assistance. When ACC is combined with Lane Keeping Assistance (LKA) in vehicles, the vehicle would be considered as a level 2, conditional automation vehicle.

In January 2019, Ontario approved the road usage of public usage of commercial vehicles equipped with SAE level 3 – conditional automation technology. At the time of the current research study, level 3 still remains the current status quo for the most advanced system available and approved for road usage. Level 3 conditional automation, an upgrade from the combination of ACC and LKA, includes complete monitoring of the environment through sensor systems. The next frontier for the automotive industry is a major leap, as level 4 high automation requires the fallback agent of the automated system to be the system itself, further removing the human from the control loop. The technological challenge of reaching the next level isn't necessarily the roadblock to mass adoption. Level 4 high-level automation means that the human driver relinquishes complete control of the longitudinal driving task, latitudinal driving task, and environment monitoring task to the system. The human driver transitions into the role of a passive passenger with supervisory control tasks, where the human is still required to monitor the status of the system. The ethical implications when a system failure occurs and the legal and regulatory ramifications are more difficult issues to solve in comparison to the technological challenges for level 4 high automation to be approved for road usage.

Table 3. SAE classification of levels of automated driving systems, SAE, 2014



Level	Name	Narrative definition	Execution of steering and acceleration/ deceleration	Monitoring of driving environment	Fallback performance of *dynamic driving task*	System capability (*driving modes*)	BASt level	NHTSA level
***Human driver* monitors the driving environment**								
0	**No Automation**	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a	Driver only	0
1	**Driver Assistance**	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes	Assisted	1
2	**Partial Automation**	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes	Partially automated	2
***Automated driving system* ("system") monitors the driving environment**								
3	**Conditional Automation**	the *driving mode*-specific performance by an *automated driving system* of all aspects of the *dynamic driving task* with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes	Highly automated	3
4	**High Automation**	the *driving mode*-specific performance by an *automated driving system* of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes	Fully automated	3/4
5	**Full Automation**	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**	-	

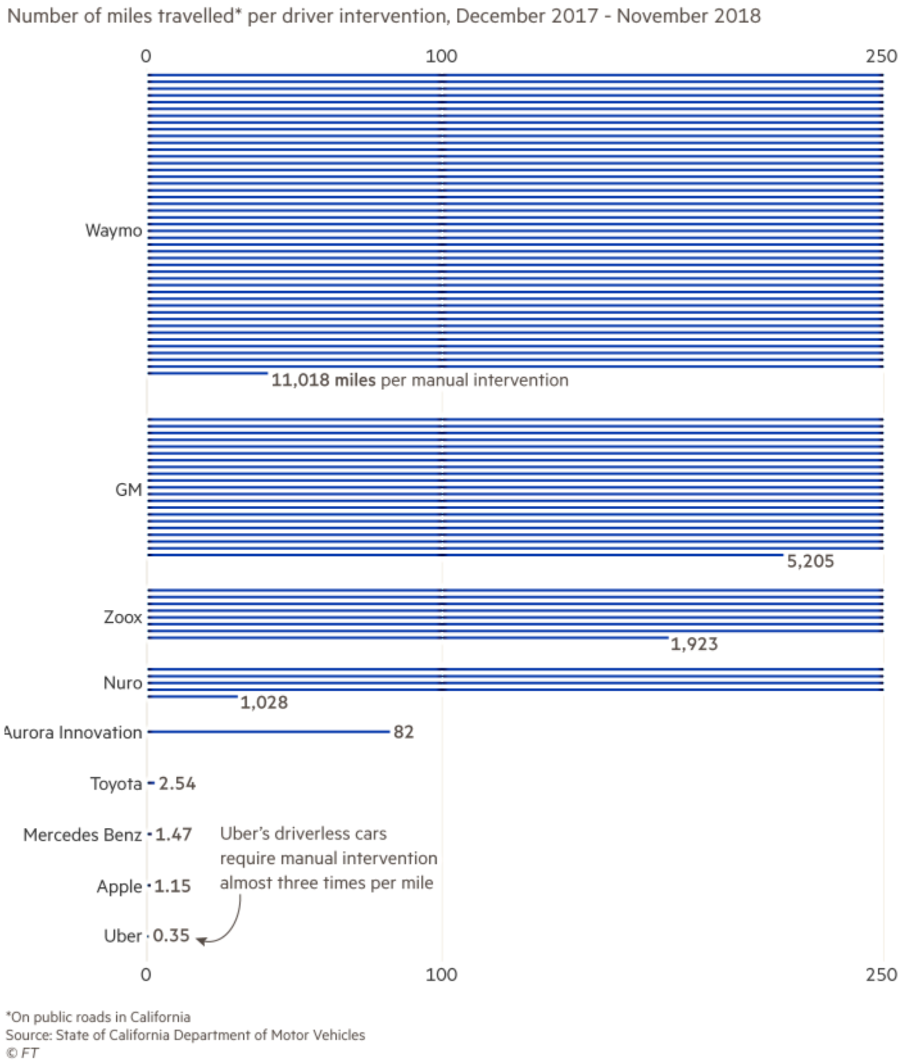
2.2. The Race Towards Level 5 Full Automation

A recent study from Kyriakidis et al (2015) suggests that the general public expects Level 5 full automation to be included in more than half of all vehicles by 2030. Tesla announced in April 2019 that level 4 high automation will be available in 2020 in all Tesla vehicles. Elon Musk, co-founder and CEO of Tesla, declared that steering wheel, acceleration, brake pedals, and other manual driving hardware features will gradually be rolled out and disappear completely as full automation become available worldwide. (FT, 2018) Waymo, the Alphabet subsidiary for the research and development of autonomous driving, announced in 2017 that it has been testing Level 4 high automation in Chandler, Arizona since October 2017. (FT, 2018) In fact, as the industry leader in developing the software for fully self-driving cars (level 4), Waymo has already begun testing of cars without any steering wheels and acceleration/deceleration pedals since 2018.

Considered an impossible feat a decade ago, the industry's view on level 4 and level 5 full automation are currently revolved around how, when, and under what condition can the actualization and adoption of the technology occur. In 2018, there were 48 automotive and technology companies that filed a testing report to California's Department of Motor Vehicles, including GM, Daimler Mercedes Benz, Apple, and Samsung (Financial Times, 2019). Indicated in

Figure 1., Waymo's autonomous driving division has already achieved an average distance per interruption of 11,018 miles in 2018. This is double its the previous year's report Waymo filed with the California Department of Motor Vehicles.

Two of key criteria that indicates the current maturity level of the automated vehicle system developed by different companies are the number of test miles driven on physical roads and the number of disengagements, both systemic and driver initiated, recorded during testing (manual interruptions required from fallback safety driver). In 2018, Waymo has accumulated more than 5 million miles tested and GM's Cruise has accumulated more than 1 million miles tested. (Forbes, 2018; Financial Times, 2018) Tesla has accumulated 1.2 billion miles on Autopilot, its in-vehicle automation software; however, it's important to note that Autopilot is only a level 2 partial automation program while other industry leaders such as Waymo and Cruise are testing Level 3 and 4 automation systems. (Electrek, 2018). Great strides have been made in vehicle-testing, but the road to a full understanding of the safety concerns is still far into the horizon. In 2016, the collision fatality rate in the US is 1.18 per 100 million miles driven. (USDOT, 2016) Before real-world test miles reach the 100 million miles mark, the jury is still out on the validity of road readiness and proclaimed safety benefits predicted by industry and academics experts alike.


Number of miles travelled* per driver intervention, December 2017 - November 2018

*On public roads in California
Source: State of California Department of Motor Vehicles
© FT

Figure 1. Industry leaders in the number of miles driven per manual intervention, FT 2019

2.3.1. Human Factors' Point of View on Potential Issues with Level 4 and Level 5 Automation

Several safety-critical issues arise from a human factor point of view as software and automotive industry leaders accelerate towards the goal of level 4 high automation - and eventually level 5 full automation. The research community has agreed that humans are not good at supervisory control tasks, where the task demands vigilance and sustained attention for a long period of time. (Warm et al., 2008) The upgrade of level 2 partial automation to level 4 high automation will see a shift of the human operator transitioning towards monitoring the system and appropriately intervening during perceived system failures and errors. The potential danger of the introduction of level 4 high automation is the concept of the operator being "out of the loop." Sven Beiker conducted an experiment where the drivers were blindfolded in an automated vehicle and examined their reaction time to return to an acceptable level of situation awareness and driving performance. Once the blindfolds were removed, it took an average five to six-second for the participant to assess the driving task and up to a minute to return to the same level of driving performance as a non-blinded folded driver.

In fact, Parasuraman et al. (2000) predicted that a high level of automation will result in decreased situation awareness, increased complacency, and skill degradation. This is particularly true in situations where automation is not "full," and the human operator must re-enter the loop; a loss of situation awareness can lead to errors, poor decision making, and awkward transitions as the human resumes the task from the automation. (Kaber and Endsley, 2004; Endsley and Kaber, 1999; Endsley and Kiris, 1995; Neubauer et al., 2012)

The past history from other industries, such as aviation, healthcare, and defence research, have shown that this loss of awareness has led to serious accidents. These accidents may occur in extremely small-time frames, where reaction time, situation awareness, and vigilance of the operator aren't sufficient enough to effectively intervene and prevent system failures. The catastrophic result of such failures includes severe human physical injuries, such as brain and spinal cord damages, and death of innocent passive human agents that's part of the system.

2.3.2. Public Acceptance and Human Machine Ethics of Level 4 and Level 5 Automation

As previously noted, the most difficult bottleneck issue of the broad adoption of fully automated vehicles isn't the technical issues of sensor superiority or algorithmic efficiency, but rather the geographical-dependent legal and regulatory changes and public perception of automated systems. This includes the shift in responsibility involved with the driving task as human drivers transition from being fully responsible and in control of the vehicle to passive passengers with zero control over the vehicle. (Lay, McHale, and Stevens 1996)

Continental AG (Sommer, 2013), in a survey spanning Germany, China, Japan, and the US found that although 59% of participants considered automated driving useful, 54% also responded they do not believe the technology will perform reliably. Howard and Dai (2014) found that of 107 participants in California, the least attractive element of automated driving, with 70% of participants strongly agreeing, is the liability concerns of the vehicle.

Kyriakidis, Happee, & De Winter (2015) found on a Likert scale of 5 (1 – disagree strongly, 5 – agree strongly) more than 65% of the participants strongly agree or agree that "the idea that fully automated driving systems may be introduced on a widespread scale worries me because of the general safety and reliability of such technology."

Goodall (2014) argues that the moral and ethical concerns stemming from automated vehicles should be addressed by the control algorithm in explicit and precise detail. In morally ambiguous situations, crash avoidance strategies should be carefully considered and trained through a multitude of real-life scenarios to build a database of recommended actions.

For example, if the situation arises where the vehicle has to make the decision to put itself and the passenger inside at a small risk to reduce the greater risk of other road agents, what should be the vehicle's predetermined course of action? Under normative ethics (what the individual or society should do instead of preferring to do), utilitarian, to maximize the cumulative benefit of the course of action, would minimize the risk outcome – the difference between the potential risk incurred to the vehicle and passengers within and the risk towards the road agents. This means either the vehicle does not change its course of action and the pedestrian is struck to protect the vehicle and passengers, or the vehicle or the passenger in the vehicle suffers damage - even if the result is as severe as quadriplegia - to protect the life of the pedestrian. Under descriptive ethics (beliefs and ethics as distribution among a society), Sandberg and Bradshaw-Martin have argued that the vehicle should act as moral proxies for the owners of the vehicle and adhere to the moral standards of the owner. This means allowing the owner to edit the vehicle preferences for ethical and moral decisions as an option embedded in the system. In the scenario above, whether the pedestrian is struck or not depends on the moral inclinations of the owners, whether the result is safe or not. Legally, there is no duty to act, in general, unless one is contractually obligated to rescue (e.g. paramedic), in a special relationship with the victim (e.g. parent-child), or responsible for the action (Goodall, 2014). Hence, vehicles legally don't have to protect the pedestrian from harm unless they are determined to have a special relationship. Moreover, aside from collision avoidances, there are intriguing yet profound ethical issues that stem from ethical thought experiments, such as the trolley problem and self-sacrifice, that must be solved with the realization of automated vehicles. (Lin, 2015)

If a pedestrian is killed by a fully automated vehicle, from a legal point of view, who's liable for the system failure? Is the owner of the system, the designer of the system, or the corporation responsible for the production of such a system? Issues like this demand governing and regulatory bodies to update the legal frameworks to further improve the public perception of fully automated vehicles. Network effect and economies at scale effect will only take place once adoption has reached a critical turning point and this requires the legal framework and current public perception to be significantly enhanced.

From a design point of view, the external intent communication with and monitoring of the vehicle is as significant as the implicit legal and ethical ramification of vehicle actions. Efforts should be applied both on the implicit vehicle decisions as well as the explicit vehicle communication of such decisions to human agents involved, both internal and eternal of the vehicle. The research community has extensively explored critical out-of-loop issues with increased automation in vehicles such as mental overload, mental underload, over-trust, loss of situation awareness, and skill degradation of the driver. However, these issues are all focused on the human agents (i.e. driver and passenger) inside the vehicle. (Bellet et al. 2003) There is significantly less research effort focused on the human agents outside the vehicle, whose safety concerns should be addressed with equal levels of scrutiny as the former user group.

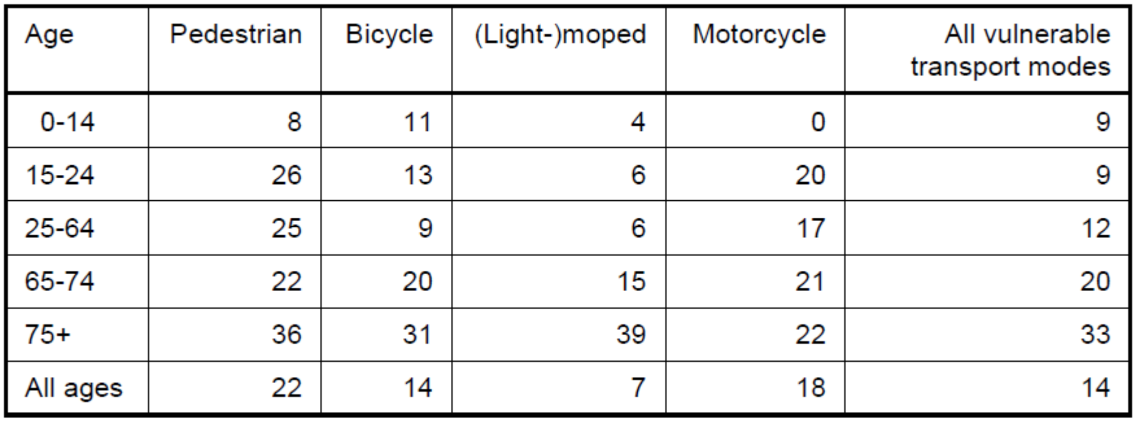
2.3.3. Automated Vehicle and Pedestrian Interaction

Wegman and Aarts (2006) classify pedestrians and cyclist as vulnerable road users. Specifically, task capability, amount of external protection (e.g. shell), and the difference in relative speed are the criteria used to define the vulnerability of various road users. For example, the inexperienced young pedestrians and the declining physical capability of the elderly pedestrians, in conjunction with a lack of external protection, identifies these two groups as vulnerable road users. The Netherland's Institute for Road Safety Research, SWOV, (2012) uses a subsection of the Wegman and Aarts' definition, mainly focusing on the level of protection and task incapability to define vulnerability. The current research uses this definition to identify the most vulnerable group of road users to understand its interaction with automated vehicles.

The vehicle protects the driver and passenger inside with its skeleton; hence, they are the least prone to injuries and fatalities. The lethality rate or case fatality rate (the number of fatalities per 100 serious road injuries) charted against age and mode of transport from 2005 to 2009 in Netherland is shown in Table 4. Lethality rate of vulnerable road user groups based on data over the period 2005 – 2009, sources: Ministry of Infrastructure and the Environment (lenM) and Dutch Hospital Data (DHD) (SWOV, 2012). The average lethality rate for vulnerable road user is 14 per 100 serious road injuries. For all age groups, the highest lethality rate is with pedestrians at 18.

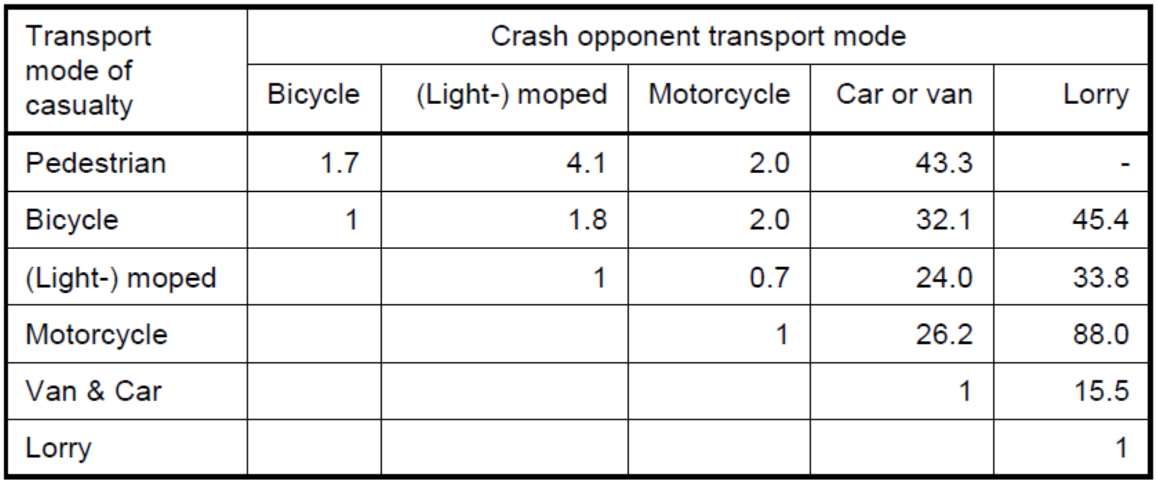
Another important factor to determine the vulnerability of road agents is the inequality factor. The inequality factor is defined as the ratio of the number of fatally and severely injured human agents between the two road users involved in the collision. Table 5. Inequality factor in serious two-vehicle crashes, 2005- 2009 (Sources: lenM and DHD). shows the pair-wise inequality factor of serious crashes between 2005 and 2009 in the Netherlands. The inequality factor between pairs of unprotected road users (i.e. pedestrian-bicycle collision, motorcycle-bicycle collision) is all far below 10. As soon as one of the road users is a car, van, or lorry, the inequality factor increases by a factor of 10. Again, pedestrian evidently is the most vulnerable road agent group consistently with the highest inequality factor against any other road user. The pedestrian-car inequality factor is as high as 43.3, which means for every fatality or severe injury for a driver or passenger inside the car, more than 43 pedestrians are fatally or severely injured. (SWOV, 2012)

Table 4. Lethality rate of vulnerable road user groups based on data over the period 2005 – 2009, sources: Ministry of Infrastructure and the Environment (lenM) and Dutch Hospital Data (DHD)



Age	Pedestrian	Bicycle	(Light-)moped	Motorcycle	All vulnerable transport modes
0-14	8	11	4	0	9
15-24	26	13	6	20	9
25-64	25	9	6	17	12
65-74	22	20	15	21	20
75+	36	31	39	22	33
All ages	22	14	7	18	14

Table 5. Inequality factor in serious two-vehicle crashes, 2005- 2009 (Sources: lenM and DHD)



Transport mode of casualty	Crash opponent transport mode				
	Bicycle	(Light-) moped	Motorcycle	Car or van	Lorry
Pedestrian	1.7	4.1	2.0	43.3	-
Bicycle	1	1.8	2.0	32.1	45.4
(Light-) moped		1	0.7	24.0	33.8
Motorcycle			1	26.2	88.0
Van & Car				1	15.5
Lorry					1

Furthermore, every 16 out of 100 fatal death on America's roads are pedestrians. The safety benefits predicted by researchers and industry experts were challenged by the recent case of the first pedestrian fatality from a collision with AVs in 2018. On March 18, 2018, in Tempe, Arizona. Uber's autonomous vehicle division was conducting testing of its "autonomous mode". The vehicle was travelling at 38 mph (61 kph) under the speed limit of 45 mph. Elaine Herzberg, who was a 49 years old female, was wheeling her bicycle on the side of the road and suddenly intercepted into the vehicle's path of travel and a collision occurred. Elaine passed away due to fatal injuries at the hospital after being rushed to the emergency room. The safety driver behind the wheels did not anticipate the collision did not take over manually. ("A pedestrian has been killed by a self-driving car", 2018) This is the first recorded pedestrian fatality in a collision with an automated vehicle, and unfortunately, it won't be the last. With the introduction of fully automated vehicles, the safety-critical issues that exist with vulnerable road users are now more prevalent than ever.

The dynamics of road users has evolved. The previous control and responsibility of the human driver shifted towards automated driving; however, the other road users have not changed as drastically in their behaviours. The most vulnerable road user remains the pedestrian; hence, the critical safety problem space of pedestrian and autonomous vehicle interaction remains of high interest from both the research and industry communities alike.

3. Design

3.1. Review of Relevant External Human-Machine Interfaces (eHMI)

In 2019, vehicles on the road still remain predominantly manually controlled by a human driver; as a result, driver-pedestrian communication has been widely explored and relatively well understood. The intent communication, especially in ambiguous situations that require negotiation, between the human driver and the pedestrian are communicated through non-verbal gestures. A combination of both vehicle signals, such as honking and high beam lighting, and human gestures, such as hand signals, head nodding, eye contact, arm-waving, are used to communicate the intent of the driver to the pedestrian to facilitate a safe and comfortable interaction. (Farber, 2016; Sucha, 2014) In the near future, as SAE level 4 high automation increases in road adoption, AVs may have to include eHMIs to augment the communication used today between human driver and pedestrians as the control and legal responsibility of the human driver decreases.

Many external human-machine interfaces (eHMI) has been suggested from both academia and industry; however, no conclusion has been agreed upon about the design principles that should be used. This section conducts a review of relevant eHMI designs from both industry and academia to understand the current design landscape.

From the Delft University of Technology, Bazilinskyy et al (2019) created a comprehensive literature review of the existing design paradigms for eHMIs for AVs to communicate with pedestrians. Designs from both the industry and academia were reviewed. It was argued that interface designs from the industry have higher public visibility compared to academic research on eHMI. Research papers are predominately peer-reviewed and receive a limited number of view counts. In contrast, online articles and promotion videos created by companies, receive millions of views due to the high public visibility that their brands attract. Knowledge is an important factor in enhancing the trust and acceptance of AVs. (Khastgir, Birrell, Dhadyalla, & Jennings, 2018) Therefore, higher exposure to the public leads to a better understanding of the state of affairs with automated driving. In turn, this facilitates an easier public acceptance of automated vehicles.

Understanding the strength and weaknesses of various design concepts speeds up the iterations of design cycles. It's difficult to analyze the industrial design of eHMI since there exists such a dynamic range of representations of the design concepts. 22 existing design concepts, in the form of a video, photo, or patent drawing, were measured (N = 1466) in the perceived clarity of each design's intent communication. Concepts from Google's Waymo, Uber, Nissan, Mitsubishi, Semcon, Volvo, Jaguar Land Rover, Daimler Mercedez-Benz, BMW, Ford, and Drive.ai were included. The design concepts were classified under four major themes: anthropomorphic vs. non-anthropomorphic gestures, textual vs. non-textual (iconographic) messages, egocentric vs. allocentric messaging (perspective), and text-colour congruency. It was found the design with the highest perceived clarity rating was found to be the combination of the text of "Go Ahead" and the anthropomorphic feature of eyes. Egocentric messaging was found to facilitate a safer crossing experience and were less ambiguous to allocentric messaging. In terms of text-congruency, green "WALK" was found to be the most persuasive of all the designs.

3.1.1. Anthropomorphic Gestures

Humans rely on nonverbal communications such as facial expressions, eye contact, and head movement in natural communications. (Bazilinskyy et al., 2018) Research in human-computer communication has found that a lack of nonverbal cues may lead to misunderstandings in the meaning and importance of the message. This hinders negotiation and may lead to critical safety concerns. (Kiesler et al., 1984; Hiltz & Turoff, 1978)

Anthropomorphism is commonly used in the domain of robotic design as robots with human-like features and behaviours, such as smiling and waving, were found to increase human agents' trust, likability, and willingness to respond to the robotic agent. (De Visser et al., 2016, Salem et al., 2011) Various anthropomorphic eHMIs have been suggested, including eye gaze (Chang et al., 2017; Pennycooke, 2012), smiling (De Clercq et al., 2019), and facial shape (Mahadevan et al., 2018; Mirnig et al., 2017). These displays are placed in the front of the vehicle - on the grille or as part of the windshield display.

Non-anthropomorphic eHMIs often use lamps or light bars to create a display that acts as a warning design. Bockle et al (2017) created an eHMI that consisted of four different states, indicating the intent of the vehicle to stop, not stop, waiting for the pedestrian, or starting to drive. Lagstrom and Lundgren (2015) created a horizontal light display bar on the front of the vehicle that shrank as the vehicle started moving and expanded as the vehicle prepared to yield. Lastly, De Clercq et al., (2019) created a concept where the light on a light bar moved from left to right. The effect of anthropomorphism on safety and intent communication are not clearly defined. Bazilinskyy (2019) found that of 22 different industrial designs, eHMIs with anthropomorphic features received median clarity ratings, and suggested that anthropomorphism may not be as convincing to the pedestrians.

3.1.2. Textual vs. Non-Textual Messages

Icons have been found to be more effective than text in communication as it overcomes the language barriers that exist in different languages (Bazilinskyy et al., 2019; Krampen, 1965) In traffic sign, icons are found to be more conspicuous, more legible at a distance, and have higher understandability compared to text. (Kline et al, 1990). In contrast, icons have also been found to take longer to interpret and are more prone to errors in communication (Huang and Bias, 2012). The effectiveness of an icon depends on semantic distance, concreteness, familiarity, and representation. (McDougall, Curry, & De Bruijin, 1999)

There are three different representations of the type of iconography: a literal representation (e.g., a male figure for male washrooms), an abstract representation (e.g., a bed for hotels), and an arbitrary representation (e.g. a cross for a hospital). (Lodding, 1983) Several eHMIs have included familiar icons in their designs, such as the walk/don't walk pedestrian sign (De Clercq et al., 2019; Fridman et al., 2017) a red upraised hand (Deb et al., 2018), and zebra crossing (Dietrich et al., 2018).

Bazillinksyy et al (2019) found that for non-textual/icon designs, familiarity and a literal representation were found to be more effective in clarity.

3.1.3. Messaging Perspective

It has been argued that people initially adhere to their own perspectives and an attempt to change people's perspectives is a time and effort-intensive task (Keysar, 2007). The perspective of messages in pedestrian communication can be classified as egocentric, where the vehicle communicates messages from the pedestrian's point of view, or allocentric, where the vehicle communicates from its own point of view. Examples of egocentric messaging include 'Walk' (Deb et al., 2016), 'Go' (Vlakveld and Kang, 2019), and ‘Cross now’ (Matthew et al., 2017). Examples of allocentric messaging include ‘Braking’ (Deb et al., 2016), and ‘Stopping’ (Nissan, 2015).

Ackermann et al (2019) argue that “direct instructions to cross the street are preferred over status information about the vehicle”. While Volvo (Volvo Cars, 2018) and Cefkin (2018) argued that displays for pedestrians should not give explicit instructions but instead communicate its current state and leave the decision making to the pedestrians. Bazilinskyy et al. (2019) found that egocentrically-biased messaging created a safer crossing experience and was less ambiguous in intent communication compared to allocentric messaging.

3.1.4. Text-colour congruency

Colour needs careful deliberation in the design of eHMI for AVs. The Stroop task paradigm (Stroop, 1935) has famously established that it takes longer to identify the colour of a set of words if the print colour of the word is incongruent with the semantics of the word (i.e. the word ‘red’ printed in yellow). This extends to colour-related words as well (Dalrymple-Alford, 1972). For traffic lights, it’s firmly established that red means stop and green means go. (Mulligan, 1936). The Stroop-interference patterns have been found to relate to traffic light colours and traffic instructions as well. Higher accuracy and quicker response rate were found when a colour is congruent with the position of the signage. (Bazilinskyy et al., 2019; Kandil et al., 2017). Bazilinskyy et al. (2019) found that green ‘WALK’ was the clearest and persuasive concept of the three different colour congruency schemes shown in the figure. 2.

Figure 2. Colour test congruency, Bazilinskyy et al., 2019

3.1.5. Key Literature for Experiment Design

The previous work from the lab on the AV pedestrian interaction problem space was an important influencing factor in the way the current experiment was designed. Li et al (2018) conducted an online survey using Amazon Mechanical Turk that investigated external displays for autonomous vehicles' interaction with a pedestrian (AVIP) at uncontrolled crossings. The study included 99 participants across the U.S. and Canada. The design conceptually mapped gap distance to a tri-colour scheme, akin to traffic lights, with light strips that spanned the front grill, the top of the windshield, and the side grill of the car. A VR environment was used to create the designs and the representation of the designs were shown to participants in a video format. The designs were divided by the Green-Yellow-Red and White-Red(flashing)-Red(constant) scheme that was used. The three colours correspondingly mapped to "safe to cross-zone", "warning zone", and "do not cross zone". The "warning zone" display (yellow and flashing red in the two design cases) flashed to increase salience while the other two remained constant.

The interface was evaluated on the warning design principle of "perceived urgency" as the displays changes according to the level of situational urgency. The participants answered survey questions that measured their perceived urgency level of the situation and their decision on whether or not to cross. Participants had higher levels of perceived urgency to the two designs compared to no warning design at all; however, both designs received similar responses with no major distinctions in perceived urgency results.

Interestingly, 76.3% of the participants responded that the primary decision factor for attempting to cross in front of the vehicle was overt kinematics of the vehicle (the perceived relative speed of the vehicle to the pedestrian) rather than the external designed display. This is similar to findings from a previous study by Clamann et al. (2016) where the researchers found that although participants believe external display should be included in autonomous vehicles, they mostly relied on their own mental model of gap estimation and inferences on the approaching speed of the vehicle for crossing decisions. Pedestrians were also found to be generally risk-averse at unpatrolled crossing sections and were more likely to wait for the vehicle to pass before crossing.

Fridman et al. (2017) created a formative evaluative assessment of 30 different designs for external autonomous vehicle-to-pedestrian displays. The researchers explored external visual display that could potentially replace traditional vehicle-to-pedestrian communication signals, such as eye contact and hand gestures, to ease the transition from level 2 partial automation to level 4 high automation. The designs were all on the front face of the vehicle, including the driver side of the windshield, the front headlights, the grill, and light projection in front of the vehicle. The design rationale and principles involved weren't explicitly mentioned. The designs split into two groups based on the criteria of safety. The two top designs that attempted to communicate that it was not safe to cross were one that projected a red hand icon and one that projected the letters “DON’T WALK” in bold red letters in a san-serif font. The two top designs that attempted to communicate that it was safe to cross were one that projected a green human walking icon and one that projected the letters “WALK” in bold green letters in a san-serif font. The top 4 designs were all projected on the driver side of the windshield.

The key takeaway of this study is in its evaluation method as a minimalistic and straightforward scenario design process. Akin to vehicle crash tests, real-world simulations that involve actual human pedestrians and vehicles of eHMI designs are a necessity before general road use for the highest level of assurance for critical safety levels. These research experiments are inherently time-exhaustive and require heavy efforts in creating stimuli conditions. The characteristic cost for such studies makes them suitable for testing and evaluation in the final phases of designing displays for autonomous vehicle to pedestrian communication. VR studies still encompass many of the benefits of an actual field study, however, they are still very effort intensive experiment to set up and design for choosing design concepts in early stages.

This study setup aims directly at creating an assessment that is fit for early-stage design evaluation, one that employs a low cost and low time intensive prototyping method for high flexibility and rapid design iteration cycles. The stimulus was the 30 display designs, all superimposed on to the same photograph of a single sedan in front of an uncontrolled zebra crossing where the driver side cannot be seen.

Figure 3. Most effective designs found in terms of perceived safety from Fridman et al., 2017

The designs were developed into short animations that composed of static animated images that flashed in a loop. Before showing the stimulus, participants read texts that explained the scenario context where a vehicle is approaching them as a pedestrian and decide if safe to cross. For each design, the participants were asked to consider if it was safe to cross and respond with “yes” or “no” or “not sure”. The study used Amazon Mechanical Turk to gain a wide-character array of global participants and benefitted from a rapid recruitment cycle. In several Turk runs spanning only hours, 200 participants were involved and they were paid $15 per hour as opposed to the usual $2 per hour.

3.2. Pedestrian's Mental Model of Crossing Decisions

A mental model is defined as theoretical models developed by individuals to understand and predict physical systems' behaviours (Gentner and Stevens 1983). Research has shown that there exists a multitude of factors that influence pedestrian's crossing decision making. (Habibovic et al., 2018). They include overt kinematics, vehicle speed, gap estimation between the vehicle and the pedestrian, traffic density and size of the gap between the vehicles, road features (i.e. geometry and signs), weather and light conditions, crossing speed, presence and behavior of other road, demographics of drivers and pedestrians, and experiences, knowledge, motivations, and cognitive state (Habibovic et al., 2018). Previously, the three key factors found to attribute to crossing decisions are gap estimation, inferences on the approaching speed, and overt kinematics of the vehicle. (Várhelyi, 1998; Sun et al., 2015; Schneemann and Gohl, 2016; Clamann et al, 2016; Li et al., 2017)

Despite a large number of design concepts suggested, there was little literature that investigated designs aimed at strengthening the key mental model factors that influence crossing decisions. Clamann et al. (2016) compared iconography and speed information dynamically displayed in a simple numerical format and found that neither displays had a significant impact on the crossing decision process. There were no studies found that explored the different explicit textual and non-textual representations (e.g. literal, abstract, and arbitrary representations) of speed, vehicle kinematics (slowing down or speeding up), and gap estimation.

As a result, the current study created three novel design concepts introduced based on the lack of visual experimentation by displaying the key mental model factors. An external speedometer display of the vehicle, a speed change indicator (i.e. decelerating/ accelerating), and gap estimation count down timer were included (Appendix A2). A combination of iconography, text, anthropomorphic features and colour were compared and measured in perceived safety, urgency, usefulness, understandability, emotion comfort, as well as the influence on crossing decisions.

3.3. Warning Design

Mapping the perceived urgency level to the situational hazard level is crucial for effective warning designs. (Baldwin et al., 2012; Chapanis, 1994) There are three main modalities for warning designs: visual stimuli, auditory stimuli, and tactile stimuli. This research focused on visual stimuli first. For visual stimuli, signal words or words choice, signal visual colour, and visual pulse rate are important signal parameters to consider during the design process. (Baldwin et al., 2012)

Annoyance is directly correlated with urgency where increased levels of perceived urgency mean increased levels of annoyance. When the signal is too abrasive or forceful or overloaded with too much information, the user may be annoyed and distracted. This may lead to impaired reaction time to critical events, decreased performance, and reduced trust levels in the system (Baldwin, 2012). The goal of an effective warning design is to find the optimal balance between the level of annoyance and level of urgency for the user through the design stimuli. A more urgent signal is tolerated or perceived as less annoying if the situation is critical enough to warrant such level of urgency.

Examples of high situational urgency include vehicle collision warning and heart rate flat-lining and examples of low situational urgency include email alerts and navigational cues.

The effect of warning design on memory, comprehension, and behavioural compliance has been widely researched. Wogalter et al., (2012) created a guideline for warning design and evaluation. Salience, word choice, layout, and pictorial symbols were important factors to consider during the visual design process. This process was used as a general guideline to validate previous designs as well as generate the novel designs of an external speedometer, gap estimation countdown, and speed change indicator. Understanding the task behaviour is crucial to designing interactive warnings; the mental model, especially the critical recall information and decision points, of the task elements and sequences, should be investigated and understood as aforementioned in the previous section on pedestrian's mental model.

Salience increases noticeability – the measure of the amount of attention the warning design will attract against another competing visual stimulus. Bold text type is suggested since they contrast against the background. The type increases the salience of the warning design as long as the stroke width doesn't obscure individual letters. All of the current design used bold letters with the font Highway Gothic. Furthermore, colour attracts attention and increases salience (Gill et al., 1987) and are perceived with higher readability. (Kline et al., 1993). This was the consideration that led to the investigation of colour versus non-colour designs.

Wording is suggested to be considered from a brevity and comprehensiveness point of view. Wogalter et al. suggests that wording should be broken down into four message components in the following order: it should attract attention, identify hazards, explain the hazardous consequences, and directions to avoid the hazard. Previous studies have shown that the most effective wording for crossing designs are “WALK/STOP” (Bazilinskyy et al., 2019) and this was included as the *de facto* word choice when creating the visual design stimuli included in the current study.

Layout is suggested to be considered from a visual decluttering sense. Visual clutter near the warning design significantly affects the warning detection time (Godfrey et al., 1991).

Pictorial symbols have been shown to improve the memory recall of warning information (Young and Wogalter, 1988). They are the most effective at communicating simple, direct concepts and less effective at communicating abstract concepts. (Wogalter, 2002; Murray et al., 1998) In terms of comprehension, there have been international standards set for pictorial warning design to meet for the general public’s safety. ISO 3864-2 (Organization of International Standards, 2016) and ANSI Z535 (American Nation Standards Institute, 2017) sets the minimum acceptable levels of comprehension by the general population at >67% and >85%, respectively.

Auditory warnings could be used as an enhancement signal when the visual signal is too crowded or there is simply too much information to convey using visual signals alone. None of the designs included in the current study was multimodal; each design only focused on the visual stimuli. Multimodal design using tactile stimuli (i.e. airflow through AC, mobile device vibration) and auditory stimuli (i.e. varying tone frequency, flash frequency of alarms, the gender of voice, levels of repetition) could be explored in the future through higher fidelity experiment setup using AR/VR headsets or real-life vehicular prototypes.

3.4. Colour and Non-colour Considerations

The colours used in today traffic system derives from a long line of colour experimentation that started with railroad systems in the 1840s in the UK. The colours used to represent "proceed" and "stop" have become standardized into the familiar system of red on top, indicating do not cross, and green on the bottom, indicating safe to cross, consistent with this idea (Mulligan, 1976). Winkileman et al (2016) found that information displayed in a familiar manner to the user is easier to process and understand. The general population of Canada and the U.S. has been publicly educated on the meaning of these colours in the context of street safety. As a result, these colours are effective as visual indicators for safety to the average North American. Colour hexadecimal (hex) code for red is #FF0000 with RGB value of (255, 0, 0), yellow is #FFCC00 with RGB value of (255, 204, 0), fluorescent yellow-green as #99FF00 with RGB value of (153, 255, 0), and green as #009900 with RGB value of (0, 153, 0).

The non-colour designs come from the need to test for the design impact from a pragmatic application point of view. As of 2019, The Ontario Highway Act (1990) does not permit the use of certain colours, such as green and red, to be emitted from lamps attached to conventional commercial vehicles. The non-colour designs remove all the banned lamp colours available for use as of today. Two colour groups were left in terms of the light colour, white light and yellow light.

According to Schreuder (1976), yellow and white headlamps were equally effective in regards to illumination for road safety, with yellow light causing less discomfort to the eye compared to white light. Current lamps are made up of light-emitting diodes (LED) and laser diodes (LD). Yellow light is created by filtering out the blue-violet portion of the light output of a lamp, and this reduces the luminance of the light by 15%. (Devaux, 1970) The reduction is mitigated or countervailed by the increase in visual acuity with yellow colour instead of white colour in weather with visual distractions (i.e. snow). (Bullough and Rea, 2001)

Under United Nations Economic Commission for Europe (UNECE), World Organization for Harmonization of Vehicles (WP.29), a global vehicle regulating body with 62 participating countries that focus on vehicle safety, requires new vehicles to be equipped with headlamps emitting white light. The same regulation is supported by the North American SAE standard J583. Previously, headlamps were permitted to be either white or selective yellow (hex: #FFBA00); in France, selective yellow was the mandatory colour to be used before 1993.

From a design aesthetics' perspective, in some cases, white denotes "emptiness". This emptiness does not denote a state of nothing. Rather, it is indicative of a state which will be populated with content in the future. White is a very useful form of communication as a "colour". (Hara, 2009)

As Ontario laws for lights emitters, or lamps, forbids the colours red, green, and blue to be used in the front of the majority of vehicles. Exceptions to the ban only include specialized vehicles, such as police vehicles and ambulances. Again, this restriction was considered during the functional requirement gathering process and a number of the design cases reflect and abide by the current regulations by only using white or yellow lights in the visual display. However, the impact of driverless vehicles has already prompted discussion for future changes in various rules and regulations of the traffic act; hence, for better understanding of the full impact of design factors, red and green colours were included to test its efficacy in communicating mode and intent of autonomous vehicles.

3.5. Gap Acceptance and Time to Arrival

The Highway Capacity Manual (TRB, 2010) defines critical gap as the time, in seconds, below which a pedestrian will not attempt to start a street crossing. Gap acceptance is defined and can be measured by the binary crossing decision of the pedestrian on the last moment to cross the road before the oncoming vehicle (Beggiato et al., 2017). Time to arrival (TTA) is defined as “the ability to estimate the time remaining before something reaches a person or particular place.” (Tresillian, 1995) To maximize the safety of pedestrian crossings, gap acceptance and time to arrival be equal. However, there’s a multiple of factor that influence TTA estimation and gap acceptance, including vehicle speed, vehicle size, weather and light conditions, road type and width, observation age, observation duration, as well as individual differences in age, gender, risk profile, and cognitive and motor abilities (Petzoldt, 2014; Beggiato et al., 2017). For example, higher speed influences pedestrians to make riskier decisions (smaller time gaps). It’s also been found that older pedestrians are more conservative in their crossing behaviours and size-arrival-effects – where larger objects appear closer to collision compared to smaller ones (DeLucia, 2013)– exist in time gap estimations. (Beggiato et al., 2017)

In spaces, such as a parking lot, with little or no formal communication (i.e. regulatory guidance in markings and signs in terms of the right of way), a higher need for informal communication is needed over. Informal communication includes the human-human interaction pattern that exists in current driver-pedestrian interactions, such as eye contact, hand gestures, head and body movements. Another important example of informal communication is anticipatory behaviour - small actions that make intention predictable to parties involved (Lagström & Lundgren, 2015). These actions help facilitate intent communication between driver and pedestrian. Example of anticipatory behaviours includes when the pedestrian slows down or speeds up, or when the pedestrian places a foot on to the path of travel of the vehicle. It's harder to comprehend informal communication signals than formal signal where the interpretation is dependent on situational context, culture, and personal understanding of social cues (Beggiato et al., 2017).

Both the intent of the pedestrian and the intent of the vehicle needs to be communicated clearly to avoid miscommunication-induced accidents and fatalities. The ambiguity in these negotiations requires additional communication channels as the control and responsibilities of the driving operator gradually decreases as vehicle automation level gradually increases. A visual display of gap estimation and time to arrival was the basis for the design of the count down timer. (figure of design)

From past studies, the average walking speed of a pedestrian was determined to be 1.2 m/s. The average critical crossing gap for pedestrians, from analysis of interlaced time gap and the distribution of the binary decision of crossing, was found to be 4.43s with a crossing width of 4m (Wang, 2010). Terwilliger et al. (2019), through a large-scale simulation and collection of 200,000 miles worth of real-world level 3 autonomous driving data, found that TTA was consistently underestimated at higher velocities and that TTA is adaptive for pedestrian as the vehicle dynamic changes (anticipatory behaviour of slowing down and speeding up). These heuristics were used during the design of the count down timer, where the timer starts at 6 secs at 150m away. The design itself was a low fidelity prototype where only speeds of 40km/h and under was displayed to the test subject. The count down timer, when moving onto the next stage of design evaluation and rebuilt in higher fidelity prototype in a VR/AR environment, will be programmed to dynamically change as the speed and distance changes from the pedestrian. The current research does not include this consideration and the variables and equations used to create the dynamic model of a count down display was not explored.

3.6. Symbolism of Speed

Applying the linguistic theory of pragmatics (theoretical definition of visual designs based on its functional purpose and the user's needs), a comprehensive review of the taxonomy and visual evolution of analogue and digital dials for speed reading was conducted by Mitchell (2010). The visual machine interface of the information graphics – the symbolism, visual variables (shape, size, colour, etc.), reference points, scale, and iconography used - for the reading of speed and speed-related status information was acutely explored.

For the current research, as the technological shift towards fully automated vehicles exacerbates, the user groups expand to include pedestrians as well. As aforementioned, from the point of view of pedestrians, a gap exists in the need for an external display of speed on autonomous vehicles. Understanding the visual language of speed and how speed has been presented in the past is the first step in the exploration for the external displays of speed included in this research.

Historically, speedometers are important to drivers in four main ways: avoiding being fined for speeding, to keep passengers safe, to change gears in manual shift vehicles, and to set the cruise control (Mitchell, 2010). Whereas for the car buyers and sellers, the design of speedometer is viewed alternatively as a branding tool for novelty or a display tool of the power and advanced technology inherent to the vehicle's engine. Often, designs of speedometers need to find the optimal balance between the safety and legal compliance concerns and the novelty and technological advancement the design shows.

From a design language perspective, analogue speedometers are multi-modally written – the modes include dial shape, scale marks, numbers, letters, dial needle design and colours (Mitchell, 2010). In parallel, the functional aspect of an analogue speedometer is realized through an acceleration sensor that acts on the principles of eddy currents, where the dial needle moves in relations to the speed of the vehicle. In Appendix C, a taxonomy of design features of speedometers from Mitchell (2010) is included.

There are many design features to consider for speedometer designs: dial shape, reference point placement, rotation direction, scales, typefaces, number of placements, needles, and colour considerations (Mitchell, 2010). For this research, dial shape, reference point placement, rotation direction, scales, and typeface were focused as key design features required to empathize with the needs of pedestrians as users.

The first analogue speedometer dates back to Schulze's patent of the eddy current speedometer in 1902; it was standardized in 1910 as a common feature in commercial vehicles. One of the earlier designs of a speedometer, shown in Figure 4. 1904 Cowey "Recording Speed Indicator" along with linear speedometer designs from the 1950s, is the 1904 Cowey speedometer. In terms of dial shape, the technology used led to the aesthetic corollary of a circular shape as the *de facto* speedometer shape. Functionally, eddy current – the electromagnetic force created by a change in speed in a conductor under the influence of an external magnetic field – is a centrifugal force; the mechanics would be more efficient if the force is rotational instead of latitudinal.

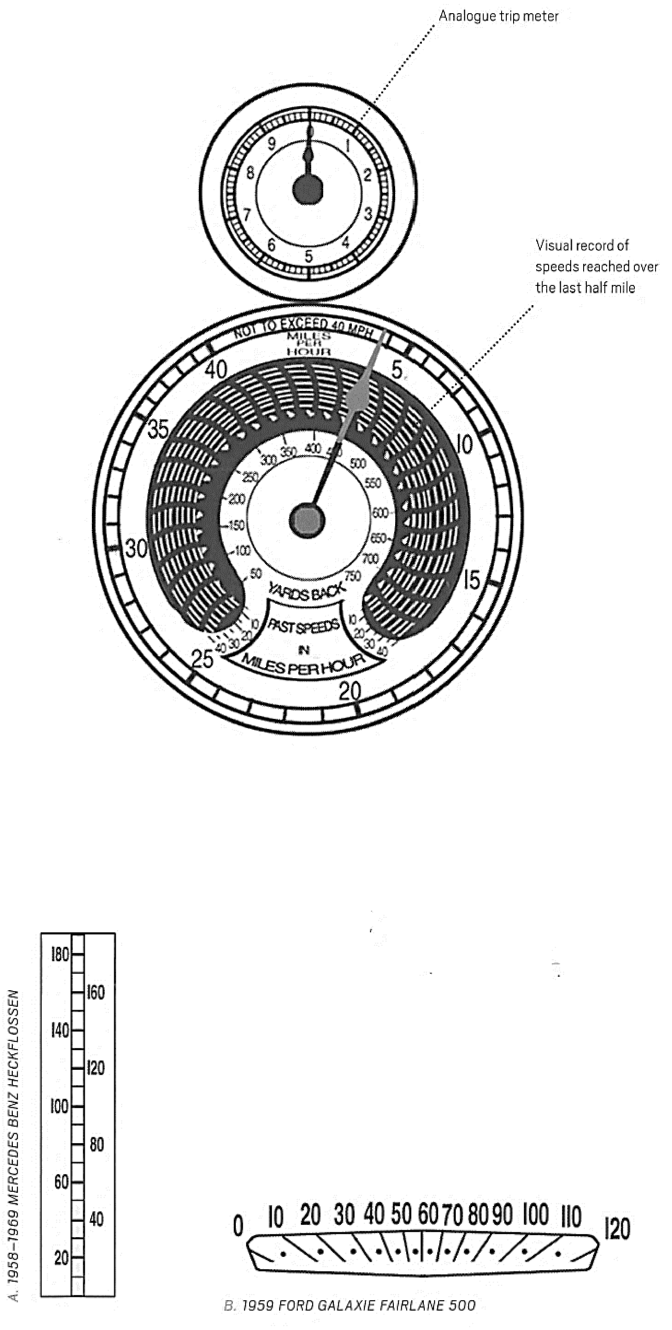


Figure 4. 1904 Cowey "Recording Speed Indicator" along with linear speedometer designs from the 1950s

Interestingly, that horizontal and vertical linear speedometer designs (figure 4a,b) was mass-produced by Benz and Ford in the 1950s. However, they were never popularized since they were less intuitive to read. Speed is only indicated by the position in a linear speedometer, whereas in a circular speedometer, two visual variables of angle and position are used to display speed. For the current study, the circular dial shape was used in designing the external speedometer for pedestrians, since the general public is highly familiar with the circular shape of speed displays.

Another crucial design consideration for circular dial designs is the location of the speed reference points. The reference speed that determines the starting and ending location for the reference points are initial, top, and maximum road speed limits (Mitchell, 2010). In the English language, speed is linguistically represented through a vertical dimension where speed increases are commonly referred to as "sped *up*" and speed decreases as "slowed *down*". Aligning the visual and linguistic representation, the most effective starting and ending points are placed at the 7:00 and 5:00 clock positions respectively (figure X). 7:00 and 5:00 o'clock positions also abide by a lateral visual symmetry between the where the reading order goes from the left to right hemispheres on the dial, allowing increase and decrease in speed to be visually balanced. Alternatively, if the dial had a starting point at 12:00, the visual representation would misalign with the linguistic representation of speed at the beginning of every trip. This is evident in the early designs. For example, in the Cowey speedometer, an increase in speed from absolute zero would lead the dial to move downwards (versus "speeding *up*") before reaching approximately 20 mph. This misalignment in the representation of speed could cause delayed reaction time, lower SA, and decreased vigilance levels.

Physically, the driver is viewing the speedometer on the instrument panel behind the steering wheel from a top-down angle. As a result, the most common road speed limits (60 – 120 kph) are placed at the top of the circle; allowing it to be the most natural position to read to abide by speed limits.

The current external speedometer design follows the best practices of today: the display is symmetrical in reference point placement and starts and ends from 7:00 to 9:00.

Another design consideration is the direction of movement for the dial rotation. Today, all speedometers abide by the conventional left to right, clock-wise movement – following the well-known pattern of clock movements.

For scales, the speedometer visual measurement indication is defined in terms of ratios (Stevens, 1951). Starting from the true zero reference point, scales have an equal increment in speed values in increasing or decreasing quantity. All measurements are equal, but the visual representation of the measurements doesn't have to be equal. At a higher speed, the speedometers marking could be halved in the space they occupation (Mitchell, 2010). For example, the speedometer embedded in the 2004 Audi S4 (figure x.), there two ratio scales included were used to represent speed, where higher speed is more compactly shown.

For the scale in the external speedometer design, 6 segments are shown in the out circular ring to augment visual understanding of speed and proximity. The segments follow the gap estimation count down timer and ideally would be 1 second for each increment. Since the average gap acceptance was between 4-5 seconds, 6 was used as a good numerical count for the number of segments. These segments are divided into three sections of red, yellow, and green to project situational urgency levels and safety levels to cross. The three sections are to mimic the well-establish pattern of traffic signal of red-yellow-green since information displayed in a familiar manner to the user is easier to understand. Oborne's reviewed the ergonomics of numerical increments and found that systems increment of 1s and 10s are the easiest to use. (1995) The current design uses a 1s increment where the speed is updated in real-time.

Figure 5. Current designs for external speedometer

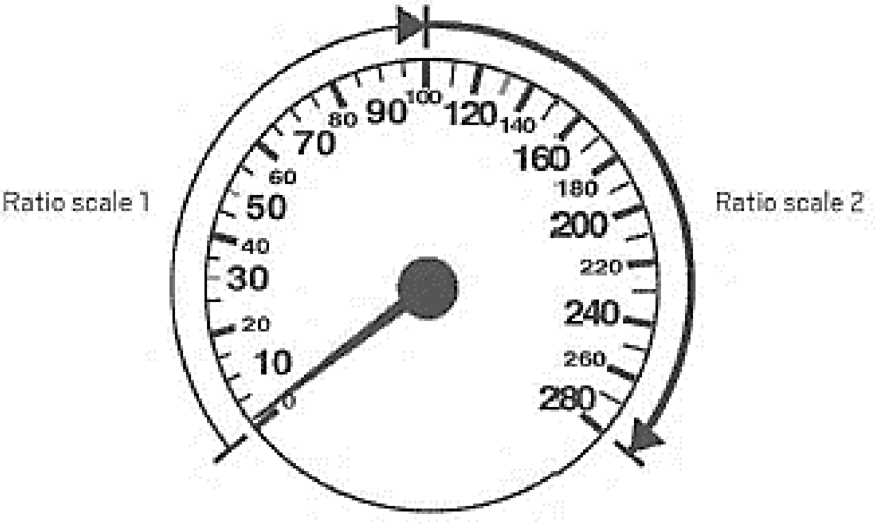


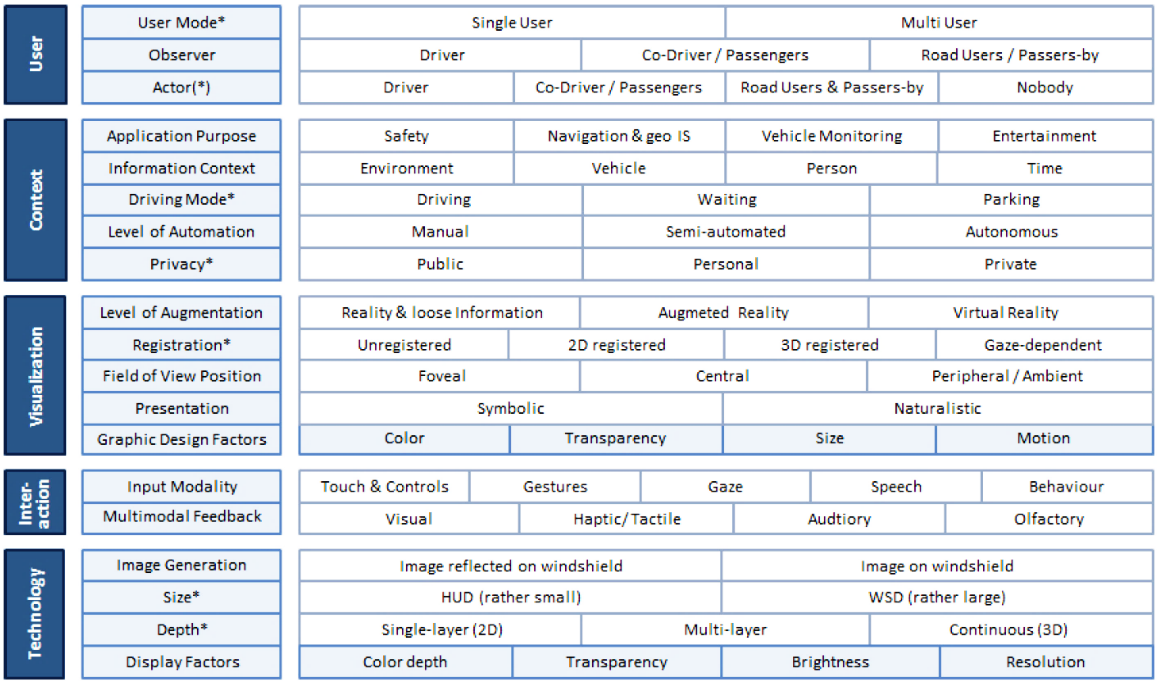
Figure 6. Two ratio scales included in the 2004 Audi S4

Figure 7. Modern Analogue Speedometer from the 2006 BMW M6

The last consideration for the design was the typeface. For instruments, simple san-serif typefaces are recommended by Woodson et al (1992) and were followed for the current designs in external speedometers.

3.7. Design Space: Functional Requirements

As the vehicles transitions beyond just a mode of transport and into a space for entertainment and leisure, the design space the eHMI displays needed a framework to define areas to explore. The design space was investigated primarily to understand today's display technology, current physical dimensions, and existing interaction schemes to contextualize the designing of the current eHMI visuals.



User	User Mode*	Single User	Multi User			
	Observer	Driver	Co-Driver / Passengers	Road Users / Passers-by		
	Actor(*)	Driver	Co-Driver / Passengers	Road Users & Passers-by	Nobody	
Context	Application Purpose	Safety	Navigation & geo IS	Vehicle Monitoring	Entertainment	
	Information Context	Environment	Vehicle	Person	Time	
	Driving Mode*	Driving	Waiting	Parking		
	Level of Automation	Manual	Semi-automated	Autonomous		
	Privacy*	Public	Personal	Private		
Visualization	Level of Augmentation	Reality & loose Information	Augmeted Reality	Virtual Reality		
	Registration*	Unregistered	2D registered	3D registered	Gaze-dependent	
	Field of View Position	Foveal	Central	Peripheral / Ambient		
	Presentation	Symbolic	Naturalistic			
	Graphic Design Factors	Color	Transparency	Size	Motion	
Inter-action	Input Modality	Touch & Controls	Gestures	Gaze	Speech	Behaviour
	Multimodal Feedback	Visual	Haptic/ Tactile	Audtiory	Olfactory	
Technology	Image Generation	Image reflected on windshield	Image on windshield			
	Size*	HUD (rather small)	WSD (rather large)			
	Depth*	Single-layer (2D)	Multi-layer	Continuous (3D)		
	Display Factors	Color depth	Transparency	Brightness	Resolution	

Figure 8. Windshield design space proposed by Haeuslschmid et al., 2016

Heauslschmid et al (2016), seen in Figure 8. Windshield design space proposed by Haeuslschmid et al., 2016, proposed four dimensions to categorize design spaces: content, visualization, interaction, and technology. The design space framework was intended for both interior and exterior designs. As a result, not all the dimensions correlate to the current research direction of external displays. This framework was useful as a guide for the creation of functional requirements for the proposed designs. In terms of context, the designs are for the application of pedestrian safety and involve both environment and vehicle information, such as speed and gap estimation. The driving mode includes driving and waiting and the level of automation is for semi-automated (level 3) to fully autonomous (level 5). The privacy of the information is set to be publically accessible as pedestrians is a part of the general public.

In terms of visualization, colour, transparency of display, and size were considered. The display is 2D and no augmented reality or virtual reality technology was considered in this design cycle. Colour is an important consideration as discussed previously since the current Ontario Highway Act does not allow the universally accepted colour schemes of red and green to be used in frontal lamps. Both colour and non-colour were considered in order to understand the level of impact colour could potentially have. The transparency of displays was set to be 100% visible, or as close to an opaque display quality as possible, to clearly indicate the symbolic visuals of text, icons, or anthropomorphic displays. The size of the displays is set to cover the entire windshield panel for enhanced visibility to the pedestrian. Unlike the display information in head-up displays (HUD), the information displayed is not for the driver or passenger and therefore not limited to the size of most HUDs. For comparison purposes, BMW HUDs had HUDs the size of 7.5 x 17.5 cm in 2012 (Heauslschmid et al, 2016).

Multimodal feedback, mainly auditory feedback, were not yet considered for efficacy in an experimental setup. In the future, auditory feedback will be considered as secondary extensions to the primary visual output on the windshield for greater accessibility for the visually impaired.

In terms of available display technology, there are mainly two ways to generate an image on a windshield display. One way is that the image is generated directly on the windshield itself. A laser projector or an embedded transparent display panel, such as OLED (organic light-emitting diode, the same ones used in mobile phone screen displays), is integrated into the windshield. The second, more prevalent method, is to use a series of reflection mirrors to project the display on to the windshield, called combiners. This is the technology used in most head-up displays (HUD), seen in Figure 9. HUD display diagram from Continental Automotive GmbH.

The technology exists to implement the designs investigated in this research; however, the cost and implementation difficulty remains high. Hence, the designs remain as early exploratory concepts with variables such as technology stack required, cost to implement, and manufacturing only quickly examined and not thoroughly investigated.

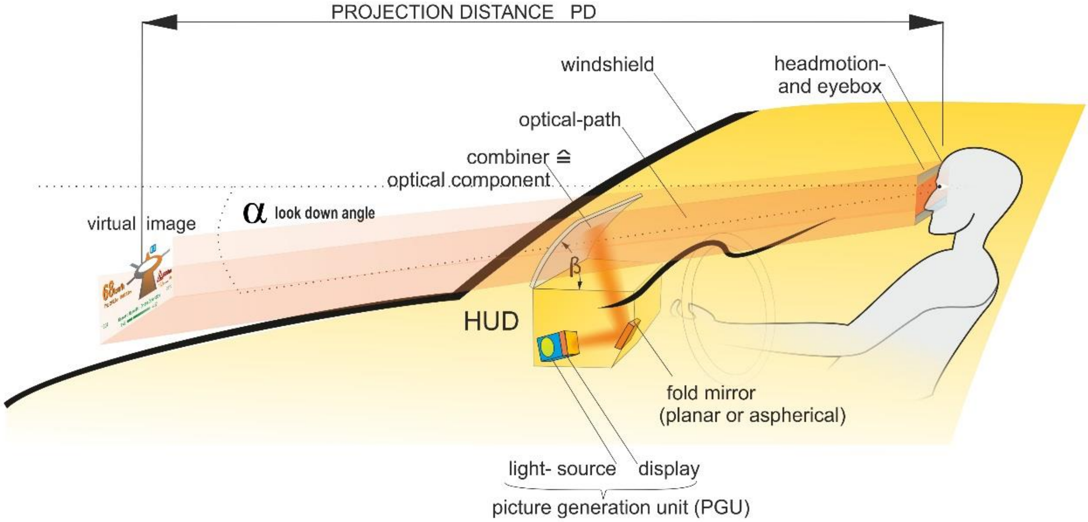
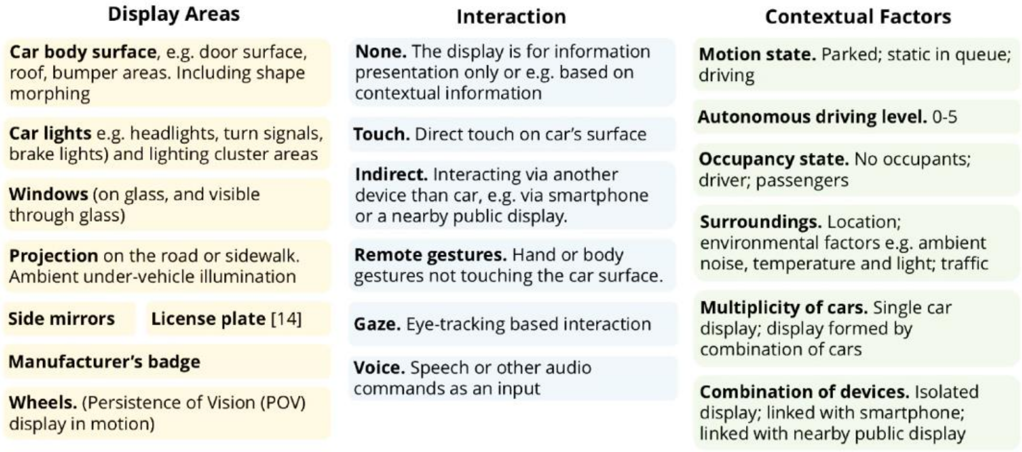


Figure 9. HUD display diagram from Continental Automotive GmbH

The focus of the design space is the windshield. When pedestrians attempt to establish communication with AVs, the windshield remains the focal point of visual contact as the driver in the past conventional vehicle is direct seen behind it. This is a critical communication channel that replaces the human-human communication that existed in manual vehicle communication. However, in order to maximize sensor integration and create more physical space for connected devices to be installed, certain designs also exploited underutilized space. Colley et al (2017), seen in figure 6., appends Haeuslschmid et al's framework by focusing the audience more towards other road agents, such as pedestrians. Using Colley et al's design space for external car displays, design concepts such as road projection were investigated.



Display Areas

Car body surface, e.g. door surface, roof, bumper areas. Including shape morphing

Car lights e.g. headlights, turn signals, brake lights) and lighting cluster areas

Windows (on glass, and visible through glass)

Projection on the road or sidewalk. Ambient under-vehicle illumination

Side mirrors

License plate [14]

Manufacturer's badge

Wheels. (Persistence of Vision (POV) display in motion)

Interaction

None. The display is for information presentation only or e.g. based on contextual information

Touch. Direct touch on car's surface

Indirect. Interacting via another device than car, e.g. via smartphone or a nearby public display.

Remote gestures. Hand or body gestures not touching the car surface.

Gaze. Eye-tracking based interaction

Voice. Speech or other audio commands as an input

Contextual Factors

Motion state. Parked; static in queue; driving

Autonomous driving level. 0-5

Occupancy state. No occupants; driver; passengers

Surroundings. Location; environmental factors e.g. ambient noise, temperature and light; traffic

Multiplicity of cars. Single car display; display formed by combination of cars

Combination of devices. Isolated display; linked with smartphone; linked with nearby public display

Figure 10. Physical design space for eHMI displays (Colley et al., 2017)

3.8. Design Types, Design Features, and Design Cases

The full breakdown, including the images of the designs, of designs can be found in Appendix A2. The designs were also uploaded to YouTube and the link are in the Appendix A3. There were three novel design features that were included in this study that hasn't been discussed and evaluated previously. External speedometer, speed change indicator, and gap estimation count down timer. Designs that performed strongly in other studies (Li et al., 2019; Fridman et al., 2017;) from both academia and industry were included to be presented in the same fidelity and contest for comparison of performance. These features were also mixed and matched with the most effective designs from the literature this research is inspired by, from Li et al.   and Fridman et al.

4. Method

4.1. Recruitment: Amazon Mechanical Turk

The entirety of the survey, from participant recruitment to data collection, was conducted online using a combination of Amazon Mechanical Turk and Survey Gizmo. Amazon Mechanical Turk (MTurk) is an online crowdsourcing marketplace for virtual tasks to be carried out by a distributed workforce and the workers on MTurk are commonly referred to as "Turkers". Starting in the early 2010s, this online marketplace has been rising in popularity as a recruitment tool for behavioural, sociological, and psychological research studies. (Paolacci & Chandler, 2014) In contrast to small sample size common to traditional campus recruitments, MTurk provides access to a participant population of hundreds of thousands of Turkers globally, and anywhere between 2500 – 7500 participants at any given moment. (Difallah et al., 2018) Although recently, a negative shadow has been cast to the academic community for economic inequality - as the median hourly wage for Turkers is only ~ $2.00 USD. (Hara et al., 2018). Driven by the rapid speed of data collection, low recruitment cost, and comprehensive sample representation, the use of MTurk as the recruitment platform was an effective data collection approach for early design assessments.

The online survey itself was created on SurveyGizmo for reliability and ease of editing and revisions. This made it possible for the survey to become scalable with potentially thousands of participants.

A common issue MTurk experiment runs into is the quality of the data collected; the lack of real-time participant monitoring may allow the subset of the sample population that is primarily motivated by monetary incentives to auto-click through the survey questions. In fact, since the inception of MTurk in 2006, there exist many bot farms, from locations like India, that use an automated script to exploit the MTurk system for financial gains (Dreyfuss, 2018). This suboptimal effort will inevitably deteriorate the quality of the collected data. As a result, three preventative measure was taken.

Firstly, Turkers were filtered through two criteria before accessing the survey. Turkers must have completed more than 500 Human Intelligence Tasks (HIT) prior to the current survey and also have had a HIT approval rate of 95% and above. Secondly, a catch-stimuli question, "Please select B as your answer choice.", was used to make sure that the participant is paying full attention into the survey during the middle of the survey. Thirdly, a question using Google's reCAPTCHA system was added into the survey as well to prevent automated bots from answering the survey.

4.2. Experimental Design

The experiment was a within-subject design with 29 levels. There were 28 different stimuli and one baseline scenario with no designs included. The stimuli were structured into 14 (Design cases: see appendix x) x 2 (Coloured vs. Non-coloured Designs) factorial design. Refer to the Design section for a comprehensive explanation of the design requirements and rationales included to arrive at the design cases.

Inspired by Fridman's approach (2016) to early design stage assessment and its effective use of Amazon Mechanical Turk, the experiment design was aimed at creating a quickly executable experiment that can be rapidly repeated for future design considerations. Akin to vehicle crash tests, real-world simulations that involve actual human pedestrians and vehicles of autonomous vehicle-pedestrian (AVP) display design are a necessity before general road use for the highest level of assurance for critical safety levels. These research experiments are inherently time-exhaustive and require heavy efforts in creating stimuli conditions. The characteristic cost for such studies makes them suitable for testing and evaluation in the final phases of designing displays for autonomous vehicle to pedestrian communication. VR studies still encompass many of the benefits of an actual field study, however, they are still very effort intensive experiment to set up and design for choosing design concepts in early stages.

This study setup aims directly at creating an assessment that is fit for early-stage design evaluation, one that employs a low cost and low time intensive prototyping method for high flexibility and rapid design iteration cycles

The study was a video experiment where the stimulus was 29 display designs, all superimposed on to the same base video of a single sedan in front of an uncontrolled crossing where the driver side cannot be seen. The designs were developed into short animations that composed of static animated images that flashed in a loop.

4.3. Ambiguity in Scenario Design

Road signs, traffic lights, and zebra crossings all explicitly indicate the de facto right of way for various road users. This study aims to understand the efficacy of different designs and attempts to isolate the effect of the designs only. The external factors, such as road signals, impact the pedestrians' decision-making process. For instance, when pedestrians observe an incoming vehicle in front of a zebra crossing, and there is an all-way stop sign for the vehicle's direction of travel, the pedestrian would fully expect the vehicle to stop in front of them and yield to them the right of way. Only if when the vehicle has no indication of the intent to slow down and yield would the pedestrian change their initial mental assumptions and start to observe more of the vehicle's overt kinematics and its distance from the zebra crossing.

The experiment design specifically attempts to maximize ambiguity in vehicle-pedestrian communication by removing as much road signs and indications as possible. (insert stats about jaywalking probabilities and the fatality rate of jaywalking) The participant would be 'jay-walking' and the by-laws would categorize their actions as illegal. However, (stat about court cases and how vehicles still have to yield in jaywalking scenarios since they can't win the legal case), this ambiguity imposes a mandatory negotiation between the vehicle and the pedestrian, as the ethical ramification of life and death is at hand. This specific scenario allows for a relatively isolated evaluation of the efficacy of the communication design and the designs alone.

A beige Toyota Camry was used as the vehicle to represent the commonality of future AVs. This also enhances the familiarity of the experiment setting as it's one of the most common commercial road vehicles available today.

An interesting condition included in this experiment is the weather that was presented in the scenarios to the participants. Most of the previous studies on the AVs' display efficacy to pedestrians were conducted in weather conditions that were optimal and would be considered the baseline for weather conditions - usually with sunny, clear skies with close to 100% visibility and no adverse weather conditions, such as rain, fog, or snow present. The current study was conducted in Toronto, Canada during the winter month of February after a night of freezing rain.

The series of photographs were taken on the afternoon of Saturday, February 9, 2019, where the weather was -3 degrees Celsius after a night of freezing rain. There were frozen snow and ice residues of ~ 0.5 inches in thickness left on the vehicular road. The sidewalks were clear and had very little to no snow or ice. A series of three images were captured; one where the vehicle is approximately 150m away, one where the vehicle is approximately 75m away, and one where the vehicle is directly in front of the pedestrian. (explanation of the distance)

These photos were taken around the block of Regent Street and Cole Street in a residential area of downtown Toronto, Canada. The street was a one-way street with no visible road signs, zebra crossing, or traffic lights in the section photographed. The photos were taken on a Canon EOS 5D Mark III with a dimension of 5760 px by 3840 px. They were downsized and converted into short animation videos of 1080p videos with a dimension of 1920 px x 1080 px that were subsequently uploaded on YouTube for ease of online access. Experiment materials that were used to conduct the survey are included in Appendix B: Survey Questionnaire.

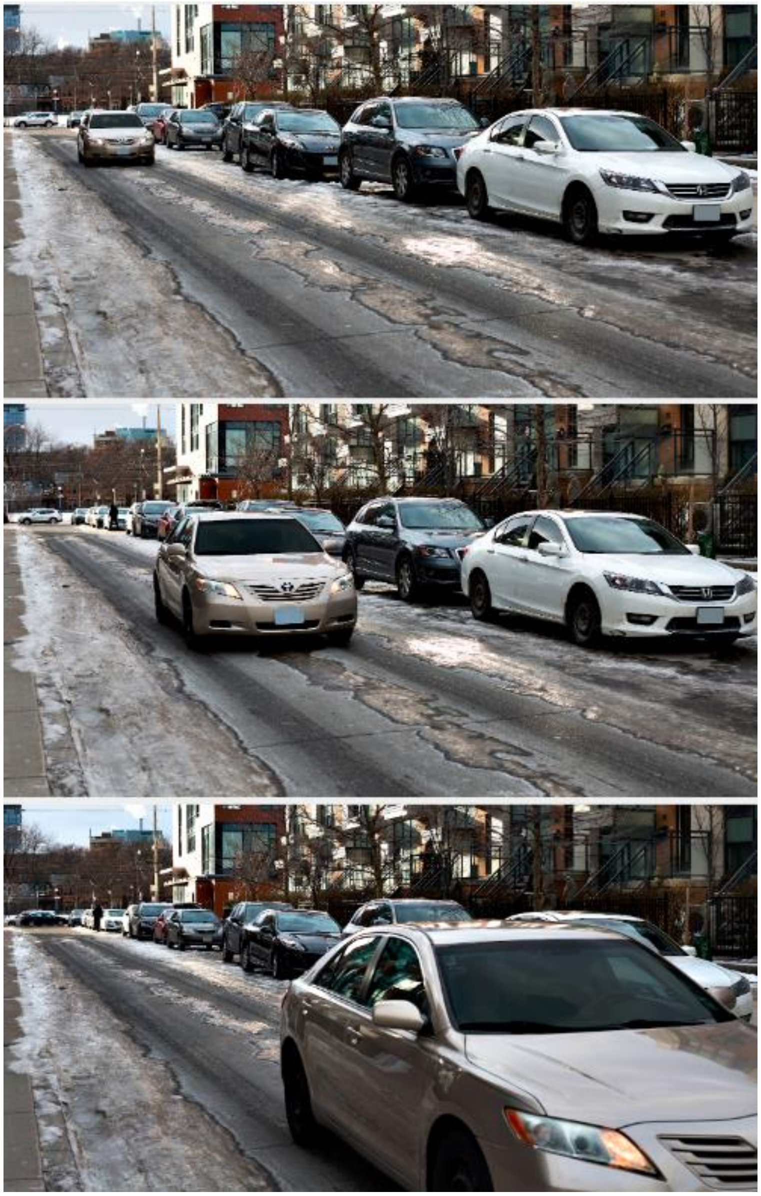


Figure 11. Visual stimuli: baseline condition

4.4. Procedure

The entirety of the survey, from recruitment to payment, was conducted online using a combination of Amazon Mechanical Turk and Survey Gizmo. Participants were recruited using the MTurk platform. Individuals who have an MTurk account (known as "workers") had access to a list of available studies (known as "human intelligence tasks" or "HIT") to which they can voluntarily participate.

When participants clicked on the title of the current study, they were redirected to the study's SurveyGizmo survey page. An information letter that outlines the procedure for completing this study and a consent form for obtaining their consent was presented.

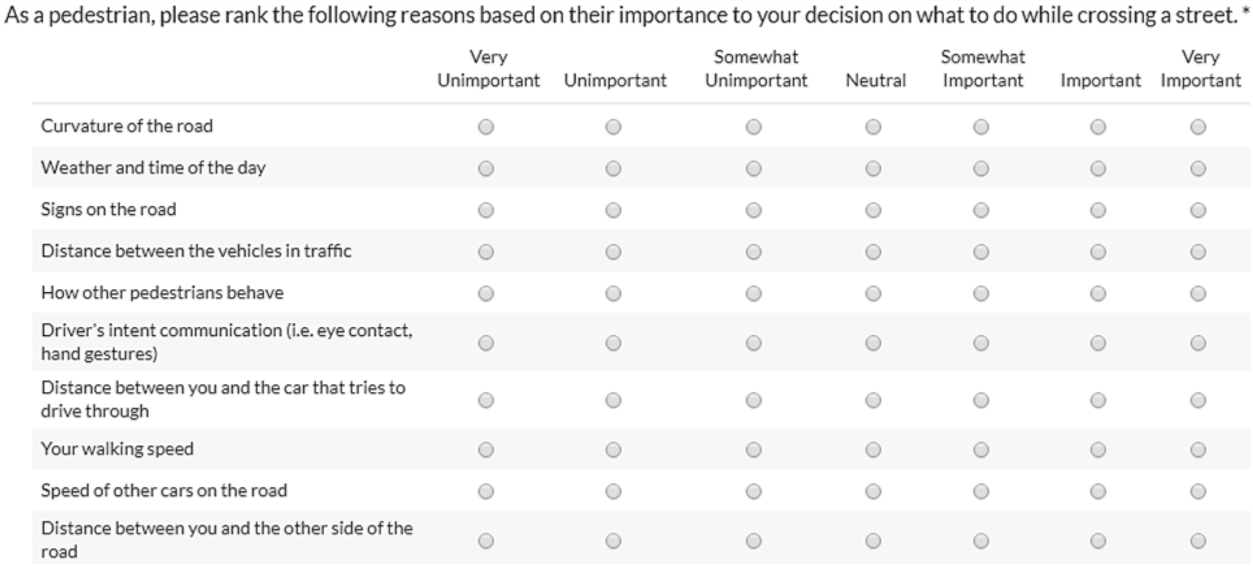
Before the actual study survey, a screening questionnaire was used to ensure all participants were adults above the age of 18 years old, were proficient with reading and writing in English, had normal vision with or without corrective lenses, and recent experience as a pedestrian in the week prior to partaking the survey. Vehicles behaviours and roads signs have a different interpretation depending on the language and culture of different regions. This study attempts to understand the crossing behaviour in an English-speaking country; hence, it requires proficiency with reading and writing in English from the participants. The study also attempted to understand the visual impact of various designs on pedestrian behaviours, therefore the participants must have had recent experience as a pedestrian to understand the context of the design and must also have normal vision (with or without corrective lenses) to be able to visually identify each design clearly. The age range was limited to 18 years old or older for the legality of payment purposes on Amazon Mechanical Turk. The participants' driving experience was not required to participate in this study.

To better understand the background of the participants, as well to enrich the data analysis for design preference and crossing behavior patterns, gender, familiarity with metric system or empiric system, years of driving experience, distance driven over the past month, and ownership of vehicles with Advanced Driving Assistance Systems (ADAS, i.e. lane departure warning system, automatic lane centring) were collected as part of the demographics survey.

Before the assessment of the design stimuli, a pre-experiment questionnaire was conducted to understand the pedestrians' predisposed crossing behaviour prior to viewing the experiment stimuli as well as to understand their trust and perception of autonomous vehicles. The first part of the questionnaire includes questions that self-reports the current behaviour patterns the participant exhibits as a pedestrian crossing street in real-life scenarios. The participants were presented with a hypothetical scenario that involves a pedestrian crossing in front of a moving car. The scenario explicitly stated that there were no stop signs, no traffic lights, no yield to pedestrian signs, and no painted yield bars. A car approaches from the left-hand side and the participant was a pedestrian attempting to cross the road (jaywalking). It was a scenario specifically set up to introduce ambiguity as to the right of way, hence negotiation and communication between the vehicle and the pedestrian must occur.

One of the key questions to understand the pedestrians' natural crossing preferences asked the participant to rate on a Likert scale of 7 about the criteria included in Table 6. Questionnaire for mental model factors that influence pedestrian crossing decision

Table 6. Questionnaire for mental model factors that influence pedestrian crossing decision



As a pedestrian, please rank the following reasons based on their importance to your decision on what to do while crossing a street. *

	Very Unimportant	Unimportant	Somewhat Unimportant	Neutral	Somewhat Important	Important	Very Important
Curvature of the road	○	○	○	○	○	○	○
Weather and time of the day	○	○	○	○	○	○	○
Signs on the road	○	○	○	○	○	○	○
Distance between the vehicles in traffic	○	○	○	○	○	○	○
How other pedestrians behave	○	○	○	○	○	○	○
Driver's intent communication (i.e. eye contact, hand gestures)	○	○	○	○	○	○	○
Distance between you and the car that tries to drive through	○	○	○	○	○	○	○
Your walking speed	○	○	○	○	○	○	○
Speed of other cars on the road	○	○	○	○	○	○	○
Distance between you and the other side of the road	○	○	○	○	○	○	○

Subsequently, the participant was asked a series of questions regarding their own opinions about autonomous vehicles in general. This includes the trust towards autonomous vehicles, preference in similarity in vehicle design, and information to gauge the public opinion towards autonomous vehicles. The full questionnaire is included in Appendix B: Survey Questionnaire.

For the stimuli assessment, the participants were asked to evaluate a series of designs that aims to facilitate the intent and mode communication of autonomous vehicles to pedestrians. The following scenario description was presented to the participant: "It's a Sunday afternoon in the wintertime. Snow and ice have been cleared from the road. Suppose you are looking at the images from the perspective of a pedestrian that stood on the curb and decided to cross the street. An autonomous vehicle is coming towards you. As the pedestrian, you find yourself in the middle of the roadblock and you intend to cross the road to the other side. There are no stop signs, no traffic light, no yield to pedestrian sign, and no painted yield bars. The autonomous vehicle detects your intent to cross the road and attempts to communicate with you its intent. There will be different external display designs shown on the vehicle. Each communication design will be different. After showing you the design under the scenario context, we would like to ask you to ask your opinion about the communication design."

29 YouTube videos approximately 5 seconds in duration were shown. The videos, described in the experiment design section, consisted of three main images: one where the vehicle is approximately 150m away, one where the vehicle is approximately 75m away, and one where the vehicle is directly in front of the pedestrian.

The survey was expected to take 45 - 60 minutes to complete.

4.5. Measures

The current study includes an early design stage formative evaluation. Formative evaluation is the type of evaluation where the designs and assessment of the designs are conducted in parallel. It's an iterative process where designs are improved cyclically. Depending on the fidelity of each design prototype, the design cycle can be iterated quickly and effectively under such evaluation methods as the strength and weakness of each design is contextually identified.

In contrast, summative evaluation is the type of evaluation for the final product at the end of a design cycle. It requires high fidelity of the prototype or the market-ready product to be tested with the intended targeted user group within a real-world context. It could be used in conjunction with formative evaluation. Errors and weaknesses identified in the summative evaluation may be costly to change and could possibly have been identified in early design and development stages.

The experiment attempted to optimize the measures such that the data dimensions were rich enough to determine the effectiveness of the design while maintaining the conciseness of the evaluation of each design scenario. This was important as the experiment attempts to include a large number of stimuli to understand the strength and weakness of each design stimuli; hence, scalability was key to selecting the measurements of the stimuli.

One of the key considerations for the primary task measures was participants' mental fatigue in correlation with the duration of the study. Mental fatigue induced by continuous, repetitive simple tasks could warrant skewed data regarding the designs. As mental fatigue increases, attention, in the form of task performance accuracy and reaction time, decreases; the participants are more likely to base their response to questions on irrelevant information as opposed to the intended primary task. (Faber et al., 2012).

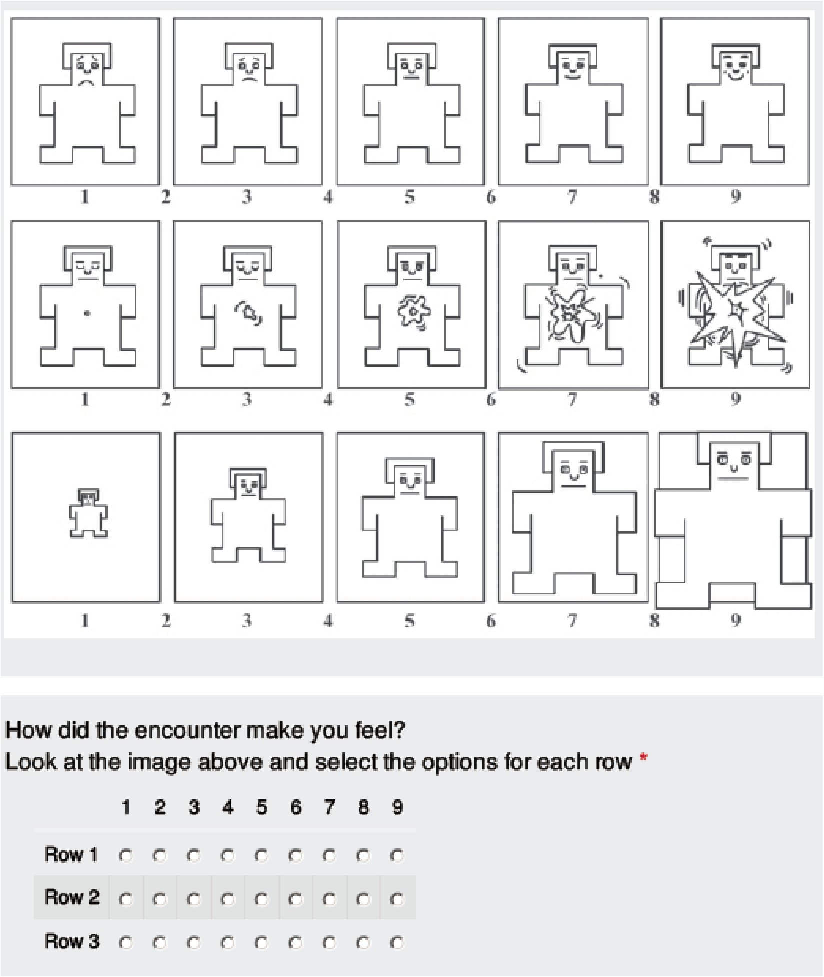
This led to a conscious effort to optimize data dimensions. A number of standardized measures for subjective mental workload (i.e NASA TLX; Hart and Steveland, 1988), usability (System Usability Scale; Brooke, 1996), trust (Scale items in final Human-Computer Trust instrument; Madsen & Gregor, 2000) and other measures were considered but weren't included into the experiment as per this consideration.

The primary measures included in this study derives from theories and empirical studies in the field of warning design and evaluation. Typically, many of the warning studies are subjectively evaluated in several dimensions on a Likert-type scale. These dimensions include the likelihood of compliance, the likelihood of injury, the severity of the injury, hazardousness, perceived urgency, noticeability, comprehensibility, coherence, reaction time, and knowledge recall. (Wogaltor et al., 2002)

For subjective measures, perceived safety, perceived urgency, perceived usefulness, perceived understandability, and emotional response were collected from participants after each scenario.

Perceived urgency was to be rated on a scale of 1 to 100 for maintaining data consistency for result comparison with a relevant study from the same lab (Li et al. 2018).

Perceived safety, perceived understandability, and perceived usefulness were rated on a Likert scale of seven. The emotional state was measured using a Self-Assessment Manikin (SAM) questionnaire. (Lang, 1980; Hodes, Cook, & Lang, 1985) SAM was an emotion assessment tool that uses graphic scales, depicting cartoon characters expressing three emotion elements in rows, and they are, in order: pleasure (level of contentment with the interaction), arousal (level of calmness felt during the interaction), and dominance (level of control felt during the interaction).



How did the encounter make you feel?
Look at the image above and select the options for each row *

	1	2	3	4	5	6	7	8	9
Row 1	○	○	○	○	○	○	○	○	○
Row 2	○	○	○	○	○	○	○	○	○
Row 3	○	○	○	○	○	○	○	○	○

Figure 12. SAM questionnaire included in the experiment survey

For objective measures, observation of when the pedestrian would cross was translated to a survey question form. The participants were asked when they would feel safe about starting crossing the street. Three choices were provided: "before the car passes me - between the first and the second image", "before the car passes me - between the second and the last image", and "after the car passes me".

The subjective and objective measures were repeated for all 29 design cases. A full counterbalance of the stimuli was not possible due to the online logic setup of the experiment. This will be expanded upon in the limitation of the study in section 6. Experiment materials that were used to conduct the survey are included in Appendix B: Survey Questionnaire.

5. Results

5.1. Hypothesis

This section discusses the hypothesis of measurements. Each hypothesis was synthesized from derivations of past literature findings and a generalized understanding of design preference appropriated from other design fields. To reiterate, the first research question that investigated the impact AV eHMIs have on the pedestrians' crossing behaviour and the pedestrian's general perception of AVs. The second research question investigated the degree of intent communication AVs, specifically SAE level 4 high automation enabled vehicles, for optimal engagement with pedestrians - to ensure optimal levels of perceived urgency, safety, clarity, usefulness, and emotional comfort. Hypotheses were categorized into the influence of colour on pedestrian perception of designs, the preference of novel versus current existing designs, the influence of additive features on pedestrian's perception of designs, and the correlation of measurements as an attempt to arrive at a single informative measurement for future research.

5.1.1. Influence of Colour

It is hypothesized that designs equipped with the full spectrum of colours, specifically red and green - currently banned in Ontario as frontal vehicle lamp colours for commercial vehicles, would receive higher perceived levels across all quantitative measures compared to the non-colour equivalent of the same design case. In other words, coloured designs are expected to outperform their non-colour counterpart.

5.1.2. Novel vs. Existing Designs

Four hypotheses about the quantitative measurement for the pedestrian perception of AV eHMI designs of both novel and existing designs were made.

1. For the three novel design concepts included in this study, generally referred to as “explicit reference-based designs”, will outperform existing design paradigms (iconography, explicit text indication, anthropomorphic features, etc.) in quantitative measures of pedestrian’s perception of designs. The novel singular information design types include design type B – speed display, design type C – time-to-arrival countdown, and design type D - speed change indicator.
2. For existing designs, across all measures, icons (design type E) and egocentric text of “walk/stop” (design type H) were hypothesized as the clearest, most useful, and applicable AV-pedestrian eHMI designs known to date.
3. eHMIs with anthropomorphic features may not be as convincing to the pedestrians.
4. Pedestrians will be generally risk-averse when interacting with AV eHMIs.

5.1.3. Influence of Additive Features

On the influence of additive features, it is hypothesized that additive features, mainly a combination of a primary and a secondary information source, will have higher perceived quantitative measurement performance compared to singular information source displays. Design type A to I - 8 design types with singular information sources - will generally have lower perceived measurement ratings compared to design type J to N - 5 design types with dual sources of information.

5.1.4. Correlation of Measurements: Most Informative Measure

A balanced design approach is appropriate for discovering the right level of design fidelity between overly-minimalistic designs and information-overload designs. For speed in discovering this balance, heuristic evaluation is useful. When synthesizing dependent variables in an attempt to reduce data dimensions, it is hypothesized that perceived safety will be the most useful quantitative heuristic evaluative measurement to understand the perception of various AV eHMIs.

5.2. Demographics and Perception of Autonomous Vehicles

In the experiment design, all participants were adults above the age of 18 years old, were proficient with reading and writing in English, had normal vision with or without corrective lenses, and recent experience as a pedestrian in the week prior to partaking the survey.

The study had 100 qualified participants that met the experiment criteria. The average participants' age was 36 years old with an age range between 18 to 64, with 54 male participants and 46 female participants. 90% of the participants were from the United States, with the rest of the participants from Bangladesh, India, Honduras, Puerto Rico, Argentina, and Kenya. 80% of the participants responded that they cross the street at least 5 times a day. Moreover, 81% of the participants had more than 5 years of driving experience, and 78% drove more than 100 km/ month.

Table 7. Age demographics of Participants

Age	% of Participants
18 - 24 years old	5%
25 - 35 years old	55%
35 - 44 years old	21%
45 - 54 years old	11%
55 -64 years old	8%
65 years or older	0%

44% of the participants owe a vehicle with ADAS. On a Likert scale of 7 (1 – very unimportant, 7 – very important), participants rated that the three most important factors, from a list of ten factors derived from literature (see Appendix B), in their decision-making process are: speed of other vehicles on the road (M = 5.93, SD = 1.13), gap estimation (M = 5.85, SD = 1.40), and the width of the road (M = 5.74, SD = 1.08). Between hand gesture, body movement, and eye-contact in intent communication from the driver, hand gesture was perceived as the most important (M =5.64, Likert scale of 7).

Table 8. Participant preference on mental model factors of pedestrian's crossing decision

Factors in the Mental Model of Pedestrian's Crossing Decision	Mean (1 – 7)	Standard Deviation
Speed of other cars on the road	5.93	1.13

Distance between you and the car that tries to drive through (gap estimation)	5.85	1.4
Distance between you and the other side of the road	5.74	1.07
Distance between the vehicles in traffic	5.5	1.41
Your walking speed	5.47	1.22
The curvature of the road	5.32	1.73
Signs on the road	5.29	1.43
Driver's intent communication (i.e. eye contact, hand gestures)	5.25	1.13
Weather and time of day	5.09	1.27
How other pedestrians behave	4.88	1.4

Prior to stimuli exposure, seen in table 9, participants were generally risk-averse in their crossing decision, with 61% of participants crossing only after the vehicle passes.

Table 9. Pedestrian crossing preference prior to stimuli exposure

Pedestrian Crossing Preference Prior to Stimuli Exposure	**% of Participants in Agreement**
Keep the current walking speed, to cross before the car	12%
Walk faster, to cross before the car	27%
Walk slower and make a complete stop, to let the car drives through first	38%
Walk slower but not make a complete stop, to let the car drives through first	6%
Keep the current walking speed, to let the car drives through first	17%

On the public perception of autonomous vehicles, 87% of participants agreed that AVs should communicate with pedestrians in the future. Five options were provided to understand the participants' preference on how AVs should interact with pedestrians before interacting with the experiment stimuli: light indicator, eHMI, voice interface, ground projection, or no communication signals. 59% of the participants preferred lights, 47% preferred eHMIs, 45% preferred voice interfaces, 32% preferred ground projection, and only 5% preferred no communication signals.

Table 10. Participant preference on autonomous vehicle communication mode

Autonomous Vehicle Communication Mode Preference	**% of Participants in Agreement**
Using lights to interact with pedestrians	59%
Using intelligent interfaces via a screen placed onto the external parts of the autonomous car	47%
Using voice interfaces to interact with pedestrians	45%
Projecting information to the ground in front of the autonomous car	32%
No additional communication signals needed	5%

Five options were provided to understand the participants' preference on the type of information AVs should provide to pedestrians before interacting with the experiment stimuli: indication of autonomous mode, vehicle's immediate future action (slow down, speed up), speed, guidance to pedestrians (please stop, go ahead), or no information. 62% preferred information about the vehicle's immediate future action, 51% preferred guidance, 38% preferred explicit information about speed, 33% preferred indication of autonomous mode, and only 6% preferred no information.

Table 11. Participant preference on information to be presented by autonomous vehicle to pedestrians

Information to be Presented by Autonomous Vehicle	**% of Participants in Agreement**

Showing explicit information about whether the car is an autonomous car or a conventional car	33%
Showing explicit information about the autonomous car's next action (slow down, speed up)	62%
Showing explicit information about its current speed	38%
Showing explicit information about the guidance to pedestrians (go ahead, please stop)	51%
Showing no information just like a normal family car	6%

In trust towards future autonomous vehicles, the most important factors for participants that affect their trust of AVs are AVs ability to handle emergency traffic incidents (53%), AVs ability to interact with pedestrians (52%), and the level of intelligence of the system (50%). The least important factor that affects trust is the liability delegation of the vehicle – who's ultimately responsible and in control, the autonomous vehicles or the drivers/passengers in the autonomous vehicles (17%). 61% of participants believe AVs are trustable, 25% believes they are not trustable, and 14% are neutral towards them in trust.

Table 12. Participant preference on factors of trust towards autonomous vehicles

Factors of Trust towards Autonomous Vehicles	**% of Participants in Agreement**
The stability of the autonomous system	45%
The level of intelligence of the autonomous system	50%
The ability of the autonomous system to interact with people in the autonomous car	45%
The ability of the autonomous system to interact with pedestrians	52%
The ability of the autonomous system to handle emergency traffic incidents	53%
Time to respond to instructions	42%

The ability of the autonomous system to diagnose and process errors	39%
The safety of the autonomous system	43%
Who takes the ultimate responsibility in control, the autonomous car or the drivers/passengers in the autonomous car	17%

On the design preference of future autonomous vehicles in terms of similarities with today's vehicles, 57% of participants believe there should be a clear distinction between the design of future autonomous vehicles and the design of conventional vehicles of today.

A one-way repeated measure analysis of variance (ANOVA) was used to analyze the differences between the different designs' impact on pedestrian's perceived urgency, perceived safety, perceived understandability, perceived usefulness, and emotional response through the SAM measures (perceived pleasure, perceived arousal, and perceived dominance). This was an efficient way of understanding the comparative differences between the designs' efficacy across measures.

A two-way repeated measure ANOVA was used to look at both the impact of colour and the design on the pedestrians' measure responses. In the current 14 (design case) x 2 (colour vs non-colour) within-subject factorial design, the independent variables are design case and colour availability. However, two of the 28 test conditions, namely condition 15 – gap estimation count down timer in red-yellow-green, and condition 16 – numerical speed display without colour, wasn't balanced with a counterpart in the second independent variable of colour. As a result, a 13 (design cases) x 2 (colour vs. non-colour) repeated measure was conducted across all measures.

5.3. Perceived Urgency

It's difficult to determine whether high or low perceived urgency can be considered effective in designs unless it is mapped to the contextual/ situational urgency level. The contextual crossing situation presented was a high urgency situation. Even though there was no occurrence of an emergency or unexpected event, a motorized vehicle travelling at the speed limit of 40 kph still induces a high fatality rate in unprotected pedestrians. Hence, a high level of situational urgency was hypothesized.

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in perceived urgency against different design cases. There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. The assumption of sphericity had been violated, $\chi^2(405) = 1539.84$, $p < 0.05$. Epsilon (ε) was 0.648, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way repeated measures ANOVA. A repeated-measures ANOVA showed that there was a statistically significant difference between different designs in the perceived urgency ratings, $F(8.048, 796.723) = 12.951$, $p < 0.0005$, partial $\eta^2 = 0.116$. Therefore, we can reject the null hypothesis that the design case has no impact on perceived urgency.

After examining the main effect of design preferences on perceived urgency, a 13 (design cases) x 2 (colour vs. non-colour) repeated ANOVA measure was conducted to understand the main effect of colour on perceived urgency and if any statistically significant interaction effect exists between different design cases and colour in relations to perceived urgency.

There was a statistically significant main effect of colour on perceived urgency, displays, $F(1,99) = 24.861$, $p < 0.0005$. All coloured design cases, consisting of colour combinations based on a red-green-yellow colour scheme, had higher perceived urgency rating compared to its non-coloured counterpart in design cases, as seen in Figure 13. Mean rating of perceived urgency sorted by eHMI effectiveness (0 – not urgent at all, 100 – extremely urgent). There was also a statistically significant interaction effect between colour and design cases, $F(8.78,868.96) = 3.107$, $p = 0.001$.

Perceived urgency of the baseline case (case 1) with no design presented had a mean (*M*) of 49.86 with a standard deviation (*SD*) of 30.25. The higher the urgency rating, the more urgent the pedestrian felt about the interaction with the AV. The three design cases with the highest levels of perceived urgency of the pedestrian include case 3 – walk/stop icon display ($M = 68.77$, $SD = 21.50$), followed by case 11 – external speedometer with walk/stop text display ($M = 68.29$, $SD = 19.78$) and case 7 – walk/stop text display ($M = 67.95$, $SD = 22.21$). Two designs that had lower perceived urgency rating than baseline include case 20 – virtual driver with anthropomorphic features ($M = 46.53$, $SD = 32.41$) and case 23 – single line road projection ($M = 44.36$, $SD = 31.306$).

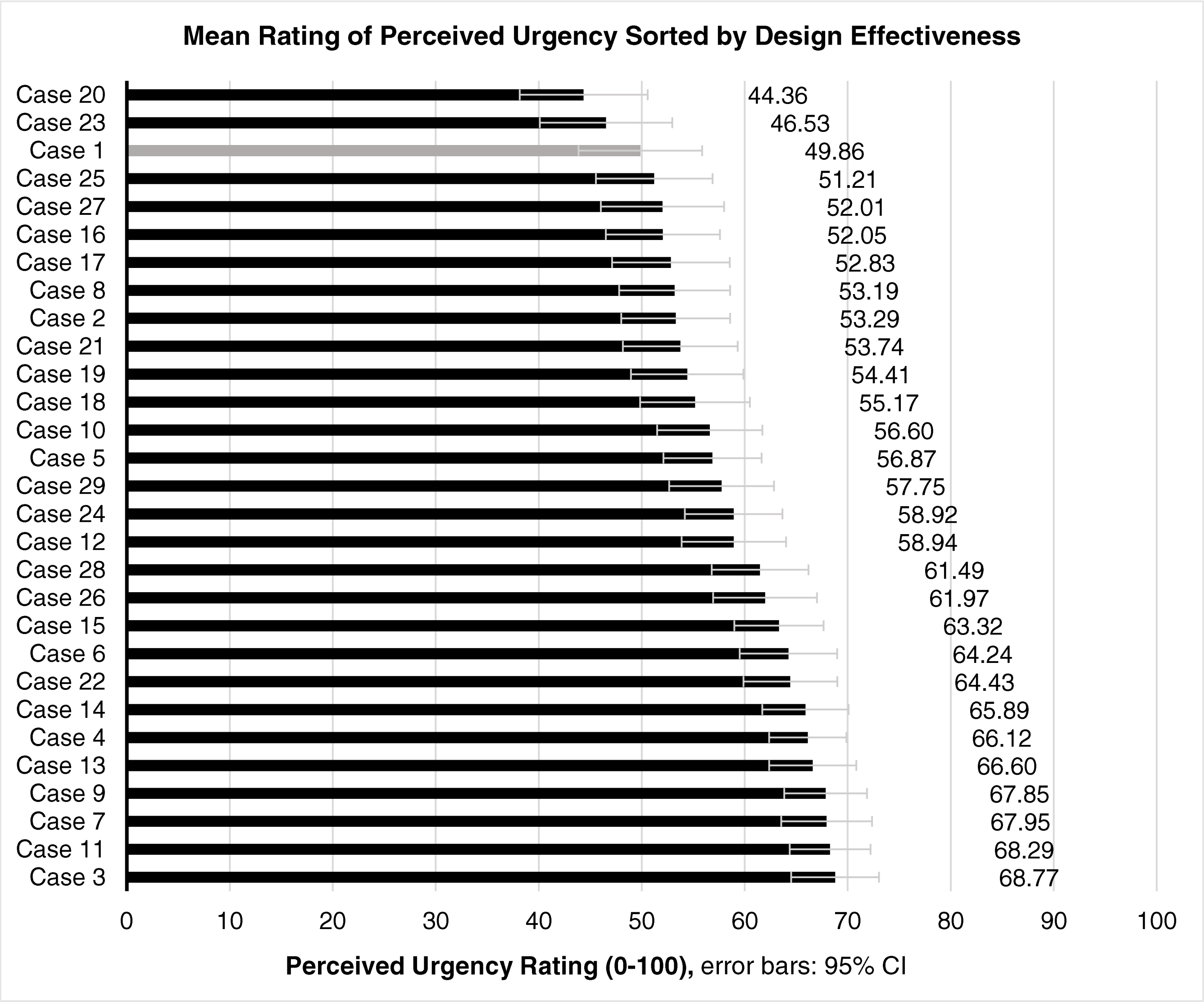


Figure 13. Mean rating of perceived urgency sorted by eHMI effectiveness (0 – not urgent at all, 100 – extremely urgent)

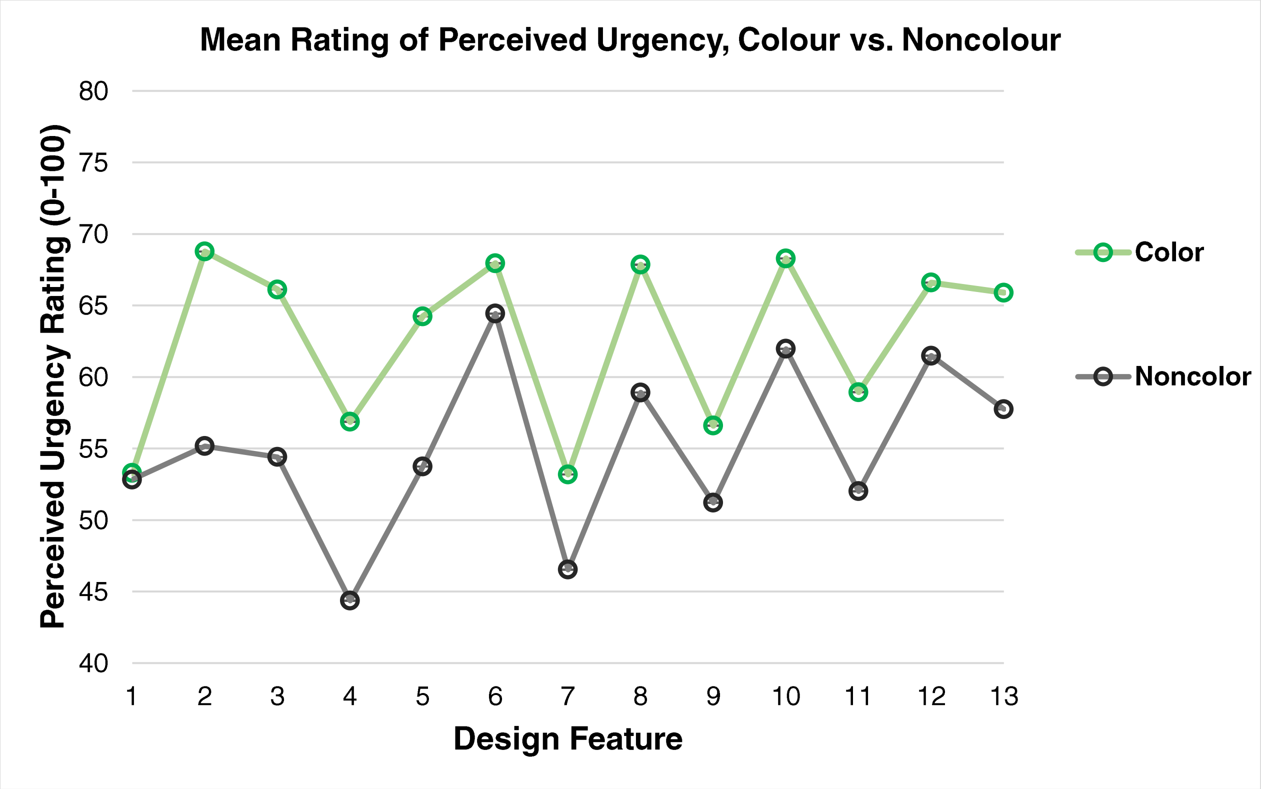


Figure 14. Mean rating of perceived urgency of coloured vs. non-coloured design cases

5.4. Perceived Safety

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in perceived safety against different design cases. There were no outliers. The assumption of sphericity had been violated, $\chi^2(405) = 1073.48$, $p < 0.0005$. Greenhouse & Geisser correction for the one-way repeated ANOVA was applied and $\varepsilon = 0.463$.

The result showed that there was a statistically significant difference between different designs and perceived safety ratings, ($p < 0.05$) $F(11.374, 1126.03) = 15.99$, $p < 0.0005$, partial $\eta^2 = 0.139$. Therefore, the null hypothesis that the design case has no impact on the perceived safety is rejected.

After examining the main effect of design preferences on perceived safety, a 13 (design cases) x 2 (colour vs. non-colour) repeated ANOVA measure was conducted to understand the main effect of colour on perceived safety and if any statistically significant interaction effect exists between different design cases and colour in relations to perceived safety.

There was a statistically significant main effect of colour on perceived safety, displays, $F(1,99) = 32.68$, $p < 0.0005$. All coloured design cases had higher perceived safety rating compared to its non-coloured counterpart in design cases, as seen in figure 15. There was no statistically significant interaction effect between colour and design cases, $F(10.19, 1008.35) = 1.74$, $p = 0.09$.

Perceived safety was rated on a Likert scale of 1 – 7, where the higher the rating, the safer the pedestrian felt about the interaction with the AV. Case 1, the baseline case with no design presented, had $M = 3.72$, $SD = 1.75$. The most effective design cases in pedestrian's perceived safety is case 11 – external speedometer with walk/stop text display ($M = 5.38$, $SD = 1.21$), followed by case 7 – walk/stop text display ($M = 5.27$, $SD = 1.34$) and case 13 – walk/stop icon with gap estimation countdown timer display ($M = 5.2$, $SD = 1.25$). One design performed worse than baseline in perceived safety: case 23 – single line road projection ($M = 3.60$, $SD = 1.78$).

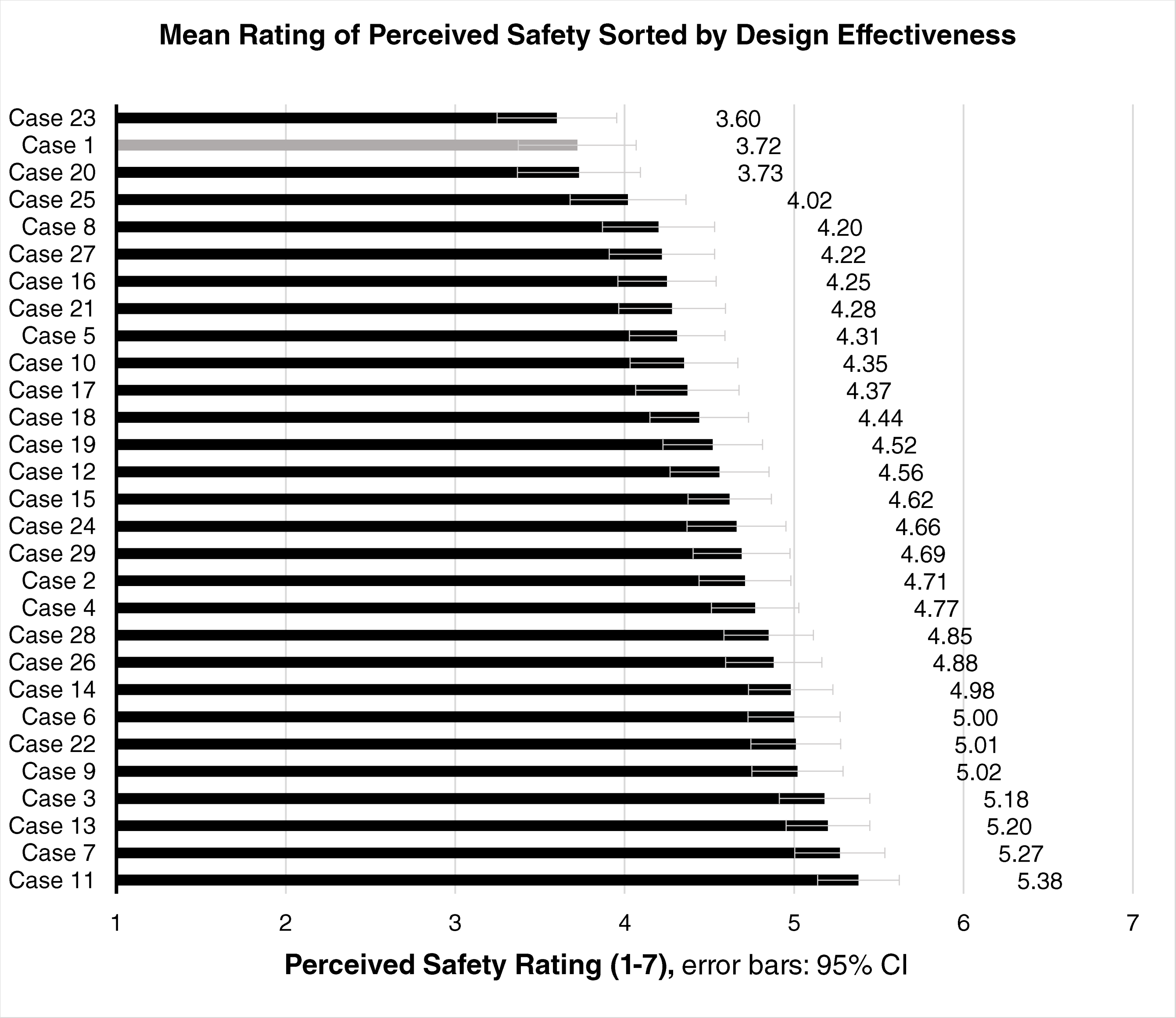


Figure 15. Mean rating of perceived safety sorted by eHMI effectiveness (1 – very unsafe, 7 – very safe)

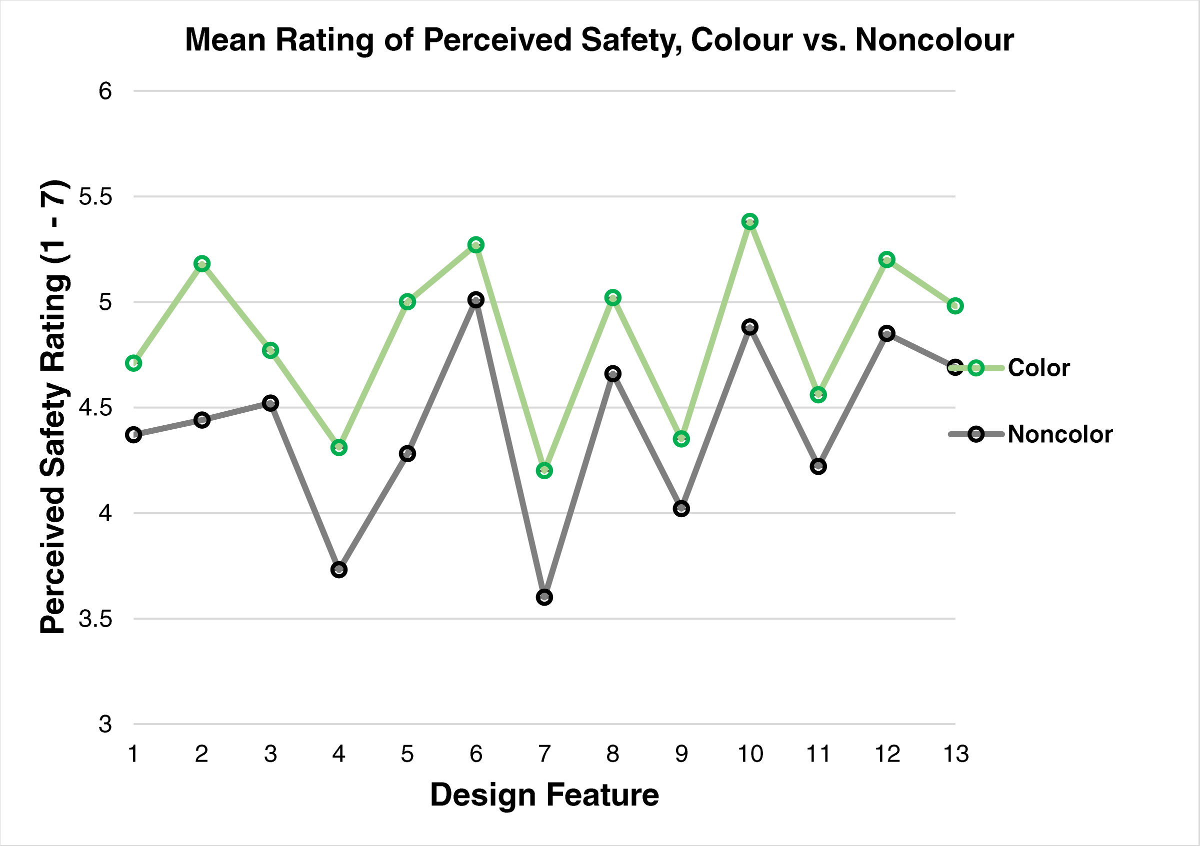


Figure 16. Mean rating of the perceived safety of coloured vs. non-coloured eHMI

5.5. SAM

SAM questionnaire breakdown into three components, pleasure-arousal-dominance. All three dimensions are represented with five images to correlate to the emotional response of the test subject. The five images were transformed into a numerical data set on a Likert scale of 1 – 9 - where the centerline of each image represents an odd number (1, 3, 5, 7, 9) and the gap between the images represents an even number (2, 4, 6, 8).

6.5.1. Perceived Pleasure

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in perceived pleasure level on the SAM questionnaire against different design cases. The assumption of sphericity had been violated, $\chi^2(405) = 1149.94$, $p < 0.0005$. Greenhouse & Geisser correction for the one-way repeated ANOVA was applied and $\varepsilon = 0.407$.

A repeated-measures ANOVA showed that there was statistically significance between different designs and perceived safety ratings, ($p < .05$) $F(11.384, 1127.06) = 8.649$, $p < 0.0005$, partial $\eta^2 = 0.08$. Therefore, the null hypothesis that the design case has no impact on the perceived pleasure is rejected.

After examining the main effect of design preferences on perceived pleasure, a 13 (design cases) x 2 (colour vs. non-colour) repeated ANOVA measure was conducted to understand the main effect of colour on perceived pleasure and if any statistically significant interaction effect exists between different design cases and colour in relations to perceived pleasure.

There was a statistically significant main effect of colour on perceived pleasure, displays, $F(1,99) = 9.84$, $p = 0.002$. All coloured design cases had higher perceived pleasure ratings compared to its non-coloured counterpart in design cases, as seen in Figure 18. Mean rating of perceived pleasure of coloured vs. non-coloured eHMI There was no statistically significant interaction effect between colour and design cases, $F(9.57, 947.14) = 1.57$, $p = 0.114$.

The perceived pleasure was rated on a Likert scale of 1 – 9, where the higher the rating, the happier the pedestrian felt about the interaction with the AV. Case 1, the baseline case with no design presented, had $M = 3.72$, $SD = 1.75$. The most effective design cases in pedestrian's perceived pleasure is case 11 – external speedometer with walk/stop text display (M = 6.44, SD =1.89), followed by case 13 – walk/stop icon with gap estimation countdown timer display (M = 6.41, SD = 1.89) and case 7 – walk/stop text display (M = 6.21, SD = 2.05). One design performed worse than baseline in perceived pleasure: case 23 – single line road projection (M = 4.64, SD = 2.54).

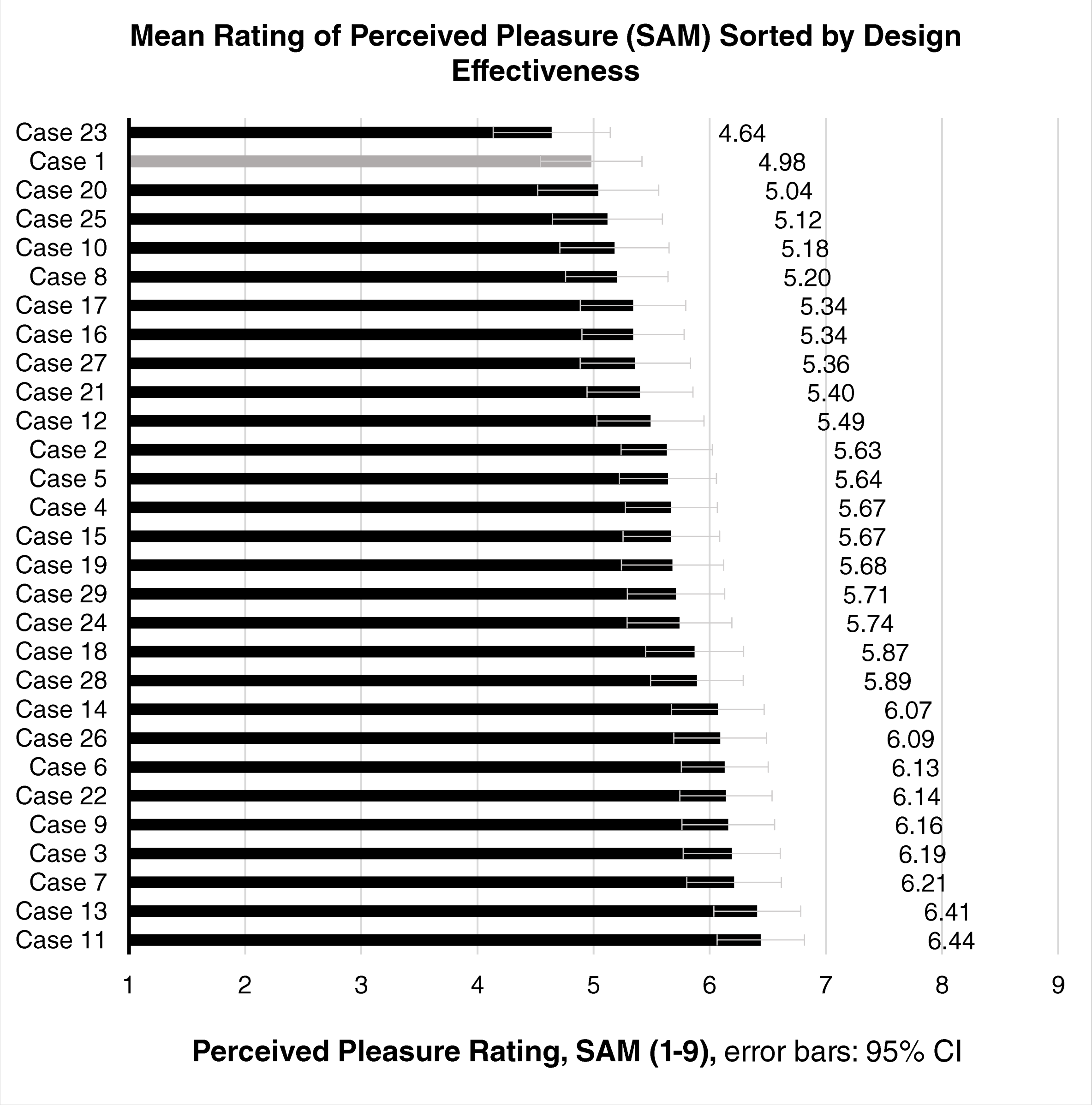


Figure 17. Mean rating of perceived pleasure sorted by eHMI effectiveness (1 – least pleasant, 9 –

most pleasant)

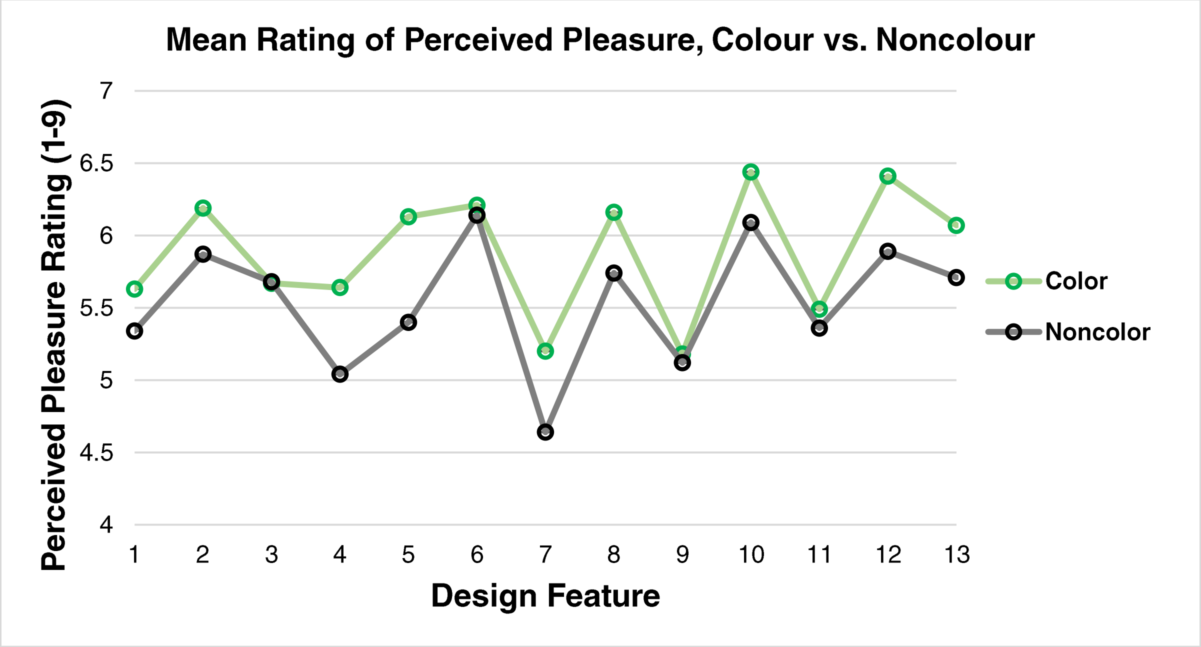


Figure 18. Mean rating of perceived pleasure of coloured vs. non-coloured eHMI

6.5.2. Perceived Arousal

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in perceived arousal level on the SAM questionnaire against different design cases. The assumption of sphericity had been violated, $\chi^2(405) = 939.79$, $p < 0.0005$. Greenhouse & Geisser correction for the one-way repeated ANOVA was applied and $\varepsilon = 0.61$.

A repeated-measures ANOVA showed that there was statistical significance between different designs and perceived arousal ratings, $F(14.45, 1430.54) = 2.563$, $p = 0.001$, partial $\eta^2 = 0.025$. Therefore, the null hypothesis that the design case has no impact on the perceived arousal is rejected.

After examining the main effect of design preferences on perceived arousal, a 13 (design cases) x 2 (colour vs. non-colour) repeated ANOVA measure was conducted to understand the main effect of colour on perceived arousal and if any statistically significant interaction effect exists between different design cases and colour in relations to perceived arousal.

Perceived arousal was rated on a Likert scale of 1 – 9, where the higher the rating, the more annoyed or uncomfortable the pedestrian felt about the interaction with the AV. In other words, a lower score means more effective designs.

There was no statistically significant main effect of colour on perceived arousal, displays, $F(1,99) = 2.40$, $p = 0.12$. Also, there was no statistically significant interaction effect between colour and design cases, $F(9.93, 982.97) = 0.732$, $p = 0.69$. Three non-coloured displays had lower, interpreted as better, mean ratings of perceived arousal (lower levels of annoyance) compared to its coloured counterpart. The designs were gap estimation count down timer (design feature 3, case 4 – colour, case 19 - non-colour), walk/stop text display (design feature 6, case 7 - colour, case 22 - non-colour), and walk/stop text display with gap estimation count down timer (design feature 12, case 13 – colour, case 28 – non-colour). For external speedometer with single line road projection (design feature 11, case 12-colour, case 27-non-colour), the non-coloured displayed had a close but still worse mean in perceived arousal rating with its coloured counterpart. Evidently, colour's hypothesized positive impact wasn't as prominent in this case. The mean and standard deviation of these four designs can be found in Table 13.

Table 13. Non-colour designs that outperformed/ had equal perceived arousal rating than its colour counterpart

Design Feature	Design Case	Mean Perceived Arousal Rating (1 – 9)	Standard Deviation
Feature 3	Case 4 (colour)	5.11	2.242
	Case 19 (non-colour)	5.09	2.288
Feature 6	Case 7 (colour)	4.83	2.495
	Case 22 (non-colour)	4.64	2.364
Feature 11	Case 12 (colour)	5.28	2.234
	Case 27 (non-colour)	5.30	2.232
Feature 12	Case 13 (colour)	4.95	2.401
	Case 28 (non-colour)	4.94	2.390

Case 1, the base line case with no design presented, had *M* = 5.20, *SD* = 2.26. The most effective design cases in pedestrian's perceived arousal is case 22 – walk / stop text display (*M* = 4.64, *SD* =2.36), followed by case 2 – external speedometer display (*M* = 4.68, *SD* = 2.29) and case 11 – external speedometer with walk/stop text display (M = 4.72, SD = 2.39).

Eight designs performed worse than baseline in perceived arousal. They are listed below starting at the top with the least effective case.

Table 14. Eight designs performed worse than baseline in perceived arousal

Ranking	Design Case	Name	Mean	Standard Deviation
1	Case 23	Non-colour External Speedometer Display with Walk/ Stop Icon	5.59	2.27
2	Case 25	Non-colour External Speedometer with Speed Change Indicator	5.40	2.33
3	Case 8	Coloured Single Line Road Projection	5.39	2.49

4	Case 20	Non-colour Virtual Driver Display with Anthropomorphic Features	5.39	2.49
5	Case 27	Non-colour External Speedometer with Single Line Road Projection	5.30	2.23
6	Case 21	Non-colour White-Yellow-White LED Display	5.28	2.50
7	Case 12	Coloured External Speedometer with Single Line Road Project	5.28	2.23
8	Case 10	Coloured External Speedometer with Speed Change Indicator	5.26	2.33

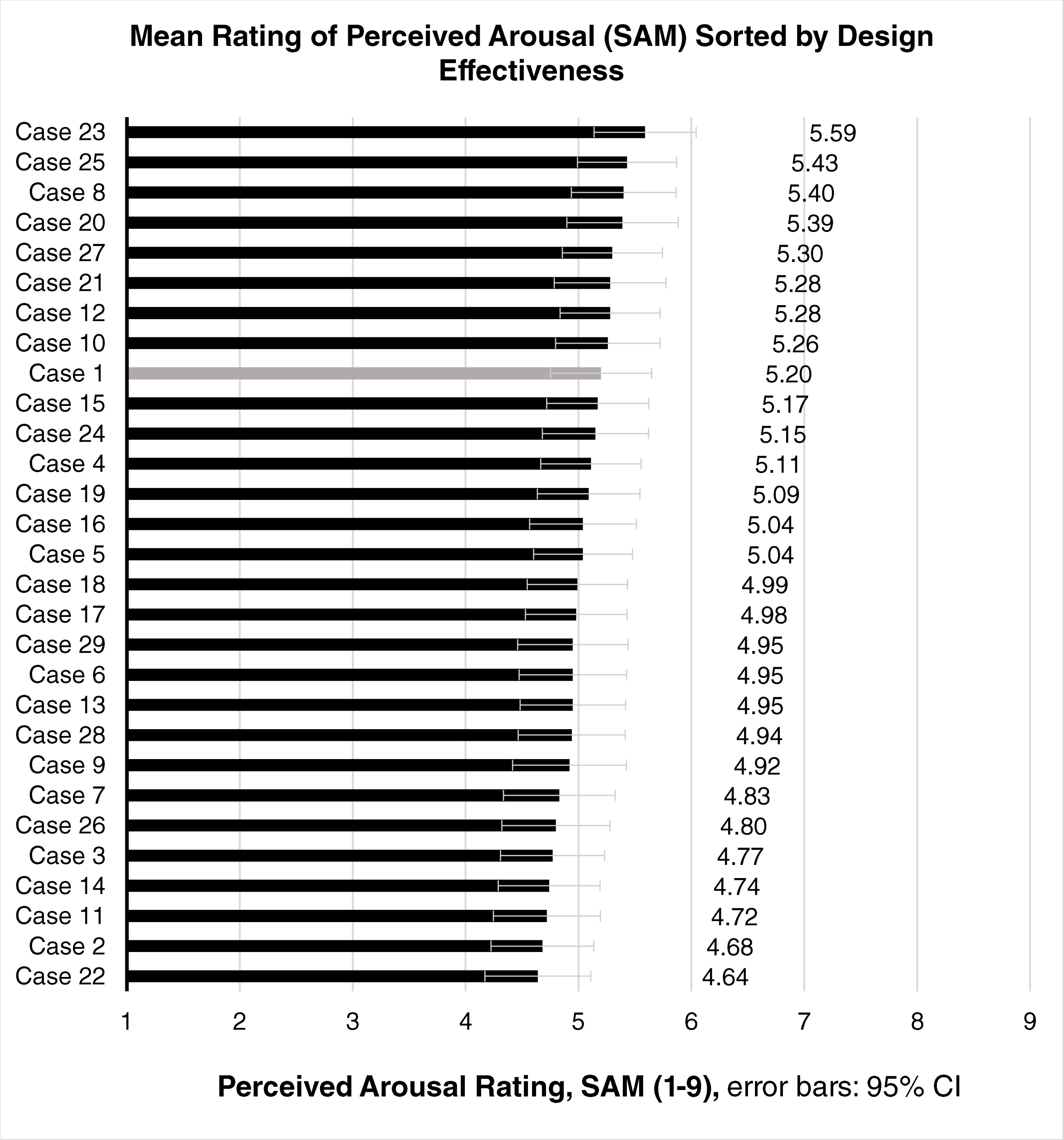


Figure 19. Mean rating of perceived arousal sorted by eHMI effectiveness (1 – least arousing, 9 – most arousing)

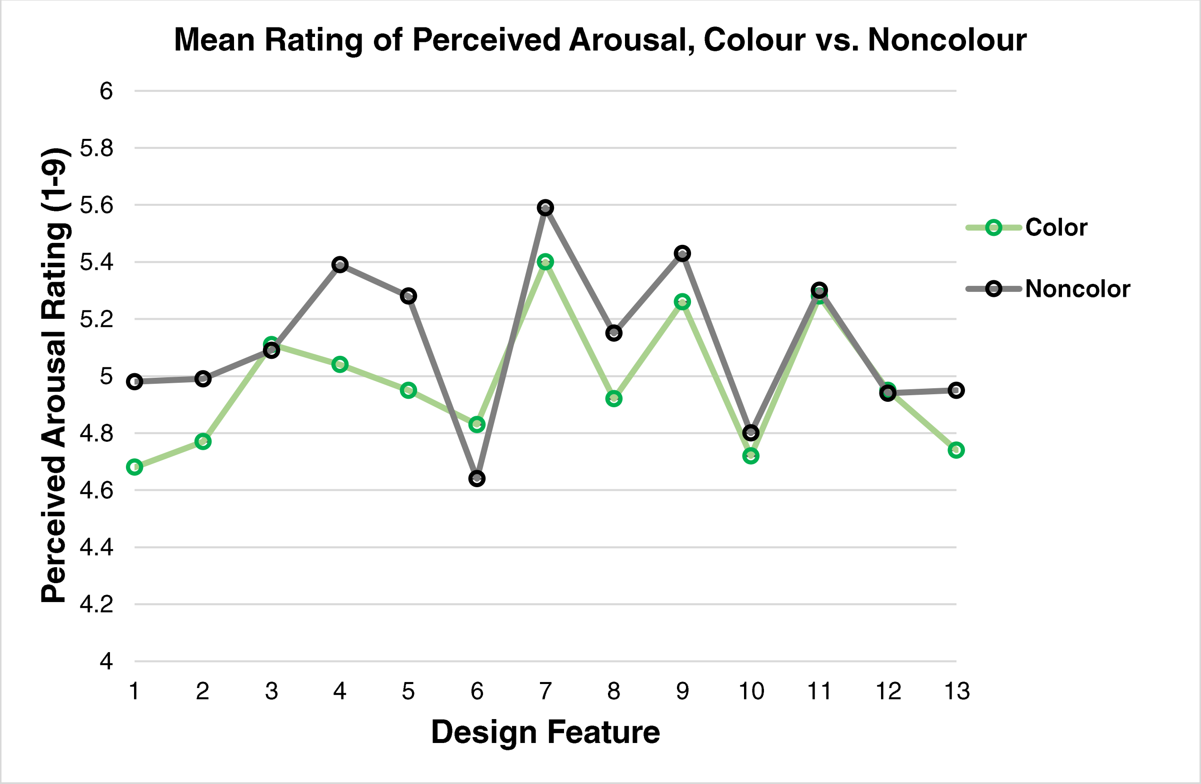


Figure 20. Mean rating of perceived arousal of coloured vs. non-coloured eHMI

6.5.3. Perceived Dominance

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in perceived dominance level on the SAM questionnaire against different design cases. The assumption of sphericity had been violated, $\chi^2(405) = 1060.9$, $p < 0.0005$. Greenhouse & Geisser correction for the one-way repeated ANOVA was applied and $\varepsilon = 0.432$.

A repeated-measures ANOVA showed that there was statistically significance between different designs and perceived dominance ratings, ($p < .05$) $F(12.106, 1196.72) = 2.609$, $p < 0.0005$, partial $\eta^2 = 0.009$. Therefore, the null hypothesis that the design case has no impact on the perceived pleasure is rejected.

After examining the main effect of design preferences on perceived dominance, a 13 (design cases) x 2 (colour vs. non-colour) repeated ANOVA measure was conducted to understand the main effect of colour on perceived dominance and if any statistically significant interaction effect exists between different design cases and colour in relations to perceived dominance.

There was a statistically significant main effect of colour on perceived dominance, displays, $F(1,99) = 4.21$, $p = 0.043$ (Figure 21. Mean rating of perceived dominance sorted by eHMI effectiveness (1 – least dominating/controlling, 9 – most dominating/controlling)). There was no statistically significant interaction effect between colour and design cases, $F(9.71, 961.21) = 1.19$, $p = 0.29$. Two non-coloured displays had higher, interpreted as better, mean ratings of perceived dominance (higher levels of control) compared to its coloured counterpart. The designs are external speedometer with single line road projection (design feature 11, case 12-colour, case 27-non-colour) and walk/stop text display (design feature 6, case 7 - colour, case 22 - non-colour). Two non-coloured designs had similar but still worse mean in perceived dominance rating with its coloured counterpart: walk/stop text display with gap estimation count down timer (design feature 12, case 13 – colour, case 28 – non-colour) and external speedometer with speed change indicator (design feature 9, case 10 – colour, case 25 – non-colour). The mean and standard deviation can be found in Table 15. Non-colour designs that outperformed/ had equal perceived dominance rating than its colour counterpart

Table 15. Non-colour designs that outperformed/ had equal perceived dominance rating than its colour counterpart

Design Feature	Design Case	Mean Perceived Dominance Rating (1 – 9)	Standard Deviation
feature 6	case 7 (colour)	5.54	2.249
	case 22 (non-colour)	5.62	2.286

Feature 9	case 10 (colour)	5.12	2.217
	case 25 (non-colour)	5.11	2.313
feature 11	case 12 (colour)	5.17	2.128
	case 27 (non-colour)	5.21	2.413
feature 12	case 13 (colour)	5.47	2.139
	case 28 (non-colour)	5.46	2.181

Perceived dominance was rated on a Likert scale of 1 – 9, where the higher the rating, the more in control the pedestrian felt about the interaction with the AV. Case 1, the baseline case with no design presented, had *M* = 4.99, *SD* = 2.27. The most effective design cases in pedestrian's perceived dominance is case 6 - Red-Yellow-Green LED display(*M* = 5.78, *SD* =1.97), followed by case 11 – external speedometer with walk/stop text display (*M* = 5.71, *SD* =2.23) and case 7 – walk/stop icon display (*M* = 5.69, *SD* = 2.21). Three designs performed worse than baseline in perceived dominance: case 17 - external speedometer (*M* = 4.99, *SD* = 2.35), case 20 -- non-colour virtual driver display with anthropomorphic features (*M* = 4.98, *SD* = 2.36), and case 23 – single line road projection (*M* = 4.87, *SD* = 2.31).

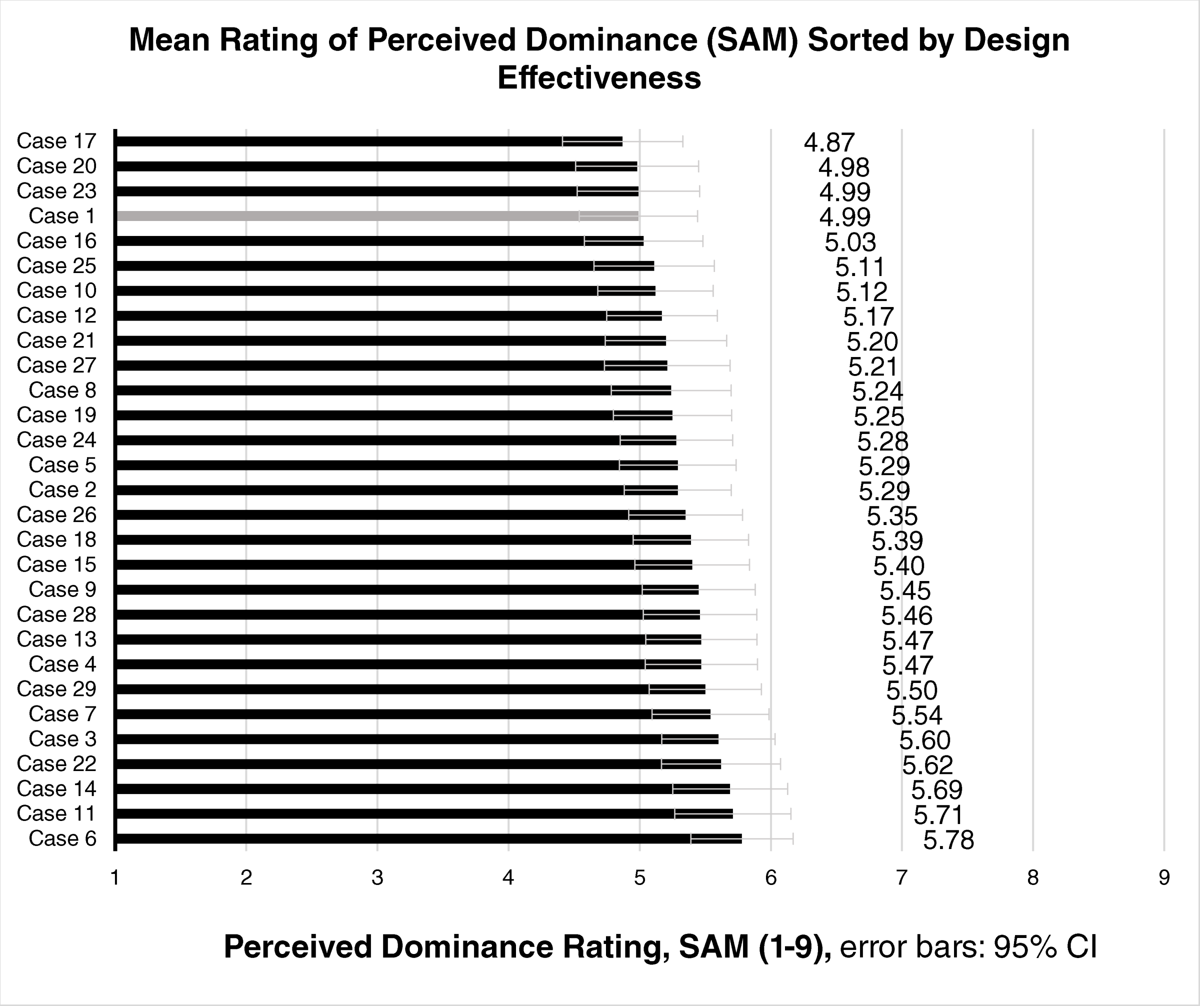


Figure 21. Mean rating of perceived dominance sorted by eHMI effectiveness (1 – least dominating/controlling, 9 – most dominating/controlling)

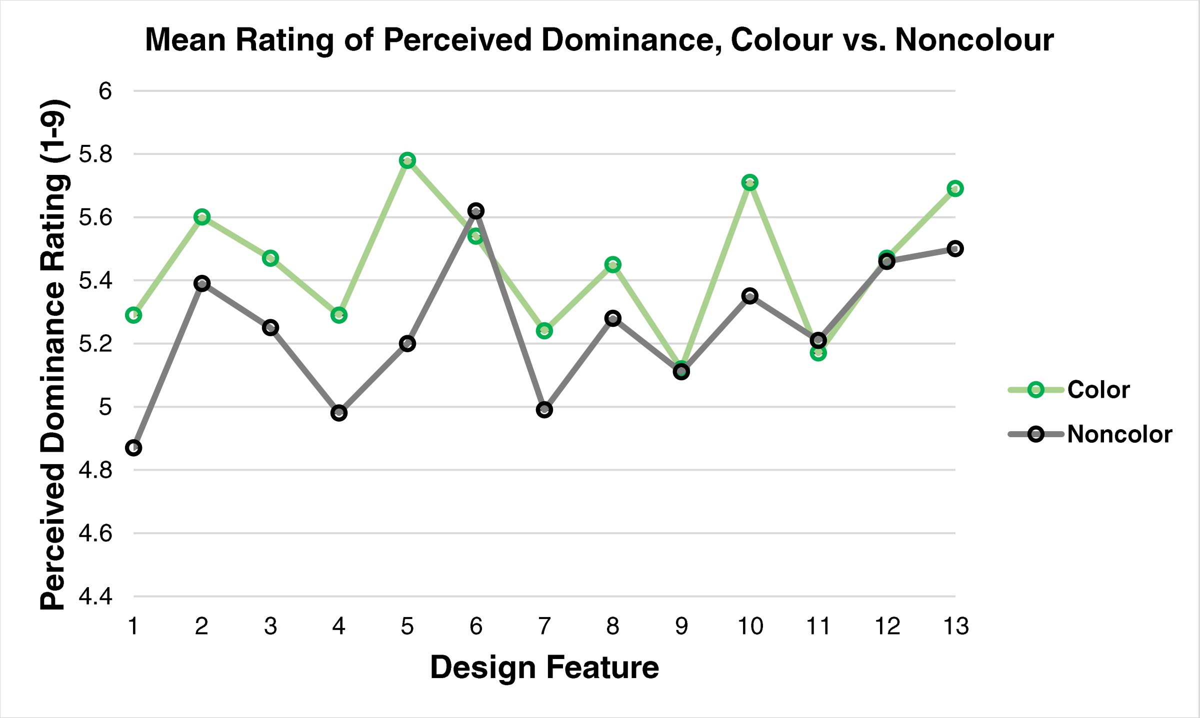


Figure 22. Mean rating of perceived dominance of coloured vs. non-coloured eHMI

5.6. Perceived Usefulness

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in perceived usefulness level against different design cases. The assumption of sphericity had been violated, $\chi^2(405)$ =1060.9, $p < 0.0005$. Greenhouse & Geisser correction for the one-way repeated ANOVA was applied and $\varepsilon = 0.432$.

A repeated-measures ANOVA showed that there was statistically significance between different designs and perceived usefulness ratings, ($p < .05$) $F(12.108, 1196.72) = 2.609$, $p < 0.0005$, partial $\eta^2 = 0.411$. Therefore, the null hypothesis that the design case has no impact on the perceived usefulness is rejected.

After examining the main effect of design preferences on perceived usefulness, a 13 (design cases) x 2 (colour vs. non-colour) repeated ANOVA measure was conducted to understand the main effect of colour on perceived usefulness and if any statistically significant interaction effect exists between different design cases and colour in relations to perceived usefulness.

There was a statistically significant main effect of colour on perceived usefulness, displays, $F(1,99) = 35.39$, $p < 0.0005$. All coloured design cases had higher perceived usefulness rating compared to its non-coloured counterpart in design cases, as seen in Figure 23. There was a statistically significant interaction effect between colour and design cases, $F(9.32, 922.78) =4.68$, $p <0.0005$.

Perceived usefulness was rated on a Likert scale of 1 – 9. Case 1, the baseline case had no designs and perceived usefulness is irrelevant since there is no use case presented to the participant. The most effective design cases in pedestrian's perceived usefulness is case 3 – walk/stop icon display($M = 5.51$, $SD =1.32$), followed by case 11 – external speedometer with walk/stop text display ($M = 5.44$, $SD =1.28$) and case 7 – walk/stop icon display ($M = 5.25$, $SD = 1.42$). Three designs with the lowest rating in perceived usefulness are: case 23 - single line road projection ($M = 3.55$, $SD = 1.80$), case 20 - non-colour virtual driver display with anthropomorphic features ($M = 3.46$, $SD = 1.86$), and case 8 – single line road projection ($M = 4.00$, $SD = 1.76$).

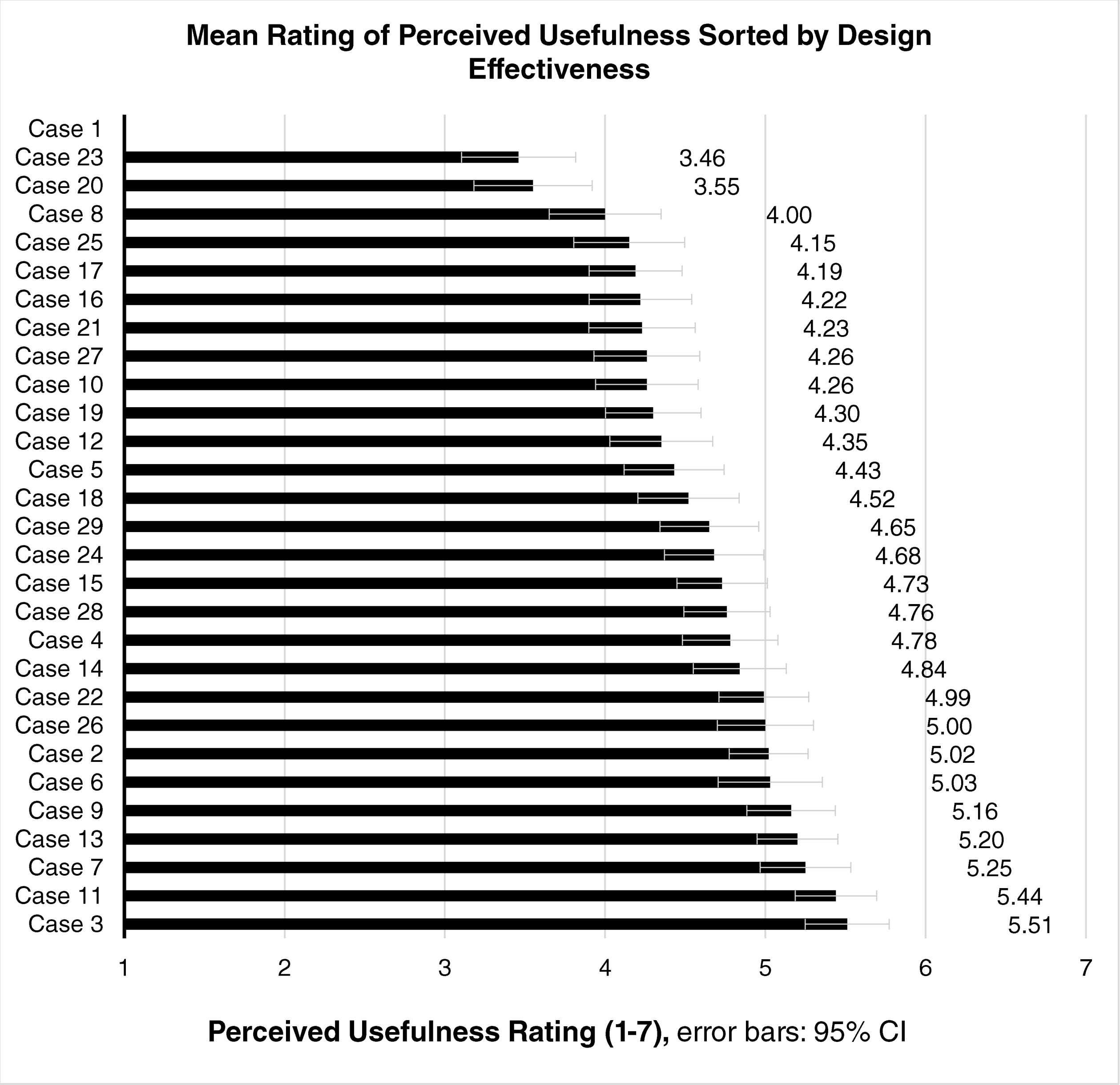


Figure 23. Mean rating of perceived usefulness sorted by eHMI effectiveness (1 – least useful, 7 – most useful)

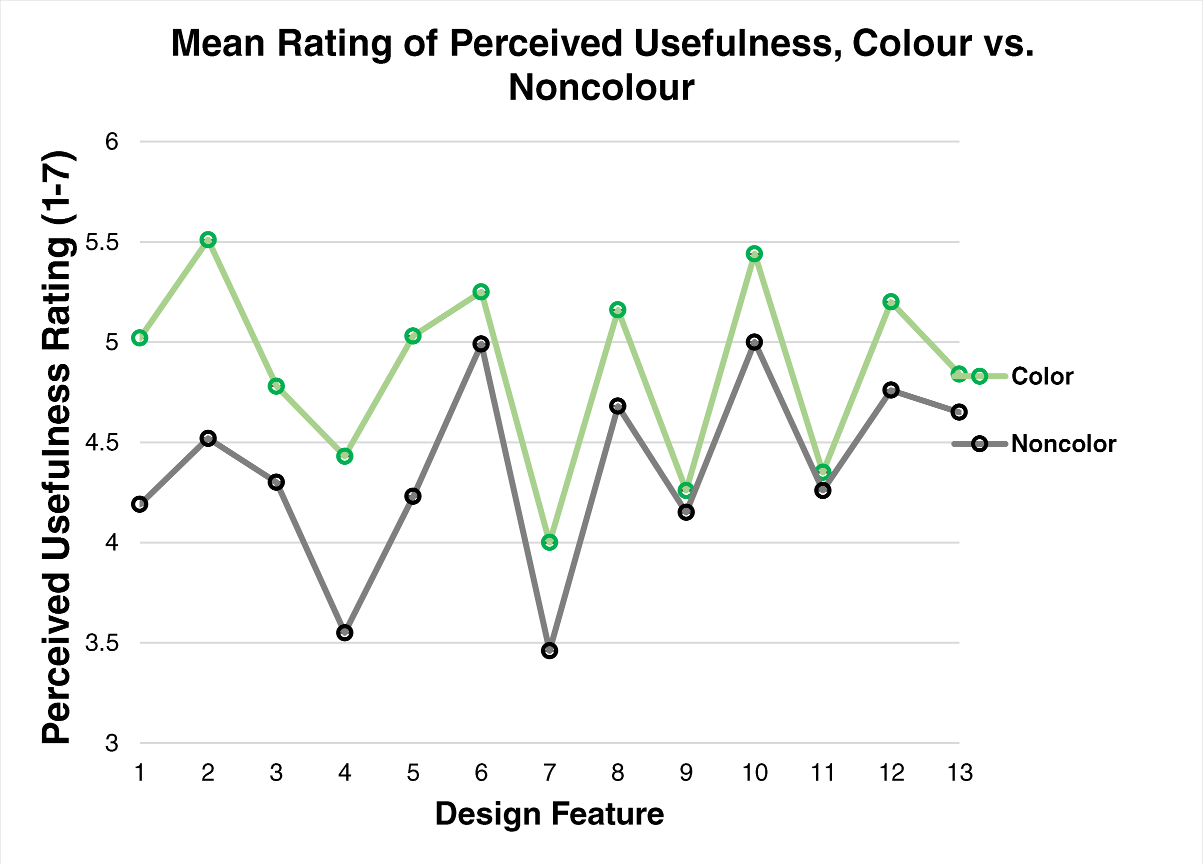


Figure 24. Mean rating of perceived usefulness of coloured vs. non-coloured eHMI

5.7. Perceived Understandability

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in perceived understandability level on the SAM questionnaire against different design cases. The assumption of sphericity had been violated, $\chi^2(405) = 1126.31$, $p < 0.0005$. Greenhouse & Geisser correction for the one-way repeated ANOVA was applied and $\varepsilon = 0.409$.

A repeated-measures ANOVA showed that there was statistically significance between different designs and perceived understandability ratings, ($p < .05$) $F(11.46, 1134.93) = 18.268$, $p < 0.0005$, partial $\eta^2 = 0.156$. Therefore, the null hypothesis that the design case has no impact on the perceive understandability is rejected.

After examining the main effect of design preferences on perceived understandability, a 13 (design cases) x 2 (colour vs. non-colour) repeated ANOVA measure was conducted to understand the main effect of colour on perceived understandability and if any statistically significant interaction effect exists between different design cases and colour in relations to perceived understandability.

There was a statistically significant main effect of colour on perceived understandability, displays, $F(1,99) = 31.84$, $p < 0.0005$. All coloured design cases had higher perceived understandability rating compared to its non-coloured counterpart in design cases, as seen in Figure 25. Mean rating of perceived understandability sorted by eHMI effectiveness (1 – least understandable, 7 – most understandable)There was no statistically significant interaction effect between colour and design cases, $F(9.29, 919.96) = 2.84$, $p = 0.07$.

Perceived understandability was rated on a Likert scale of 1 – 7, where the higher the rating, the easier it is for the pedestrian to comprehend AV's intent communication through the external eHMI. Case 1, the baseline case with no design presented, had $M = 3.70$, $SD = 1.94$. The most effective design cases in pedestrian's perceived dominance is case 7 – coloured walk/stop text display ($M = 5.62$, $SD = 1.38$), followed by case 3 – coloured walk/stop icon display ($M = 5.56$, $SD = 1.33$) and case 11 – coloured external speedometer with walk/stop text display ($M = 5.44$, $SD = 1.28$). Two designs performed worse than baseline in perceived understandability: case 23 – non-colour single line road projection ($M = 3.45$, $SD = 1.86$) and case 8 - coloured single line road projection ($M = 3.70$, $SD = 1.76$).

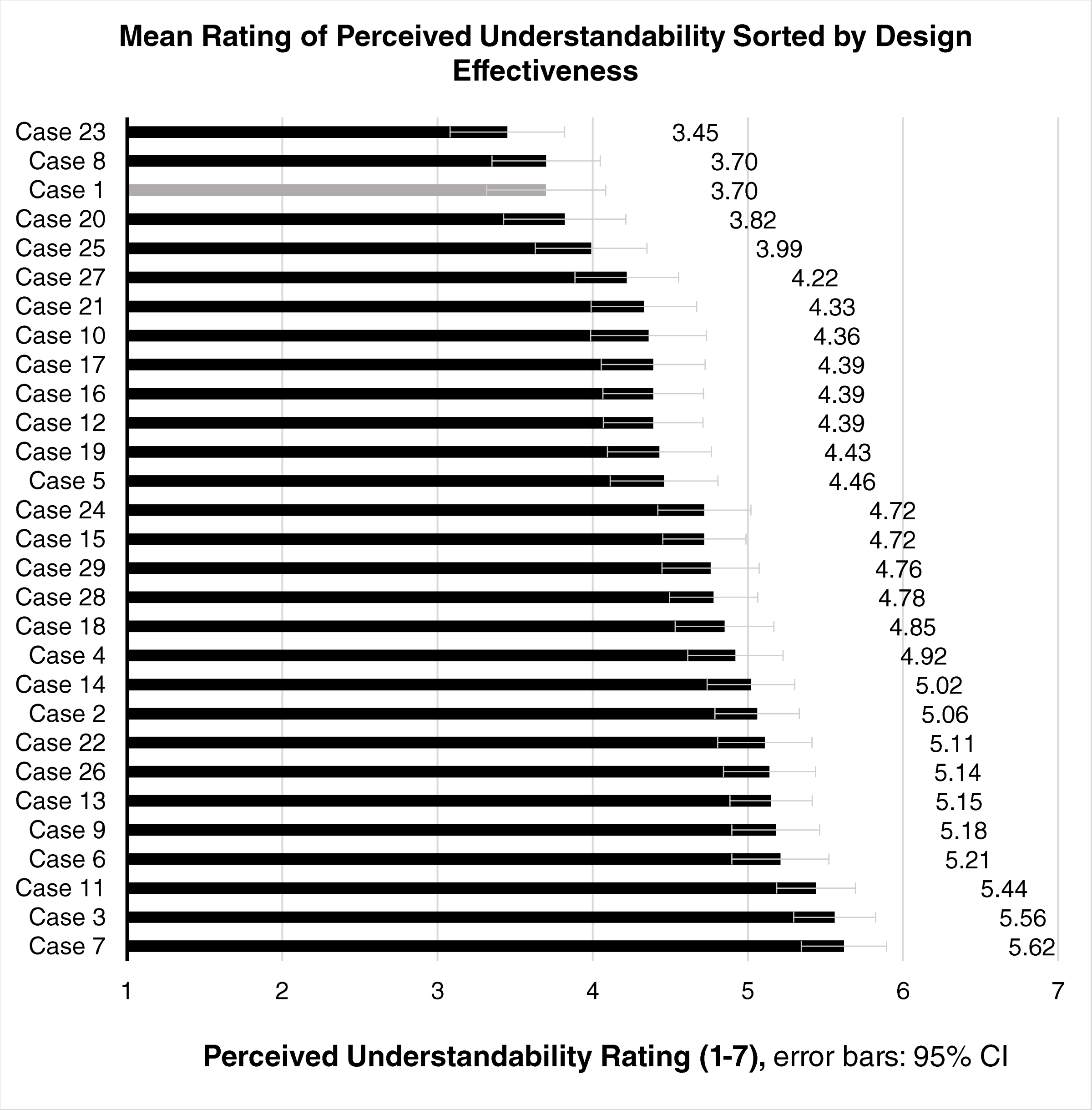


Figure 25. Mean rating of perceived understandability sorted by eHMI effectiveness (1 – least understandable, 7 – most understandable)

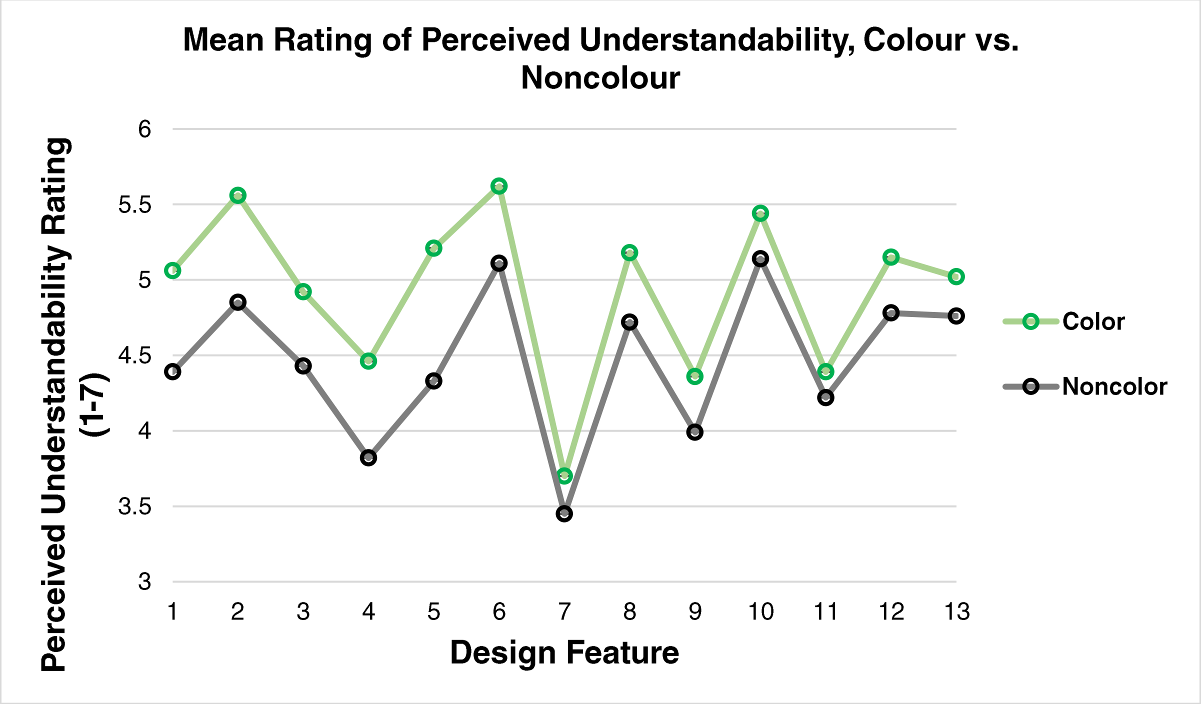


Figure 26. Mean rating of perceived understandability of coloured vs. non-coloured eHMI

5.8. Crossing Decision

The qualitative data or whether to cross before the car after the first image or after the second image or after the car has passed was transformed into an ordinal variable with three levels for data analysis. A Friedman test was conducted to determine if there were differences in the crossing decisions as a result of the different visual stimuli shown; results confirm that there was a statistically significant difference, $\chi^2(28) = 66.57$, $p < .0005$.

Table 16. Crossing decision data transformation

Response in Questionnaire	**Data Code**
Before the car passes me, between the first and the second image	1
Before the car passes me, between the second and the last image	2
After the car passes me	3

72.4% (21/29) of the design cases had a median of 3 - meaning that the majority of the participants would wait after the car has passed to proceed to cross the street. 6 cases had a median of 2 (crossing before the car passes, between the second and the last image, 2 seconds away from the pedestrian), and in descending ranking order, they were: case 26 – non-coloured external speedometer with walk/ stop text display, case 13 – coloured walk/ stop icon with gap estimation count down timer, case 7 – coloured walk/stop text display, case 6 – coloured tri-state led display, case 11 – coloured external speedometer with walk/ stop text display, and case 9 – coloured external speedometer with walk/stop icon display. Interestingly, case 28 – non-coloured walk/ stop icon with gap estimation count down timer and case 24 – non-coloured external speedometer with walk/stop icon display had a median crossing decision of 2.5. The median rating of 2.5 means, that on average, the participants were equally likely to cross after the car passes or just right before.

Table 17. Crossing decision in descending ranking order

Design Case	**Mean Rank**	**Median**
Case 26	*13.37*	*2*
Case 13	*13.59*	*2*
Case 7	*13.685*	*2*
Case 6	*13.84*	*2*
Case 11	*13.895*	*2*
Case 9	*14.085*	*2*

Case 28	*14.2*	*2.5*
Case 24	*14.26*	*2.5*
Case 22	14.575	3
Case 2	14.665	3
Case 3	14.775	3
Case 18	14.79	3
Case 29	14.84	3
Case 14	14.86	3
Case 4	14.99	3
Case 19	15.01	3
Case 10	15.085	3
Case 25	15.145	3
Case 15	15.505	3
Case 27	15.52	3
Case 12	15.62	3
Case 5	15.845	3
Case 16	15.915	3
Case 17	15.945	3
Case 20	15.99	3
Case 1	16.09	3
Case 21	16.095	3
Case 8	16.105	3
Case 23	16.71	3

5.9. Participant Feedback

The current research is a mixed-method study as the perception of design elements are not always able to be defined in precise, quantitatively measured dimensions. Enriching the quantitative dimensions, in an attempt to understand the attitudes and perceptual patterns, participants were asked the open-ended question - "What did you think about the design? Feel free to let us know your thoughts." - at the end of every design stimuli. The responses were collected from a text box in the online survey where the participants provided their personal opinions about the design.

The coding of the quality data from the participants provided a secondary measure to identify categorization and generalization about the perception of eHMIs for AVs in addition to the quantitative analysis offered by formal statistical analysis methods. It's important to note that this commentary feedback was voluntary; on average, the 28 design cases (excluding the base case of no design) had a response rate of 60.12%, N = 100.

Table 18. Code properties of qualitative data

	Count
Total responses	778
Number of codes from 1st iteration of coding	1231
Number of unique codes after 1st iteration of code analysis	466
Number of unique codes after 2nd iteration of code analysis	104
Number of categories of codes	5

The responses were coded solo with the author being the sole ethnographer. A number of qualitative coding methods were deployed to code the data in two cycles of coding. In the first cycle, attribute coding, structural coding, descriptive coding, values coding, motif coding, and narrative coding were used to identify 1231 codes (Saldana, 2013). Subsequently, in the second cycle, code patterns were identified and synthesized to eliminate repeated concepts. As a result, codes were mapped and categorized into 5 categories, shown in table 19. In general, the designs were more ambiguous (20.23% of all codes) than clear (9.99% of all codes) in intent communication.

Table 19. Five categories from code mapping

	Category	**Percentage of all codes**	**Code count**
1	Preference	63.36%	780
2	Ambiguity	20.23%	249
3	Clarity	9.99%	123
4	Colour	4.31%	53
5	Design Pattern	2.11%	26
		Total Code	1231

63.36% of the codes were about whether the design preferred or not. A negative response such as “the virtual driver seems unnecessary and takes focus away from the actual signal.”, was coded into “unpreferable”, “distracting”, “annoyance”, and “noise instead of signal”. A neutral response such as “useful, but I preferred the red and green lettering.” was coded into “neutral”, “useful”, and “lacking colour”. Meanwhile, a positive response such as “easy to understand since it has both words and icons” were coded into “preferable” and “dual signals are better than a singular signal”.

In perceptual preference, 60.38% of the codes reflected general negative feedback to the eHMI designs shown. This may not be indicative of design preferences, but rather a reflection of the general attitude towards the adoption of AVs.

Table 20. Category I. Preference

Category I. Preference	
Code	Percentage of categorical codes
unpreferable	60.38%
preferable	32.82%
neutral	6.79%

Shown in table 21, ambiguity, confusion, non-salience, information overload, and uselessness were the most frequently reoccurring themes that describe the general attitude of the ambiguity of AV's intent communication through the eHMI design shown - together accounting for 51.14% of the codes under the category of ambiguity.

Table 21. Category II. Ambiguity

Category II. Ambiguity	
Code	Percentage of categorical codes
ambiguous intent communication	17.43%
confusing	14.39%
nonsalient	7.96%
information overload	6.82%
useless	6.06%
annoyance	3.41%
creepy	3.41%
distracting	3.41%

mismatch in intent communication	3.03%
low perceived urgency	2.65%
unable to predict the future state	2.27%
silly	2.27%
confusing for multi-pedestrian scenario	1.89%
lacking intent communication	1.52%
requires further education	1.52%
uncomfortable	1.52%
risk-averse	1.52%
confusing for multi-vehicle scenario	1.14%
cognitive stress	1.14%
unhelpful	1.14%
deceptive	1.14%
irrelevant information	0.76%
ridiculous	0.76%
too subtle	0.76%
visual overload	0.76%
dangerous	0.76%
distrust	0.76%
unclear	0.76%
separate visual planes	0.76%
aesthetically lacking	0.76%
contradicting	0.38%
combined signals are ineffective	0.38%
concept of AV needs emphasis	0.38%
gimmick	0.38%
inconsistence	0.38%
language barrier	0.38%
need direction	0.38%
no behaviour impact	0.38%
noise instead of signal	0.38%
not culturally universal	0.38%

numbers not preferred	0.38%
scary	0.38%
size too small	0.38%
unnecessary	0.38%
unrealistic	0.38%
unsafe	0.38%
variability in gap acceptance	0.38%
earlier warning required for gap acceptance	0.38%
ignore	0.38%
interference with other road users	0.38%

Shown in table 22, usefulness, comprehension, clarity, safety, and understandability were the most frequently reoccurring themes that describe the general attitude of understanding AV's intent communication through the eHMI design shown - together accounting for 79.67% of the codes under the category of clarity.

Table 22. Category III. Clarity

Category III. Clarity	
Code	Percentage of categorical codes
useful	43.90%
comprehensible	12.20%
clear intent communication	11.38%
safe	6.50%
easy to understand	5.69%
helpful	3.25%
minimalistic	2.44%
salient	2.44%
simple	2.44%

informative	1.63%
dominant	0.81%
concise	0.81%
cool	0.81%
good use of multiple design space	0.81%
intuitive	0.81%
matching urgency level	0.81%
promotes safer crossing behaviour	0.81%
straightforward	0.81%
understood vehicle intent	0.81%
visible from side	0.81%

In terms of the influence of colour, shown in table 23, a lack of colour in non-colour (white) design cases and an emphasis on the importance of colour (red, green, and yellow) in coded responses accounted for 88.68% of the attitude towards colour in the eHMIs shown.

Table 23. Category IV. Colour

Category IV. Colour	
Code	Percentage of categorical codes
lacking colour	49.06%
importance of colour	39.62%
low visibility in snow-haze-fog	7.55%
message colour incongruency	3.78%

Table 24. Category V. Design Pattern

Category V. Design Pattern	
Code	Percentage of categorical codes
dual information sources are better than a singular source	23.08%
explicit reference is unpreferable	11.54%
anthropomorphic unpreferred	3.85%
anthropomorphic preferred	3.85%
comfort with similarity in design patterns	3.85%
auditory warning	3.85%
counter needed	3.85%
earlier warning required for gap acceptance	3.85%
ground projection is useless	3.85%
importance of icon	3.85%
knowing future state	3.85%
no design change	3.85%
preference of iconography	3.85%
the similarity to existing design patterns	3.85%
speedometer preferred	3.85%
street sign similarity	3.85%
text is better than an explicit reference	3.85%
timing	3.85%
the vehicle ahead as a visual cue for gap acceptance	3.85%

The coded response was not comprehensive since the response option was voluntary. However, for the design preference, the general attitude towards the influence of additive features, such as combining iconography with explicit vehicle kinematic reference, was received positively (23.08% of coded responses). For some participants (11.54%), explicit reference was not preferred on its own.

There are limitations to the current research method. For rapid iterative research cycles for a large number of design considerations, the research design prototype was created to be the lowest-fidelity but a minimally viable prototype. As a result, behaviour compliance, measurable through physiological signal measurements such as eye-tracking, heart rate, and pulse rate monitoring, as well as other modalities, especially auditory signal designs, weren't studied. These were common feedback from participants. This is expanded and discussed in section 7. Limitations and Future Research.

5.10. Summary of Design Types

The two independent variables included in the experiment design were design cases and colour (14 x 2 + baseline). In the case of the design case, the 29 different cases acted as the most detailed breakdown by differentiating designs (see Appendix A1). For the summary, the design cases were synthesized into and subsequently analyzed by design types (see Appendix A2).

A compilation of the results by the explicit ranking order is shown in Table 27. Summative Ranking: Explicit Sum of Individual Measurement Ratings. The five most effective designs in the cumulative measurement compilation of the explicit ranking order were shown in Table 25. Five Most Effective Designs in terms of Summative Ranking Novel design type J and type K (external speedometer with iconography and external speedometer with WALK/ STOP text) ranked 3rd and 4th in terms of design effectiveness.

The five least effective designs are shown in Table 26. Five Least Effective Designs in terms of Summative Ranking of All Measures Novel design type D, external speedometer ranked 3rd last and 4th last in terms of design effectiveness. In Section 6 Discussion, validation and rejection of hypotheses, findings in terms of research and design applications, and outliers in results are discussed in length.

Table 25. Five Most Effective Designs in terms of Summative Ranking of All Measures

Summative Ranking	Design Type	Dual vs Single Information Source	Novel vs. Existing Design	Design Case	Feature Name
1	E	Single 4	Existing 1	7	Walk/ Stop Text Display (colour)

2	H	Single 7	Existing 4	3	Walk/ Stop Icon (colour)
3	J	Dual 1	Novel 4	9	External Speedometer with Walk/Stop Icon Display (colour)
4	K	Dual 2	Novel 5	11	External Speedometer with Walk/ Stop Text Display (colour)
5	G	Single 6	Existing 3	6	Tri-state LED Display (colour)

Table 26. Five Least Effective Designs in terms of Summative Ranking of All Measures

Summative Ranking	**Design Type**	**Dual vs Single Information Source**	**Novel vs. Existing Design**	**Design Case**	**Feature Name**
24	I	Single 8	Existing 5	8	Single Line Road Projection (colour)
25	D	Single 3	Novel 3, Explicit Reference 3	10	External Speedometer with Speed Change Indicator (colour)
26	D	Single 3	Novel 3, Explicit Reference 3	25	External Speedometer with Speed Change Indicator (non-colour)
27	F	Single 5	Existing 2	20	Virtual Driver Display (non-colour)
28	I	Single 8	Existing 5	23	Single Line Road Projection (non-colour)
29	A	-	-	1	No Display (Baseline)

Table 27. Summative Ranking: Explicit Sum of Individual Measurement Ratings

Summative Ranking	Explicit Sum of Individual Measurement Ratings	Design Case	Perceived Urgency Rank (1 - Most Urgent)	Perceived Safety Rank (1 – Most Safe)	Perceived Pleasure Ranking (1 - Most Pleasant)	Perceived Arousal Rating (1 - Least Annoying)	Perceived Dominance Rating (1 - Most in Control) *	Perceived Usefulness Rating (1 - Most Useful)	Perceived Understandability Rating (1 - Most Understandable)
1	22	**7**	3	2	3	7	3	3	1
2	24	**3**	1	4	4	5	7	1	2
3	33	**9**	4	5	5	8	1	5	5
4	39	**11**	2	1	1	3	27	2	3
5	48	**6**	9	7	7	11	4	6	4
6	53	**22**	8	6	6	1	15	9	8
7	55	**13**	5	3	2	10	25	4	6
8	60	**26**	11	9	8	6	11	8	7
9	72	**14**	7	8	9	4	24	10	10
10	75	**28**	12	10	10	9	9	12	13
11	79	**4**	6	11	16	18	6	11	11
12	87	**2**	21	12	18	2	18	7	9
13	90	**29**	15	13	13	12	8	15	14
14	102	**24**	14	14	12	19	13	14	16
15	108	**5**	16	21	17	15	5	17	17
16	109	**18**	18	18	11	14	20	16	12
17	111	**15**	10	15	15	20	23	13	15
18	123	**19**	19	17	14	17	19	19	18
19	134	**12**	13	16	19	23	26	18	19
20	144	**17**	23	19	23	13	21	24	21
21	147	**21**	20	22	20	24	16	22	23
22	150	**16**	24	23	22	16	22	23	20
23	150	**27**	25	24	21	25	10	21	24
24	154	**8**	22	25	24	27	2	26	28
25	154	**10**	17	20	25	22	28	20	22
26	168	**25**	26	26	26	28	12	25	25
27	179	**20**	29	27	27	26	17	27	26
28	186	**23**	28	29	29	29	14	28	29
29	189	**1**	27	28	28	21	29	29	27

6. Discussion

6.1. Influence of Colour

Colour is confirmed as an important criterion to be considered for designers to make better sense of the different gender gaps that exist due to perception. Across all measures, seen in figures 14, 16, 18, 20, 22, 24, and 26, designs with colours were predominantly more effective in perceived urgency, perceived safety, perceived pleasure, perceived arousal, perceived dominance, and perceived usefulness, and perceived understandability than its non-colour counterpart. The hypothesis that colour designs were expected to outperform its non-coloured counterpart can be generally accepted. The two exceptions to this generalized rule exist in perceived arousal and perceived dominance. As previously discussed in the results section, both quantitative measures had four designs where the non-colour designs outperformed its coloured counterpart - see Table 13 and Table 15.

6.2. Novel vs. Existing Designs

Five hypotheses about the quantitative measurement for the pedestrian perception of AV eHMI designs of both novel and existing designs were made.

1. It was hypothesized that the three novel singular information source design concepts included in this study (design type B – external speedometer, design type C – time-to-arrival countdown, and design type D – speed change indicator, see Appendix A2) would outperform existing design

paradigms (iconography, explicit text indication, anthropomorphic features, etc.) in quantitative measures of pedestrian's perception of designs.

The hypothesis was rejected entirely. These three design types were included in 8 different design cases. In terms of summative ranking, design type B and C did not have any design case that was included in the top five. Furthermore, design type D - speed change indicator, both the colour (design case 10) and non-colour design case (design case 25), were the 3rd and 4th least effective cases in terms of summative ranking (table 27.)

2. Based on the literature, across all measures, icons (design type E) and egocentric text of "walk/stop" (design type H) were hypothesized as the clearest and useful AV-pedestrian eHMI designs known to date. In the results, design case 7, the coloured case for design type H, was shown to be the design case with the highest summative ranking. Meanwhile, design case 11, the coloured case for design type E, was the case with the second-highest summative ranking. Hence, the hypothesis can be generally accepted. For AV pedestrian interaction in North America, with English as a language of choice, eHMI elements of iconography and explicit text with egocentric messaging are affirmed to be the two most clear and useful single information source for AV-pedestrian eHMI design.

3. It was hypothesized that eHMIs with anthropomorphic features may not be perceived as safe and as comforting to the pedestrians than previous postulated. This was shown to be the case and the hypothesis is accepted. For the design type F – anthropomorphic features (coloured: design case

5, non-coloured: design case 20), the qualitative analysis indicated that the participants generally disliked the designs with anthropomorphic features, with coloured design resulting in 39.2% of codes indicating general discontent (N = 51, table 28), and non-coloured design resulting in 56.5% of the codes indicating displeasure with the design (N = 46, table 29).

Table 28. Design case 5 - anthropomorphic feature, colour – code count

Design Case 5 Code	Count	Percentage of Categorical Codes
unpreferable	**20**	**39.2%**
preferable	6	11.8%
neutral	4	7.8%
confusing	3	5.9%
cognitive stress	2	3.9%
creepy	2	3.9%
importance of colour scheme	2	3.9%
clear	1	2.0%
distracting	1	2.0%
mismatch of intent communication and anthropomorphic features	1	2.0%
no behavior impact	1	2.0%
requires further education	1	2.0%
ridiculous	1	2.0%
silly	1	2.0%
unclear	1	2.0%
unclear on intent communication	1	2.0%
unsafe	1	2.0%
useful	1	2.0%
visual overload	1	2.0%

Table 29. Design case 20 - anthropomorphic feature, non-colour – code count

Design Case 20 Code	Count	Percentage of Categorical Codes
unpreferable	**26**	**56.5%**
preferable	4	8.7%
creepy	2	4.3%
useful	2	4.3%
confusing	2	4.3%
dangerous	1	2.2%
distracting	1	2.2%
importance of colour	1	2.2%
low urgency	1	2.2%
low visibility in snow-haze-fog	1	2.2%
mismatch in perceived urgency	1	2.2%
mismatch of intent communication	1	2.2%
nonsalient	1	2.2%
silly	1	2.2%
unclear intent	1	2.2%

4. Lastly, on the perception of novel versus existing design patterns in AV pedestrian eHMIs, it was hypothesized that pedestrians will be generally risk-averse when interacting with AV eHMIs.

6.3. Influence of Additive Features

Single information source novel design concepts did not perform as strongly as initially hypothesized. However, using a novel design concept, such as external speedometer, as an additive feature was found to be a novel effective design concept.

Design type J – external speedometer with WALK/STOP icon display – along with design type K – external speedometer with WALK/STOP text – indicated high levels of design effectiveness through their summative ranking. Design type J and K, the coloured design case of the design type, were the 3rd and 4th most effective in their perspective summative ranking.

Furthermore, design case 11 – coloured external speedometer with WALK/ STOP text display - ranked 1st in perceived safety and perceived pleasure (most pleasant) and 2nd in terms of perceived usefulness and perceived urgency. Design case 9 – the coloured case for an external speedometer with walk/stop icon display – ranked 1st in perceived dominance, where the participant felt the most in control of their decisions to cross or not to cross.

6.4. Correlation of Measurements: Most Informative Measure

One of the key considerations of experiment design in the current study is to apply the lowest acceptable threshold in both design fidelity and research survey for scalability and reduced time and resource cost. A Spearman's rank-order correlation was run to assess the relationship between the eight quantitative measurement ranks and the summative rank to determine if there exists one single measure that is more informative than the others. This will be beneficial for future research to reduce experiment setup time and optimize research completion time and resource cost. Preliminary analysis showed that all relationships were monotonic, except for perceived dominance, as assessed by visual inspection of a scatterplot. All correlations were statistically significant and the strongest positive correlation between individual measurement rank and summative rank is from perceived understandability (table 30), rs(29) = .975, p < .0005. Perceived safety was second most informative as a single measurement for future research requiring heuristic evaluation; perceived usefulness being the third most informative.

Table 30. Correlation between Measurements

	Spearman's rho Correlation Coefficient, in Correlation with Summative Rank
Perceived Understandability Rank	0.975
Perceived Safety Rank	0.969
Perceived Usefulness Rank	0.966
Perceived Pleasure Rank	0.956
Perceived Urgency Rank	0.915
Perceived Arousal Rank	0.862
Perceived Dominance Rank	0.342

7. Limitations and Future Research

There are several limitations to this study. The current research is a controlled investigation where the message content, scenarios, and traffic context remain constant. Therefore, there is no presence of other road users, such as pedestrians or cyclists. As a result, the demonstrated behaviour of the vehicle is linear and simplistic and not a dynamic and naturalistic context. The traffic scenario is not complex and does not represent situations with bi-directional traffic and enriched visual distractions. The scenarios are only prepared for a single pedestrian point of view. This could be dangerous as egocentric messaging in a multi-pedestrian crossing situation could lead to unclarity and confusion.

Participants' cultural understanding of the local traffic rules and the nuances involved could lead to misunderstanding in the semantic meaning of designs. Of the 100 participants, 10 participants weren't geographically located in the U.S – the data originated from three different continents, from Kenya, India, Bangladesh, Argentina, Puerto Rico, and Honduras. Their response to the survey questions was not considered as outliers. However, this could change as cultural nuance possibly need a larger sample size with a more in-depth investigation into the cultural and lingual impact of design elements.

The fidelity of the designs is low as the prototypes weren't meant to be used in a formalized summative evaluation. The static images do not build a holistic situation awareness of the traffic dynamics. Relative distance and relative speed, as well as the perceived mental distance and speed, are not able to be measured as well as in a real-life experiment or VR study. The results can only be used as a model and approximation for pedestrian behaviours, to quickly funnel down the design decisions to find the most receptive designs to move forward and increase the design fidelity with. Furthermore, the data collected from the survey style test did not allow rich qualitative data to be collected.

This study was a study in perception of designs. A follow-up behavioural compliance study would be the logical next step. A realistic VR simulation that further investigates the most effective display designs in this study could be carried out to pedestrian's behavioural compliance of crossing decisions would be useful.

8. Conclusion

Under the assumption of the mass adoption of SAE level 4 high automation enabled vehicles, this thesis investigated the impact external human-machine interface has on pedestrians' crossing behaviour, as well as the pedestrian's general perception of AVs.

The current research showed a general positive influence of eHMI. The external display design improved the pedestrians' level of perceived safety, perceived urgency, perceived pleasure, perceived arousal, perceived dominance, perceived understandability, and perceived usefulness. This affirms that external displays can enhance the perceived safety and comfort levels of pedestrians when interacting with AVs. Within the same design type, colour design case, with few outliers, consistently outperformed the non-colour equivalent.

The dynamics of road users has evolved. The previous control and responsibility of the human driver shifted towards automated driving; however, the other road users have not changed as drastically in their behaviours. The most vulnerable road user remains the pedestrian. Despite a large number of design concepts suggested, few works of literature investigated designs aimed at strengthening the key mental model factors that influence crossing decisions.

An Amazon MTurk study was conducted with 100 participants, 90 from the U.S. The experiment was a within-subject design with 29 levels. There were 28 different stimuli and one baseline scenario with no designs included. The stimuli were structured into 14 (Design cases: see appendix x) x 2 (Coloured vs. Non-coloured Designs) factorial design. The current study created three novel design concepts introduced based on the lack of visual experimentation with displaying the key pedestrian crossing mental model factors of explicit vehicle kinematics, gap acceptance, and time-to-arrival.

Novel design types showed promise and open a conversation with the perception of eHMIs for AVs to improve pedestrian-AV communication. The three novel design types, as a single information source, did not perform as strongly as hypothesize. However, the novel design type of external speedometer, as an additive feature, when combined with iconography or text with egocentric messaging performed strongly. In the near future, these designs merit further investigation in behaviour compliance with a higher fidelity of experiment setup.

From the results, it can be concluded that AV pedestrian intent communication remains vital and the mode of the communication will have to adapt. The conventional design paradigm, in terms of lights and honking, used by manual vehicles of today will not be sufficient in order to improve or even maintain the lethality rate of pedestrians on roads.

The technological challenges of sensor accuracy, algorithmic efficiency, and system processing speed are tough issues to be solved. They are the foundations to enable mass adoption of AVs - for a safer, more efficient, more environmentally conscious transportation future. Great strides have been made as level 3 automation is a reality today. However, for mainstream ubiquity of fully autonomous vehicles, at level 4 and level 5 automation, to be realized, new regulatory frameworks and a greater public acceptance is the next uphill battle to be conquered. This research hopes to contribute to improve the perception of AVs and for the pedestrian to embrace new technological changes, instead of fear them.

9. References

Baldwin, C. L., Eisert, J. L., Garcia, A., Lewis, B., Pratt, S. M., & Gonzalez, C. (2012). Multimodal urgency coding: auditory, visual, and tactile parameters and their impact on perceived urgency. Work, 41(Supplement 1), 3586–3591.

Baldwin, C. L., & Lewis, B. A. (2014). Perceived urgency mapping across modalities within a driving context. Applied Ergonomics, 45(5), 1270–1277. https://doi.org/10.1016/j.apergo.2013.05.002

Bazilinskyy, P., Dodou, D., & de Winter, J. (2019). Survey on eHMI concepts: The effect of text, colour, and perspective.

Beggiato, M., Witzlack, C., & Krems, J. F. (2017). Gap Acceptance and Time-To-Arrival Estimates as Basis for Informal Communication between Pedestrians and Vehicles. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '17, 50–57. https://doi.org/10.1145/3122986.3122995

Benderius, O., Berger, C., & Malmsten Lundgren, V. (2018). The Best Rated Human–Machine Interface Design for Autonomous Vehicles in the 2016 Grand Cooperative Driving Challenge. IEEE Transactions on Intelligent Transportation Systems, 19(4), 1302–1307. https://doi.org/10.1109/TITS.2017.2749970

Böckle, M.-P., Brenden, A. P., Klingegård, M., Habibovic, A., & Bout, M. (2017). SAV2P: Exploring the Impact of an Interface for Shared Automated Vehicles on Pedestrians' Experience. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - AutomotiveUI '17, 136–140. https://doi.org/10.1145/3131726.3131765

Bohannon, J. (2011). Social science for pennies. American Association for the Advancement of Science.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry, 25(1), 49–59.

Bulligan, B. (1936). Collision in street and highway transportation. Philadelphia: Dorrance.

Bullough, John; Mark S. Rea (2001). Driving in Snow: Effect of Headlamp Colour at Mesopic and Photopic Light Levels (PDF). SAE Technical Paper Series. Archived from the original (PDF) on 2006-02-23.

Bullough, J. D., & Rea, M. S. (2001). Driving in snow: effect of headlamp colour at mesopic and photopic light levels. SAE Technical Paper.

Chang, C.-M., Toda, K., Igarashi, T., Miyata, M., & Kobayashi, Y. (2018). A Video-based Study Comparing Communication Modalities between an Autonomous Car and a Pedestrian. Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18, 104–109. https://doi.org/10.1145/3239092.3265950

Chapanis, A. (1994). Hazards associated with three signal words and four colours on warning signs. Ergonomics, 37(2), 265–275. https://doi.org/10.1080/00140139408963644

Chen, K.-M., Lue, X.-Q., Ji, H., Zhao, Y.-D., 2010. Towards the pedestrian delay estimation at intersections under vehicular platoon caused conflicts. Scientific Research and Essays 5 (9), 941–947.

Colley, A., Häkkilä, J., Pfleging, B., & Alt, F. (2017). A Design Space for External Displays on Cars. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - AutomotiveUI '17, 146–151. https://doi.org/10.1145/3131726.3131760

Curry, R., Blommer, M., Greenberg, J., & Tijerina, L. (2009). Immediate Recall of Driver Warnings in Forward Collision Warning Scenarios. Transportation Research Record: Journal of the Transportation Research Board, 2138(1), 28–33. https://doi.org/10.3141/2138-05

de Clercq, K., Dietrich, A., Núñez Velasco, J. P., de Winter, J., & Happee, R. (2019). External Human-Machine Interfaces on Automated Vehicles: Effects on Pedestrian Crossing Decisions. Human Factors, 1.87208E+13

Demiroz, Y. I., Onelcin, P., & Alver, Y. (2015). Illegal road crossing behavior of pedestrians at overpass locations: Factors affecting gap acceptance, crossing times and overpass use. Accident Analysis & Prevention, 80, 220–228. https://doi.org/10.1016/j.aap.2015.04.018

Devaux, P. (1970). State-of-the-Art Signaling and Lighting (paper, PDF). SAE 700386 Society of Automotive Engineers. Retrieved 21 May 2012

Dey, D., & Terken, J. (2017). Pedestrian Interaction with Vehicles: Roles of Explicit and Implicit Communication. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '17, 109–113. https://doi.org/10.1145/3122986.3123009

Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and Dynamics of Mechanical Turk Workers. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM '18 (pp. 135–143). Marina Del Rey, CA, USA: ACM Press. https://doi.org/10.1145/3159652.3159661

Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and Dynamics of Mechanical Turk Workers. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM '18, 135–143. https://doi.org/10.1145/3159652.3159661

Dreyfuss, E. (2018, November 29). A Bot Panic Hits Amazon's Mechanical Turk. Retrieved from https://www.wired.com/story/amazon-mechanical-turk-bot-panic/

Faber, L. G., Maurits, N. M., & Lorist, M. M. (2012). Mental Fatigue Affects Visual Selective Attention. PLoS ONE, 7(10), e48073. https://doi.org/10.1371/journal.pone.0048073

Fridman, L. (2018). Human-centered autonomous vehicle systems: Principles of effective shared autonomy. arXiv Preprint arXiv:1810.01835.

Fridman, L., Mehler, B., Xia, L., Yang, Y., Facusse, L. Y., & Reimer, B. (2017). To Walk or Not to Walk: Crowdsourced Assessment of External Vehicle-to-Pedestrian Displays. arXiv Preprint arXiv:1707.02698.

Godfrey, S.S., Laughery, K.R., Young, S.L., Vaubel, K.P., Brelsford, J.W., Laughery, K.A., Horn, E., 1991. The new alcohol warning labels: how noticeable are they? In: Proceedings of Human Factors Society 35th Annual Meeting. Human Factors Society, Santa Monica, CA, pp. 446–450.

Gill, R.T., Barbera, C., Precht, T., (1987). A comparative evaluation of warning label designs. In: Proceedings of Human Factors Society 31st Annual Meeting. Human actors Society, Santa Monica, CA, pp. 476–478.

Habibovic, A., Lundgren, V. M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., … Larsson, P. (2018). Communicating Intent of Automated Vehicles to Pedestrians. Frontiers in Psychology, 9 https://doi.org/10.3389/fpsyg.2018.01336

Hara, K. (2008). White. Lars Muller.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, 1–14. https://doi.org/10.1145/3173574.3174023

Hellier, E. J., Edworthy, J., & Dennis, I. A. N. (1993). Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. Human Factors, 35(4), 693–706.

Highway Traffic Act, R.S.O. https://www.ontario.ca/laws/statute/90h08 § c. H.8. -1990 Retrieved June 24, 2019

Johns, M., Sibi, S., & Ju, W. (2014). Effect of cognitive load in autonomous vehicles on driver performance during transfer of control. Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 1–4. ACM.

Katz, A., Zaidel, D., & Elgrishi, A. (1975). An experimental study of driver and pedestrian interaction during the crossing conflict. Human Factors, 17(5), 514–527.

Kline, P.B., Braun, C.C., Peterson, N., Silver, N.C., (1993). The impact of colour on warning research. In: Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting. Human Factors and Ergonomics Society, Santa Monica, CA, pp. 940–944.

Koh, P. P., & Wong, Y. D. (2014). Gap acceptance of violators at signalised pedestrian crossings. Accident Analysis & Prevention, 62, 178–185. https://doi.org/10.1016/j.aap.2013.09.020

Kyriakidis, M., de Winter, J. C. F., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., … Happee, R. (2019). A human factors perspective on automated driving. Theoretical Issues in Ergonomics Science, 20(3), 223–249. https://doi.org/10.1080/1463922X.2017.1293187

Lagström, T., & Lundgren, V. M. (2015). AVIP-Autonomous vehicles interaction with pedestrians. Master of Science Thesis, Chalmers University of Technology.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (I!&.), Technology in mental health care delivery systems (pp. 119-l 37). Norwood, NJ: Ablex.

Li, Y., Dikmen, M., Hussein, T. G., Wang, Y., & Burns, C. (2018). To Cross or Not to Cross: Urgency-Based External Warning Displays on Autonomous Vehicles to Improve Pedestrian Crossing Safety. Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 188–197. ACM.

Lu, Z., Happee, R., Cabrall, C. D. D., Kyriakidis, M., & de Winter, J. C. F. (2016). Human factors of transitions in automated driving: A general framework and literature survey. Transportation Research Part F: Traffic Psychology and Behaviour, 43, 183–198. https://doi.org/10.1016/j.trf.2016.10.007

Mahadevan, K., Somanath, S., & Sharlin, E. (2018). Can Interfaces Facilitate Communication in Autonomous Vehicle-Pedestrian Interaction? Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18, 309–310. https://doi.org/10.1145/3173386.3176909

Matthews, M., Chowdhary, G., & Kieson, E. (2017). Intent communication between autonomous vehicles and pedestrians. arXiv Preprint arXiv:1708.07123.

Mitchell, M. (2010). The development of automobile speedometer dials. Visible Language, 44(3), 331-366. Retrieved from https://s3-us-west-2.amazonaws.com/visiblelanguage/pdf/44.3/the-development-of-automobile-speedometer-dials.pdf.

Naujoks, F., Purucker, C., & Neukum, A. (2016). Secondary task engagement and vehicle automation – Comparing the effects of different automation levels in an on-road experiment. Transportation

Research Part F: Traffic Psychology and Behaviour, 38, 67–82. https://doi.org/10.1016/j.trf.2016.01.011

NHTSA, 2016. Traffic Safety Facts. May. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812270

Paolacci, G., & Chandler, J. (2014a). Inside the Turk: Understanding Mechanical Turk as a participant pool. Current Directions in Psychological Science, 23(3), 184–188.

Paolacci, G., & Chandler, J. (2014b). Inside the Turk: Understanding Mechanical Turk as a participant pool. Current Directions in Psychological Science, 23(3), 184–188.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 30(3), 286–297.

Pawar, D. S., & Patil, G. R. (2016). Critical gap estimation for pedestrians at uncontrolled mid-block crossings on high-speed arterials. Safety Science, 86, 295–303. https://doi.org/10.1016/j.ssci.2016.03.011

Petzoldt, T. (2014). On the relationship between pedestrian gap acceptance and time to arrival estimates. Accident Analysis & Prevention, 72, 127–133. https://doi.org/10.1016/j.aap.2014.06.019

Rasouli, A., & Tsotsos, J. K. (2018). Autonomous Vehicles that Interact with Pedestrians: A Survey of Theory and Practice. arXiv Preprint arXiv:1805.11773.

Reig, S., Norman, S., Morales, C. G., Das, S., Steinfeld, A., & Forlizzi, J. (2018). A Field Study of Pedestrians and Autonomous Vehicles. Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18, 198–209. https://doi.org/10.1145/3239060.3239064

Rothenbücher, D., Li, J., Sirkin, D., Mok, B., & Ju, W. (2016). Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 795–802. IEEE.

Samuel, A. (2018). Amazon's Mechanical Turk has Reinvented Research. Retrieved March 22, 2019, from https://daily.jstor.org/amazons-mechanical-turk-has-reinvented-research/

Schneemann, F., & Gohl, I. (2016). Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. 2016 IEEE Intelligent Vehicles Symposium (IV), 38–43. IEEE.

Schreuder, D.A., 1976, White or Yellow Light for Vehicle Head-Lamps?. SWOV Publication 1976-2E. SWOV (Dutch Institute for Road Safety Research. 1976 Retrieved 29 July 2018

Schreuder, D.A. (1976). White or yellow light for vehicle head-lamps? Arguments in the discussion on the colour of vehicle head-lamps. Publication 1976-2E. Voorburg, SWOV, 1976

Stanton, N. A., & Young, M. S. (1998). Vehicle automation and driving performance. Ergonomics, 41(7), 1014–1028. https://doi.org/10.1080/001401398186568

Stanton, Neville A., & Marsden, P. (1996). From fly-by-wire to drive-by-wire: safety implications of automation in vehicles. Safety Science, 24(1), 35–49.

Strano, M., Novak, F., Walbert, S., Palmeiro, B., Morales, S., & Alvarez, I. (2018). "Peace of Mind", An Experiential Safety Framework for Automated Driving Technology Interactions. 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 53–59. IEEE.

SWOV (2012). SWOV Fact sheet: Vulnerable road users. SWOV, Leidschendam. Retrieved from https://www.swov.nl/en/publication/vulnerable-road-users.

Terwilliger, J., Glazer, M., Schmidt, H., Domeyer, J., Toyoda, H., Mehler, B., … Fridman, L. (2019). Dynamics of Pedestrian Crossing Decisions Based on Vehicle Trajectories in Large-Scale Simulated and Real-World Data. arXiv Preprint arXiv:1904.04202.

TRB, 2010. Highway capacity manual 2010. Transportation Research Board of theNational Academies, Washington, US.

Tresilian, J.R. (1995). Perceptual and cognitive processes in time-to-contact estimation: Analysis of prediction-motion and relative judgment tasks. Perception & Psychophysics 57, 2, 231–245. http://dx.doi.org/10.3758/BF03206510.

Vissers, L. K., van der Schagen, S., & INLG van & Hagenzieker, M. P. (2017). Safe interaction between cyclists, pedestrians and automated vehicles: what do we know and what do we need to know?

Volz, B., Mielenz, H., Gilitschenski, I., Siegwart, R., & Nieto, J. (2019). Inferring Pedestrian Motions at Urban Crosswalks. IEEE Transactions on Intelligent Transportation Systems, 20(2), 544–555. https://doi.org/10.1109/TITS.2018.2827956

Wang, J., 2010. Pedestrian's critical cross gap and its application in conflict simulation. Proceedings of the International Conference on Intelligent Computation Technology and Automation, 11–12 May 2010, Changsha, China.

Wickens, C. D., Li, H., Santamaria, A., Sebok, A., & Sarter, N. B. (2010). Stages and levels of automation: An integrated meta-analysis. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54, 389–393. Sage Publications Sage CA: Los Angeles, CA.

Winkielman, P., Schwartz, N., Fazendeiro, T. A., & Reber, R. (2016). The Hedonic Marking of Processing Fluency: Implications for Evaluative Judgement. The Psychology of Evaluation: Affective Processes in Cognition and Emotion.

Wogalter, M.S., Conzola, V.C., Smith-Jackson, T.L., (2002). Research-based guidelines for warning design and evaluation. Applied Ergonomics 33 (3), 219e230. http://dx.doi.org/10.1016/s0003-6870(02)00009-1.

Wu, J.-P., Huang, L., Zhao, J.-L., 2004. The behavior of cyclists and pedestrians at signalized intersections in Beijing. Journal of Transportation Systems Engineering and Information Technology 4 (2), 105–114.

Yang, S. (2017). Driver behavior impact on pedestrians' crossing experience in the conditionally autonomous driving context.

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. Accident Analysis & Prevention, 78, 212–221. https://doi.org/10.1016/j.aap.2015.02.023

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. Transportation Research Part C: Emerging Technologies, 98, 207–220. https://doi.org/10.1016/j.trc.2018.11.018

10. Appendix A1: Design Features in Correlation with Design Cases

Design Feature	**Design Case, Colour**	**Design Case, Non-colour (White)**	**Feature Name**
NA	1	1	No Display (Baseline)
1	2	17	External Speedometer
2	3	18	Walk/ Stop Icon
3	4	19	Time-to-arrival Countdown
4	5	20	Virtual Driver with Anthropomorphic Features Display
5	6	21	Tri-state LED Display
6	7	22	Walk/ Stop Text Display
7	8	23	Single Line Road Projection
8	9	24	External Speedometer with Walk/Stop Icon Display
9	10	25	External Speedometer with Speed Change Indicator
10	11	26	External Speedometer with Walk/ Stop Text Display
11	12	27	External Speedometer with Single Line Road Projection
12	13	28	Walk/ Stop Icon with Gap Estimation Count Down Timer
13	14	29	Walk/ Stop Icon with Virtual Driver Display
14	15	NA	Gap Estimation Count Down Timer (RYG)
15	NA	16	Numerical Speed Display

11. Appendix A2: Design Types' Correlation with Design Features and Design Cases

Design Type	Dual vs Single Information Source	Novel vs. Existing Design	Design Feature(s)	Design Case(s)
A	-	-	No Display (Baseline)	1
B	Single 1	Novel 1	External Speedometer	2, 16, 17

Design Case 2

Design Case 16

Design Case 17

C	Single 2	Novel 2	Time-to-arrival Countdown	4, 15, 19

Design Case 4

Design Case 15

Design Case 19

D	Single 3	Novel 3	Speed Change Indicator	10, 25

Design Case 10

Design Case 25

E	Single 4	Existing 1	Walk/ Stop Icon	3, 18

Design Case 3

Design Case 18

F	Single 5	Existing 2	Anthropomorphic Features	5, 20

Design Case 5

Design Case 20

G	Single 6	Existing 3	Tri-state LED Display	6, 21

Design Case 6

Design Case 21

H	Single 7	Existing 4	Walk/ Stop Text Display	7, 22

Design Case 7

Design Case 22

I	Single 8	Existing 5	Single Line Road Projection	8, 23

Design Case 8

Design Case 23

J	Dual 1	Novel 4	External Speedometer with Walk/Stop Icon Display	9, 24

Design Case 9

Design Case 24

K	Dual 2	Novel 5	External Speedometer with Walk/ Stop Text Display	11, 26

Design Case 11

Design Case 26

L	Dual 3	Novel 6	External Speedometer with Single Line Road Projection	12, 27

Design Case 12

Design Case 27

M	Dual 4	Novel 7	Walk/ Stop Icon with Gap Estimation Count Down Timer	13, 28

Design Case 13

Design Case 28

N	Dual 5	Novel 8	Walk/ Stop Icon with Virtual Driver Display	14, 29

Design Case 14

Design Case 29

12. Appendix A3: Design Case and YouTube Links

Design Case	**Design Case Name**	**Colour vs No Colour**	**Youtube Link**	**Display Feature 1**	**Display Feature 2**
1	No Display	Colour	https://youtu.be/dfLF6630vfY	No Display	
2	External Speedometer	Colour	https://youtu.be/pEEvrvu4FBE	External Speedometer	
3	Icon Display	Colour	https://youtu.be/I-9EBRDggV	Walk/ Stop Icon	
4	Gap Estimation Count Down	Colour	https://youtu.be/vjpkpqPTcUA	Gap Estimation Count Down Timer	
5	Virtual Driver Display	Colour	https://youtu.be/KTUkqv4YnUg	Virtual Driver Display	
6	LED Display	Colour	https://youtu.be/quNYitgYmhU	LED Display	
7	Text Display	Colour	https://youtu.be/QB0pNAn9-0U	Text (WALK/ STOP) Display	
8	Road Projection	Colour	https://youtu.be/ZEi4_-vPt7g	Single Line Road Projection	
9	Speedometer with Icon Display	Colour	https://youtu.be/bLkMIs0WM-w	External Speedometer	Walk/ Stop Icon

10	Speedometer with Change Indicator	Colour	https://youtu.be/2s_Y84Rhcg0	External Speedometer	Speed Change Indicator
11	Speedometer with Text Display	Colour	https://youtu.be/ZwsJQ7BgHpk	External Speedometer	Text Display
12	Speedometer with Road Projection	Colour	https://youtu.be/5TQ89z8pNM0	External Speedometer	Road Projection
13	Countdown Timer with Icon Display	Colour	https://youtu.be/vlc2ClZ0-Do	Walk/ Stop Icon	Gap Estimation Count Down Timer
14	Icon Display with Count Down Timer	Colour	https://youtu.be/_q0YD8tHMoA	Walk/ Stop Icon	Virtual Driver Display
15	Count Down Timer (Red Yellow Green)	Colour	https://youtu.be/mzId3b_1fwY	Count Down Timer	Red Yellow Green
16	Simple Speed Display	No Colour (White)	https://youtu.be/dnpxfkdrhx8	Numerical Speed Display	
17	Speedometer (White)	No Colour (White)	https://youtu.be/FfzmXVHGvLw	External Speedometer	

18	Icon Display (White)	No Colour (White)	https://youtu.be/tsjTe5Zi9Xg	Walk/ Stop Icon	
19	Count Down Timer (White)	No Colour (White)	https://youtu.be/bflI6ZK4EhA	Gap Estimation Count Down Timer	
20	Virtual Driver (White)	No Colour (White)	https://youtu.be/imN5ByJeXAQ	Virtual Driver Display	
21	LED Indicator (White)	No Colour (White)	https://youtu.be/xP3T79pPdrw	LED Display	
22	Text Display (White)	No Colour (White)	https://youtu.be/jBx2nS3cVIQ	Text (WALK/ STOP) Display	
23	Road Projection (White)	No Colour (White)	https://youtu.be/FN1Yu1W9IRw	Single Line Road Projection	
24	Speedometer with Icon Display (White)	No Colour (White)	https://youtu.be/TNllzWLGjDY	External Speedometer	Walk/ Stop Icon
25	Speedometer with Change	No Colour (White)	https://youtu.be/-HkEpvUL-4s	External Speedometer	Speed Change Indicator

	Indicator (White)				
26	Speedometer with Text Display (White)	No Colour (White)	https://youtu.be/bFcaqjFK98E	External Speedometer	Text Display
27	Speedometer with Road Projection (White)	No Colour (White)	https://youtu.be/z3ejWUlsVNE	External Speedometer	Road Projection
28	Count Down Timer with Icon Display (White)	No Colour (White)	https://youtu.be/YyxxL4Qhgys	Walk/ Stop Icon	Gap Estimation Count Down Timer
29	Virtual Driver with Icon Display (White)	No Colour (White)	https://youtu.be/6Mb2f-oXBvc	Walk/ Stop Icon	Virtual Driver Display

13. Appendix B: Survey Questionnaire

Information Letter

Thank you for your time. Please read this thoroughly.

You are invited to participate in a research study conducted by David Qin, under the supervision of Catherine Burns at the Department of Systems Design Engineering of the University of Waterloo, Canada. This study aims to understand the crossing behaviour of the pedestrian when interacting with novel external display designs on fully autonomous vehicles. We want to answer two key research questions in this study. First, with driver-less, fully autonomous vehicles, to what degree should the vehicle communicate its intent with pedestrians? Second, how can we design new means of communications (e.g., visual intent designs on exterior displays) on autonomous vehicles that are effective in a pedestrian-crossing situation? The first research question explores the prioritization of different factors (e.g., display signals, the mental model of vehicle speed) of pedestrians when making crossing decisions. The second research question would provide implications for designing future communication mechanisms.

If you decide to participate in our study, you will be asked to do an online survey that is completed anonymously. The online survey includes a demographic questionnaire, a pedestrian-related questionnaire and external display design evaluation questionnaire. All these activities, including breaks, will take about 45 - 60 minutes. You will be paid $2 at the end of the study.

Your participation in this study is voluntary. You may decline to answer any questions that you do not wish to answer. You can withdraw your participation at any time by ceasing to answer questions, without penalty or loss of remuneration. Any information you provided up to that point will not be used.

To withdraw, please close the SurveyGizmo page you're on and send an email to zehao.qin@uwaterloo.ca to obtain a unique code for this HIT. Please keep your Amazon Mechanical Turk page open and enter the unique code to receive remuneration.

Your participation in this research will be completely anonymous and data will be averaged and reported in aggregate. Possible outlets of dissemination include academic conferences and journals. Although your participation in this research may not benefit you personally, it will greatly help us understand how people's perception of potential new designs on future commercial vehicles. There are no risks to individuals participating in this survey. You may experience minor levels of eye strain and muscular discomfort from sitting for a period gazing at a computer screen, but there are no risks to you beyond those that exist in normal day-to-day personal computer usage.

It is important for you to know that any information that you provide will be confidential. The website is programmed to collect responses alone and will not collect any information that could potentially identify you (such as machine identifiers). When information is transmitted over the internet confidentiality cannot be guaranteed. University of Waterloo practices is to turn off functions that collect machine identifiers such as IP addresses. The host of the system such as Amazon Mechanical Turk may collect this information without our knowledge. We will not own this information.

The data, with no personal identifiers, collected from this study will be maintained on a password-protected computer database in a restricted access area of the university. As well, the data will be electronically archived after completion of the study and maintained for seven years.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE# 40819) If you have questions for the Committee contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca

Should you have any questions about the study, please contact either David Qin (zehao.qin@uwaterloo.ca) or Catherine Burns (catherine.burns@uwaterloo.ca).

Thank you for your consideration in partaking this study.

Participation

With full knowledge of all foregoing, I agree, of my own free will, to participate in this study. By providing your consent, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities. *
I agree to participate.

(Please close the browser immediately if you don't want to participate.)

Eligibility

Are you proficient with reading and writing in English? *
[ ] Yes
[ ] No

Do you have normal vision in both eyes with or without the aid of corrective lenses? *
[ ] Yes
[ ] No

What's your age range? *
[ ] 18-24 years old
[ ] 25-34 years old
[ ] 35-44 years old
[ ] 45-54 years old
[ ] 55-64 years old
[ ] 65 years or older

As a pedestrian, how often do you cross streets per day?*
[ ] None
[ ] Less than 5 times

[ ] 5 to 10 times
[ ] More than 10 times

Demographics

Gender*
[ ] Male
[ ] Female
[ ] Other

Which measurement system are you familiar with? (If you are familiar with both systems, please choose the one you are more familiar with.) *
[ ] mph (miles per hour)
[ ] km/h (kilometers per hour)

How long have you had your current driver's license? *
[ ] I do not currently have a driver's license
[ ] Less than 5 years
[ ] 5 to 10 years
[ ] More than 10 years
[ ] I prefer not to answer

How many miles have you driven in the past month? *
[ ] 0 mile
[ ] 1 - 100 miles
[ ] 101 - 500 miles
[ ] 501 miles - 1000 miles
[ ] 1001 miles or more
[ ] I prefer not to answer

How many kilometres have you approximately driven in the past month? *

[ ] 0 km
[ ] 1 - 100 km
[ ] 101 - 500 km
[ ] 501 - 1000 km
[ ] 1001 km or more
[ ] I prefer not to answer

Do you currently own a vehicle with an Advanced Driving Assistance System? (i.e. lane departure warning system, automatic lane centring)
[ ] Yes
[ ] No
[ ] I prefer not to answer

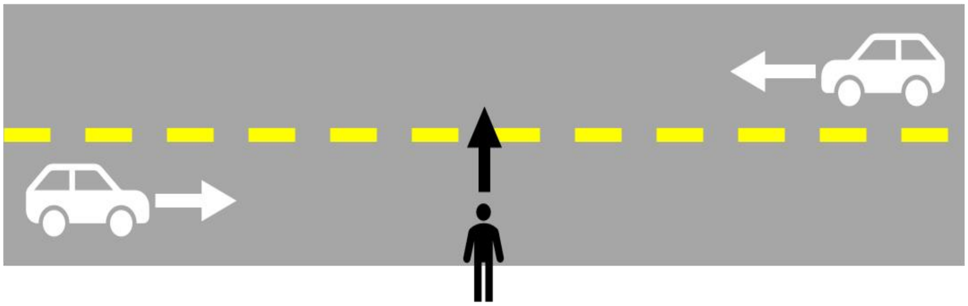
Pedestrian Experience Questions I

No Stop Sign

No Traffic Light

No Yield To Pedestrian Sign

No Painted Yield Bars



Suppose you are a pedestrian trying to cross an undivided road. There are no stop signs, no traffic lights, no yield to pedestrian signs, and no painted yield bars. You find a car approaching you from either your left- or right-hand side.

As a pedestrian that intends to cross the road, what are you most likely to do?*

[ ] Keep the current walking speed, to cross before the car
[ ] Walk faster, to cross before the car
[ ] Walk slower and make a complete stop, to let the car drives through first
[ ] Walk slower but not make a complete stop, to let the car drives through first
[ ] Keep the current walking speed, to let the car drives through first

Do you have a road crossing experience that is particularly memorable for you? Please tell us about it.

Pedestrian Experience Questions II

As a pedestrian, please rank the following reasons based on their importance to your decision on what to do while crossing a street.*

	Very Unimportant	Unimportant	Somewhat Unimportant	Neutral	Somewhat Important	Important	Very Important
Your walking speed							
Distance between you and the car that tries							

to drive through							
Speed of other cars on the road							
Distance between you and the other side of the road							
Driver's intent communication (i.e. eye contact, hand gestures)							
Distance between the vehicles in traffic							

Weather and time of the day							
How other pedestrians behave							
The curvature of the road							
Signs on the road							

As a pedestrian, please rate the importance of the following ways you would use to communicate with the driver to determine if it is safe to cross.*

	Very Unimportant	Unimportant	Somewhat Unimportant	Neutral	Somewhat Important	Important	Very Important
Eye-contact							
Hand gesture							

Body movement							

As a pedestrian, are there other ways you would use to communicate with the driver?*
[ ] Yes
[ ] No

Please tell us which other way you would communicate with the driver:

Autonomous Vehicle Questions

Should autonomous cars communicate with pedestrians in the future? *
[ ] Yes
[ ] No

In your opinion, how would a future autonomous car interact with pedestrians? (Choose all that apply)*
[ ] Using lights to interact with pedestrians
[ ] Using intelligent interfaces via a screen placed onto the external parts of the autonomous car
[ ] Using voice interfaces to interact with pedestrians
[ ] Projecting information to the ground in front of the autonomous car
[ ] Other (Please specify): *
[ ] No additional communication signals needed

If an autonomous car approaches you when you cross a street, what kind of information could be presented by the autonomous car to increase your trust in the autonomous car? (Choose all that apply)*

[ ] Showing explicit information about whether the car is an autonomous car or a conventional car
[ ] Showing explicit information about the autonomous car’s next action (slow down, speed up)
[ ] Showing explicit information about its current speed
[ ] Showing explicit information about the guidance to pedestrians (such as “go ahead” or “please stop”)
[ ] Showing no information just like a normal family car
[ ] Other (Please specify): *

In your opinion, what are the main factors that may affect your trust in autonomous cars? (Choose all that apply)*
[ ] The stability of the autonomous system
[ ] The level of intelligence of the autonomous system
[ ] The ability of the autonomous system to interact with people in the autonomous car
[ ] The ability of the autonomous system to interact with pedestrians
[ ] The ability of the autonomous system to handle emergency traffic incidents
[ ] Time to respond to instructions
[ ] The ability of the autonomous system to diagnose and process errors
[ ] The safety of the autonomous system
[ ] Who takes the ultimate responsibility in control, the autonomous car or the drivers/passengers in the autonomous car
[ ] Other (Please specify): *

In your opinion, how would you compare the design of future autonomous cars and the design of conventional cars? (Choose one)*
[ ] The more similar the better
[ ] There should be a clear distinction between the design of future autonomous cars and the design of conventional cars
[ ] The ability to interact with drivers/passengers in the autonomous car
[ ] It doesn't matter
[ ] Other (Please specify): *

What's your attitude towards autonomous cars? (Choose one)*
[ ] Very Distrustable

[ ] Distrustable
[ ] Somewhat Distrustable
[ ] Neutral
[ ] Somewhat Trustable
[ ] Trustable
[ ] Very Trustable

Scenario Instructions

Please read thoroughly

Important: In this part of the survey, we will show you a series of images. Each scenario involves a hypothetical case of a pedestrian and an autonomous car.

It's a Sunday afternoon in the wintertime. Snow and ice have been cleared from the road. Suppose you are looking at the images from the perspective of a pedestrian that stood on the curb and decided to cross the street. An autonomous vehicle is coming towards you. As the pedestrian, you find yourself in the middle of the roadblock and you intend to cross the road to the other side. There are no stop signs, no traffic light, no yield to pedestrian sign, and no painted yield bars.

The autonomous vehicle detects your intent to cross the road and attempts to communicate with you its intent. There will be different external display designs shown on the vehicle. Each communication design will be different. After showing you the design under the scenario context, we would like to ask you about your opinion about the communication design.

Each video will be played at least once. Replay the video again if conceptual clarification is required. If you decide to watch the video again, feel free to stay as long as you want on the page replay the video.

After the scenario is played, click next, where we ask you a few questions about the design.

Check if you are a survey bot

Captcha is an anti-bot service provided by Google. No respondent information is sent to Google as part of this service. Visit Google's reCAPTCHA page to learn more.

Please select B as your answer choice.
[ ] A
[ ] B
[ ] C
[ ] D
[ ] E

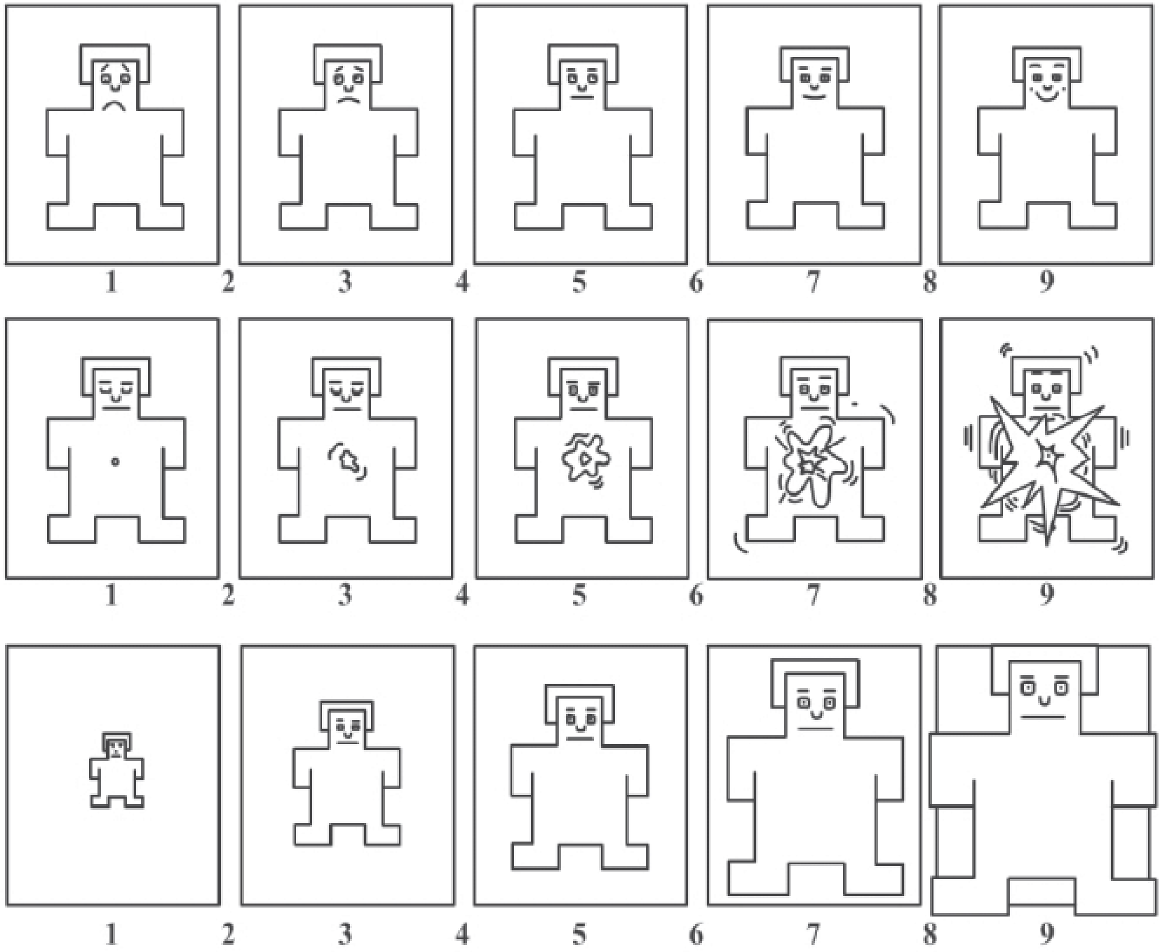
Experiment Scenario Questions

These questions were repeated after every single scenario

Suppose you were the pedestrian in the previous video that stood on the curb.

How urgent was the warning? (0 to 100)*
0 ________________________[__]_____________________________ 100

How safe did you feel in this encounter? *
[ ] Very Unsafe
[ ] Unsafe
[ ] Somewhat Unsafe
[ ] Neutral
[ ] Somewhat Safe
[ ] Safe
[ ] Very Safe



How did the encounter make you feel?

Look at the image above and select the options for each row*

	1	2	3	4	5	6	7	8	9
Row 1									

Row 2									
Row 3									

How useful do you think the warning would be?*

[ ] Very Useless

[ ] Useless

[ ] Somewhat Useless

[ ] Neutral

[ ] Somewhat Useful

[ ] Useful

[ ] Very Useful

How likely would you understand what the autonomous car tried to do? *

[ ] Very Unlikely

[ ] Unlikely

[ ] Somewhat Unlikely

[ ] Neutral

[ ] Somewhat Likely

[ ] Likely

[ ] Very Likely

Suppose you were the pedestrian in the previous video, when would you start crossing the street?*

[ ] Before the car passes me, between the first and the second image

[ ] Before the car passes me, between the second and the last image

[ ] After the car passes me

[ ] Other - Write In (Required): *

What did you think about the design? Feel free to let us know your thoughts

Feedback Letter

Dear participant,

I wish to extend a warm thank you for your participation in this study. Autonomous vehicles (AV) or self-driving vehicles (e.g., Waymo and Tesla self-driving cars) prompts the emergence of a new set of communication challenges that must be addressed to ensure the safety and effectiveness of all agents on the road. One of the most vulnerable groups in road agents is pedestrians. Prior to AVs, conventional vehicles had a human driver that could communicate with pedestrians through several human signals; signals such as eye gaze, head movements, and hand and arm gestures. With the introduction of AVs, pedestrians can no longer rely on such communication signals. In the future, when all of the control and responsibilities of the human driver gradually transfers the autonomous driving program, the vehicle's intent communication to pedestrians will have to evolve as well.

This study aims to understand the crossing behaviour of the pedestrian when interacting with novel external display designs on fully autonomous vehicles. We want to answer two key research questions in this study. First, with driver-less, fully autonomous vehicles, to what degree should the vehicle communicate its intent with pedestrians? Second, how can we design new means of communications (e.g., visual intent designs on exterior displays) on autonomous vehicles that are effective in a pedestrian-crossing situation? The first research question explores the prioritization of different factors (e.g., display signals, the mental model of vehicle speed) of pedestrians when making crossing decisions. The second research question would provide implications for designing future communication mechanisms.

To address these research questions, we designed this online survey that involves participants from a general population to respond to survey questions from the perspective of different road users. From this survey, we expect to assess possible designs under simulated conditions of future road and vehicle paradigms. Results of this survey may provide design patterns of autonomous vehicles that can improve pedestrian-crossing safety.

The data, with no personal identifiers, collected from this study will be maintained on a password-protected computer database in a restricted access area of the university. As well, the data will be electronically archived after completion of the study and maintained for seven years.

If you are interested in knowing the results of the study, please contact David Qin (zehao.qin@uwaterloo.ca) and you will be able to receive a summary of the findings once available.

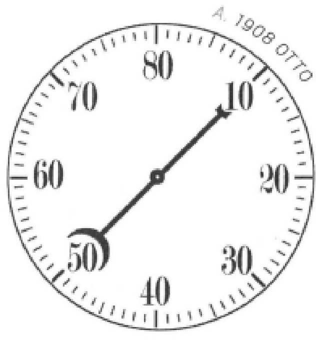
This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE#40819) If you have questions for the Committee contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

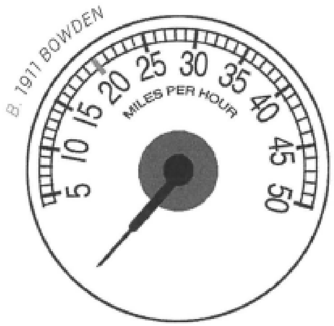
Should you have any questions about the study, please contact David Qin (zehao.qin@uwaterloo.ca) or Catherine Burns (catherine.burns@uwaterloo.ca).

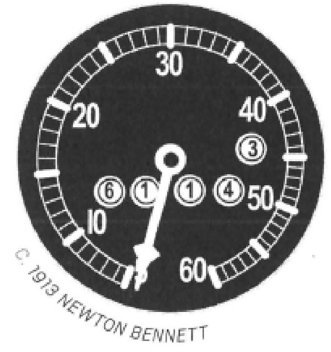
Thank you for your participation in this study.

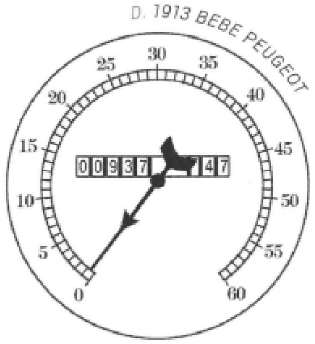
14. Appendix C: Related Speedometer Designs

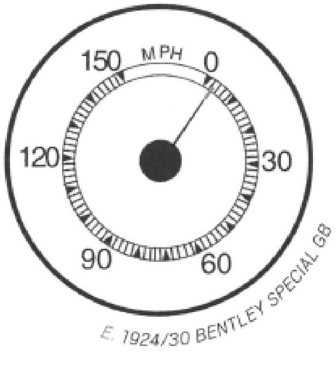
From Marilyn Mitchell, 2010, The Development of Automobile Speedometer Dials: a Balance of Ergonomics and Style, Regulation and Power



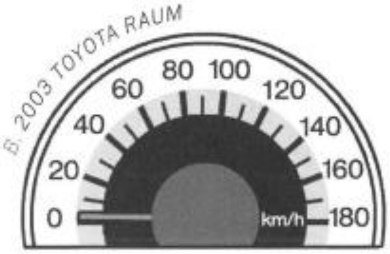








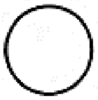

A. 2003 TOYOTA MATRIX
0
20
40
60
80
100
120
140
MPH
km/h
20
40
60
80
100
120
140
160
180
200
220
240
80


B. 2003 TOYOTA RAUM
0
20
40
60
80
100
120
140
160
180
km/h

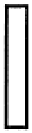

0
10
20
30
40
50
60
70
80
90
100
120
140
160
180
200
220
240
260
280
C. 2004 AUDI S4

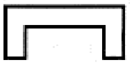
DIAL SHAPE

Circular or semi-circular



Vertical or horizontal, linear



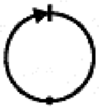


Unique

REFERENCE POINTS AND DIRECTIONS ON CIRCULAR OR SEMI-CIRCULAR DISPLAYS

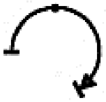
Start 7:00–9:00, end 3:00–5:00,
motorway limit near 12:00

Start and end at or near 12:00,
motorway limit near 6:00



Some early designs,
based on clocks or timers

Asymmetrical



Fits around other information,
information, and maintains
visual ratio scale for high speeds

Start near 5:00, end near 1:00

Racing

Start near 4:00, end near 8:00

Eccentric

Start and end near 6:00,
motorway limit near 12:00

REFERENCE POINTS AND DIRECTIONS ON LINEAR DISPLAYS

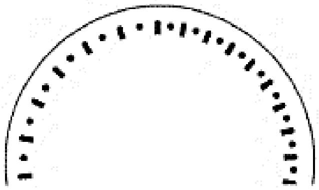
Reading order

Moves upward

TYPES OF SCALES

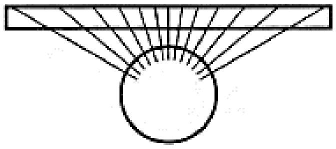
Ratio

Two or three connected
ratio scales



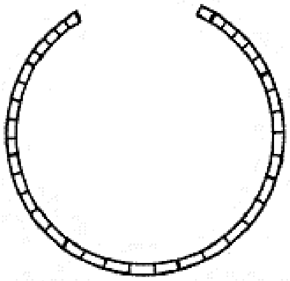
Conserves space, retains symmetry
on dials that have high top speeds

Semi-circular ratio scale
mapped onto horizontal line



Unique

Uneven



Early design based on early technology

NUMBER OF SCALES

1 (mph or km/h)

2 concentric (one for mph and one for km/h)

2 concentric (each for a different gear)

NUMBER OF SCALES MARKS

1 mark = 1 mph, major mark every 5 mph	*Early design*
1 mark = 2 mph, major mark every 10 mph	
1 mark = 5 mph, major mark every 10 mph	
1 mark = 1 km/h, major mark every 5 km/h	*Early design*
1 mark = 2 km/h, major mark every 10 and 20 km/h	
1 mark = 5 km/h, major mark every 20 km/h	
1 mark = 10 km/h, major mark every 20 km/h	*Racing*

SCALE MARK SHAPES

Major scale marks	▮ Line	● Circle	‖ Double lines	◆ Diamond	▲ Triangle
				Some earlier designs	
Minor scale marks	▮ Line	● Circle	*Marks may be joined by single or double track lines* *Identically shaped minor marks may be smaller*		

NUMBERS PRESENTED ON THE DIAL

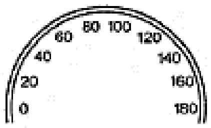
(0), 10, 20, 30, 40... mph	
(0), 20, 40, 60, 80 ... mph	
30, 60, 90, 120... mph	*unique*
10, 30, 50, 70...mph	*unique*
(0), 10, 20, 30, 40... km/h	
(0), 20, 40, 60, 80... km/h	

FONT

Serif	*Early design*
Sans-serif	*Follows ergonomic design recommendations*

PLACEMENT OF NUMBERS

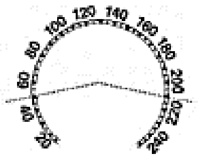
Straight up



Follows ergonomic design recommendations

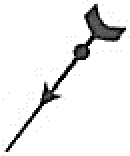
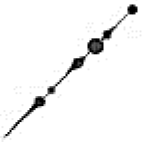
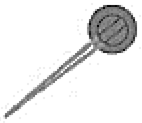
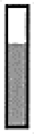
Angled

Angled and turned



Early design

POINTER

Decorative			*Early design*
Simple			*Follow ergonomic design recommendations*
Band of color			*Unique*

MOVING SCALE VERSUS MOVING POINTER

Moving scale	*Early design, Unique*
Moving pointer	*Follows ergonomic design recommendations*

COLOR

Light background, contrasting graphics and pointer

Dark background, contrasting graphics and pointer

Key reference point may have unique color