2020

Determining and Assessing Fault Attribution in Collisions Involving Autonomous Vehicles

Alexandra Kaplan
University of Central Florida

DETERMINING AND ASSESSING FAULT ATTRIBUTION IN COLLISIONS INVOLVING AUTONOMOUS VEHICLES

by

ALEXANDRA DANIELA KAPLAN
B.A. New York University, 2011
M.A. University of Central Florida, 2019

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2020

Major Professor: Peter A. Hancock

ABSTRACT

There exists considerable research concerning how humans attribute fault to each other, both in cases of accidents and those instances of intentional harm. There also exist studies involving blame attribution towards robots, when such robots have caused harm through operational failure or lack of safety features. However, relatively little work has, to date, examined the ways in which fault is attributed to self-driving vehicles involved in collisions, despite many newspaper and popular articles which both report past incidents and warn of future risk. This dissertation examined fault attribution in collisions involving autonomous vehicles by conducting three separate experiments. The first experiment placed participants in the roles of witnesses to a collision, and compared fault attributed to an autonomous vehicle to fault attributed to a regular, manually-operated vehicle, when those cars were involved in identical collisions. The second, and third experiments explored the fault that operators attributed to both themselves and autonomous vehicles when involved in a collision, whether they were the operator of the autonomous vehicle or the operator of a regular car that shared the road with automated ones. Results showed that, across experiments, perceived avoidability of the collision was the largest predictor of fault regardless of whether the participant was a witness or a driver. Additionally, participants in all three experiments thought themselves in general to be better than average drivers.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Peter Hancock, Provost Distinguished Research Professor, Pegasus Professor and Trustee Chair, for his support and assistance during my graduate school career. I would also like to thank the other members of my committee, Dr. Mustapha Mouloua, Dr. Valerie Sims, and Dr. Gerald Matthews, who have all helped me both in relation to this dissertation, and in numerous other ways. Additionally, I thank the many others who have served as research mentors such as Amanda Bond of Soar Technology, and Dr. J. Christopher Brill of the Air Force Research Laboratory.

I thank my parents for their guidance and patience. Without their support I never could have come this far, and I am grateful that they helped me carry the burdens of graduate school. Particularly, I am thankful for my mother's wisdom and advice on statistics.

Finally, I thank all the other friends, family members, fellow students, lab members, and cats who have helped me along this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Perhaps the first autonomous mode of human transportation was the horse. Trains, planes, and automobiles that have followed each proved faster and more efficient but lacked the horses' ability to choose its own path, i.e., present a degree of self-determination. This is now changing (Hancock, 2020). More and more, companies are beginning to manufacture and test autonomous vehicles for private and commercial use. The cars of the future (and, in some more limited ways, the present) range from fully self-driving to semi intelligent with some autonomous features in the way of lane-assists, self-parking abilities, and other novel capacities. In giving drivers a supervisory, rather than active role in driving, many collisions can be circumvented. However, some accidents are bound to happen, whether due to environmental factors, other humans on the road, or even problems with the autonomous vehicle's algorithmic control (Hancock, 2019). When these collisions happen, it is important to consider the question: who do people fault when autonomous vehicles crash?

Although self-driving cars are still relatively rare, some have already been involved in deadly collisions. In 2016, a former Navy Seal was killed in an accident while his Tesla Model X was in autopilot mode (MacRae, 2016). The Tesla collided with a truck broadside. This was thought to be due to the white, uniform side of the truck, which resembled open sky to the vehicle's vision system. Later that same year, a driver in China was killed when his Tesla Model X rear-ended a road-cleaning truck shortly after engaging autopilot mode (Horowitz & Timmons, 2016). Two years later, a Tesla Model S was involved in a fatal accident, after the vehicle's owner complained that its autopilot always steered towards the highway divider (Green, 2018). More recently, a pedestrian was struck and killed by a self-driving Uber in

Arizona. So far, the latter example is the only known case of a pedestrian death relating to an autonomous vehicle (Lubben, 2018). These deaths represent the most disastrous consequences of crashes involving self-driving vehicles. Many more minor incidents have occurred with relatively minimal damage to driver or vehicle (Hawkins, 2019).

The above examples of collisions have already happened with only a relatively limited number of autonomous vehicles on the road. As autonomous vehicles become more numerous, more collision situations are likely to occur. Those who design such vehicles have to consider what the vehicle's response will be in these ambiguous but foreseeable situations. If a car stops short in front of an autonomous vehicle, while another car is following too closely behind, the programmer must decide, ahead of time, whether such circumstances will require the vehicle to stop, and thus risk being rear-ended; to swerve, and thus risk collision with traffic in another lane, or to strike the stopped car and thus risk injury (and see Hancock & De Ridder, 2003). With human drivers, these decisions are made in fractions of a second on the road, using a lifetime of real-world experience. True, they are often made incorrectly, a fact which is reflected in the record of damaging events in collision and accident databases. Those who program the algorithms of self-driving cars have the advantage of a longer period to consider the spectrum of all the possible outcomes. Yet at the same time this can be seen as a disadvantage in having to determine, well ahead of time, which course of action to take. This means potentially taking on both legal and ethical responsibility for the result of such events (Nyholm & Smids, 2016). After all, if a programmer has designed a vehicle to stop immediately behind a disabled car, they are certainly at least partially responsible for the outcome if rear-ended. Designing for such situations has often been compared to The Trolley Problem (Thomsen, 1985). The latter is an

ethical-dilemma posed as a thought exercise in which a trolley driver must choose between staying on-course and striking several people, or switching tracks and striking only one (Foot, 1967). Of course, as road vehicles are not on tracks, there are more than two discrete options. Thus the associated moral quandary becomes even more complex. That, and the fact that one's choice in the hypothetical Trolley Problem does not lead to real-life lawsuits. However, the victims of collisions with self-driving vehicles can certainly seek litigation, and in some cases already have done so (O'Kane, 2019).

The fact that autonomous vehicles can, have, and will be involved in collisions underscores the need to consider where fault will be placed when these events occur. To date, perhaps the most well-publicized death has been that of Joshua Brown, who was allegedly watching a Harry Potter movie on a personal DVD player when his Tesla collided with the side of a truck (Macrae, 2016). In this incident, the Tesla's camera was unable to distinguish the difference between the side of a white truck, and the sky. Thus, it did not recognize any need to stop and drove, without braking, into the side of the truck's trailer. Here, there is an obvious failure on the part of the automation in not recognizing a truck, which a human would easily notice. It is, of course, important to note that human-driven "truck run-under" collisions are themselves relatively frequent at night-time and in the twilight hours. Here, the sensory capacities of human drivers frequently fail to register the dark-colored trailer in the low light conditions. In essence, both are similar events in which we witness failures of recognition due to issues of dynamic conspicuity (Hancock, 2019). In this particular case of autonomy's failure, there is also a failure on the driver's part to properly monitor the automation. Tesla's autopilot putatively requires constant observation and the human driver is encouraged to take over at any

sign of danger. How exactly this can be done has yet to be adequately specified. As the driver did not survive the crash, we can only assume that he was too distracted by Harry Potter's escapades to notice the truck. But it is also probable that he expected his Tesla would stop itself. And one can just as easily fault the insufficient camera as the distracted driver. When programming the cameras of an autonomous vehicle with examples of objects it may encounter on the road, and which it ought to avoid striking, 'trucks' is one of the first items to come to mind. In this case, the collision had an obvious cause (failing to notice an oncoming truck) but no obvious causer.

Types of Autonomous Vehicles

There are several types of autonomous vehicle, and not all of them are fully self-driving. Both the National Highway Traffic Safety Administration (NHTSA: 2013) and the Society of Automotive Engineers (SAE: 2016) have developed rankings of levels of automation in vehicles. The Tesla involved in the previously described collision might well be identified as a level two (see Table 1). This means that it was not responsible for monitoring itself, a shortcoming which is supposedly fully communicated to consumers.

Table 1: *SAE levels of automation in self-driving vehicles.*

SAE level	Definition
0	No automation; only the human driver is in control of the vehicle.
1	The vehicle can assist the human driver with some aspects of driving.
2	The vehicle can drive itself, with the human monitoring. The human must take over as needed, and must be the one to determine when such a takeover is needed.
3	The vehicle can drive itself and monitor itself, but will request for the human to take over when needed.
4	The vehicle can drive itself and monitor itself fully in some, but not all, conditions.
5	The vehicle can drive and monitor itself fully, in all the same situations in which a human can do so.

The higher levels of automation may truly be considered "driverless" cars. In these cases, there is no classic driver, only a passenger. The passenger cannot then be faulted for any collision. The only choice that they make is whether or not to get into the vehicle. However, at levels 2 and 3, the driver continues to share control with the autonomous vehicle. Here, there is room for doubt as to who is responsible for preventing a collision and who is at fault when these events do occur. These levels are where current automation and the automation of the near future fall. Thus far we establish that collisions can happen, and that, at least currently, control of the vehicle is shared between the driver and the automation. Fault is therefore potentially shared

between these entities or skewed towards one or the other dependent upon circumstances. There are two main groups of people who may attribute fault. First are witnesses to the collision, whether they were involved or not. The other individuals are the drivers of the car(s), who may perhaps distribute fault between themselves and the vehicles involved. There are several factors which might serve to predict in which direction fault is attributed. Of course, other forensic and legal professionals seek to apportion blame or influence its apportionment during the investigative and legal proceedings which follow (Hancock, 2019). It is to these aspects of attribution that I now turn.

The "Above Average" Effect

People tend to rate themselves as being better than average, esecially when it comes to positive traits (Alicke, 1985; Alicke & Govorlin, 2005). Young people and adolescents in particular believe that they are unique and special, and that the consequences that affect other people do not always apply to themselves (Elkind, 1967). This is also true in the domain of operating motor vehicles. People are likely to describe their own driving skills as safer than the average motorist (Matthews & Moran, 1986). In general, people believe that they are better and safer drivers than others in their situation. For example, they believe they are better able to drive while impaired by alcohol or lack of sleep than their immediate peers (Wohleber & Matthews, 2016). However, logically, not all drivers can be better than average, and this collective overconfidence can lead to poor decisions on the road. It appears that most are somewhat aware that they inflate their own abilities. In fact, people are able to understand that the rating they give their driving abilities is higher than the rating other people would give to them (Roy & Liersch, 2013). However, this does not stop them from having high opinions of their own ability. It is

necessary to consider these beliefs in conjunction with the possibility that automated vehicles will require more ability and effort, not less, from drivers. At levels 2 and 3 of automation, the driver/operator is required to remain vigilant and be able to take over when and if necessary. Fault for any collisions can then fall entirely on the human operator of the vehicle, for failing to intervene properly. However, this is to misunderstand and misconceive the inherent human capacity for sustained attention (Hancock, 2013). Blame can also fall on the operations of the automated vehicle itself. Since most people believe themselves to be better drivers than they truly are, it stands to reason that, when they are the operator of the vehicle, they will fault the vehicle more than themselves especially when they are asked to judge another operator/machine pairing. They may also fault the other driver regardless of whether any specific automation errors led to the collision. There is, of course, the legal motivation to avoid responsibility which also feeds into this assessment. Prior to hypothesizing about aspects of fault attribution it is important to consider the cognitive biases that influence how people fault themselves, other people, and non-human entities for the course of any set of events.

How People Fault Other People

Some disasters happen without any immediately identifiable cause. Yet, people still attempt to seek out causal factors. This is true of both large-scale tragedies (Veltfort & Lee, 1943) as well as personal situations such as a child's illness (Chodoff, Friedman, & Hamburg, 1964). In general, it is quite common for people to seek to identify one causal influence in accidents rather than concede the fact that something horrible occurred purely from chance alone. Blame occurs when people identify a causal influence, and believe that the perpetrator will not take action to avoid future incident (Bucher, 1957).

Causal Attribution:

Heider (1958) examined the ways in which people attempt to account for social behavior. These explanations can stem from external situational factors, or, alternatively internal factors. Heider argued that people's biases can lead them to make incorrect assumptions about the nature of another person and therefore judge them inaccurately. People tend to assume that the bad actor is operating based on an internal characteristic. That is, they think that someone who behaved poorly did so because they are a bad person, and not because they were influenced by an outside circumstance. Of course, there are flaws associated with these forms of assumptions.

Fundamental Attribution Error:

Attribution is the property via which one derives the perception of an individual's motivations from their actions, either correctly or incorrectly (Kelley, 1973). There are many who believe that by simply witnessing the outcome of a situation, they can determine what motivated the actor in that situation to behave in the way they did. They believe they can also tell whether the outcome was intentional on the part of that individual. That is to say, a motorist who is cut off by another car, might assume that the offending driver acted intentionally. Of course, this is not always the case. The other driver may have been distracted, or swerving to avoid an obstacle in the road, or had any number of motivations for their behavior that did not intentionally result in causing a near-collision (and see Lewin, 1936).

Defensive Attribution:

One influence which subsumes the need to assign fault, concerns the Defensive-Attribution Hypothesis (Walster, 1966). This idea indicates that people assign fault in collisions

because they wish to believe that the such events are controllable, and thus preventable. If it was both controllable and preventable, and happened anyway, then it follows that someone was responsible for the failure in that controlling and preventing role, either unintentionally, maliciously, or through incompetence. Assigning fault, in this theory, allows individuals to feel that they are safe from recurrence of the same incident, since it is not the circumstances themselves which led to said incident but rather another person's failure. According to Walster, "*If he decides that someone was responsible for the unpleasant event, he should feel somewhat more able to avert such a disaster"* (1966, pp. 74)." Walster found that that this assignment of fault was more common in severe accidents with serious consequences, but less likely in minor accidents. Though this was not replicated in subsequent studies (see Shaver, 1970), an overall meta-analysis has confirmed Walster's proposition (Burger, 1981).

Culpable Control:

The Culpable Control Model has examined an individual's volitional, and causal, responsibility in the actions leading to a negative outcome (Alicke, 2000; and see Figure 1). This model indicates that, when evaluating other individuals in order to potentially attribute fault for an event, people do not unconsciously assign fault but take into consideration the other person's motivation, or mental affect, as well as their behavior. Thus, they use both of these elements to determine whether or not the other person is truly at fault. The Mind to Behavior link considers whether one's actions or behaviors are intentional and whether they had any control over those actions. The Behavior to Consequence link examines whether a person's behavior really led to the consequence, and if so, to what extent was it causal compared to other situational factors?

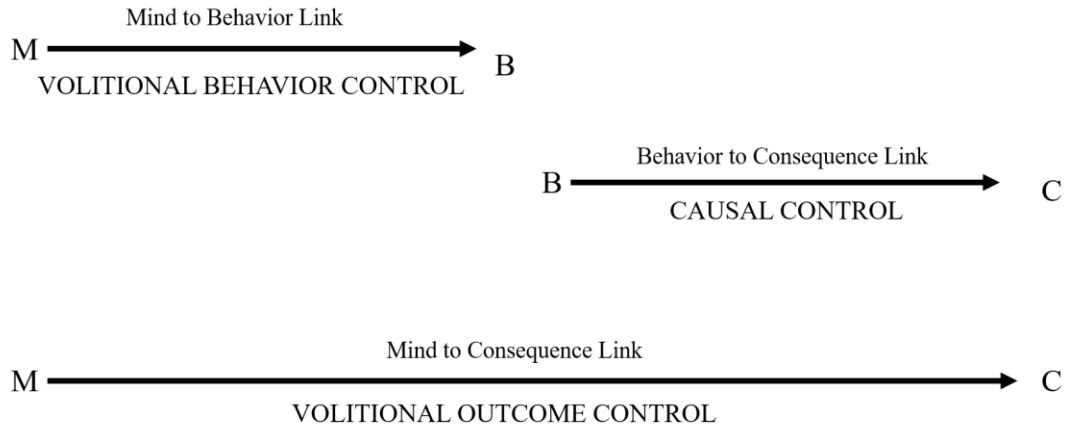


Figure 1: The link between one's mental state, behavior, and subsequent outcomes. B= Behavior, M= Mental, C = Consequence (Alicke, 2000).

One may wonder how the culpable control model can be applied to automation. The link between behavior and consequence is clear; the automation's behavior may have some role in causing an incident. However, machines lack agency in any current theory of "mind" that would give them any kind of volitional control over the outcome, or their behavior. In essence, automation is only an actor, and not a thinker. Whether this lack of control will exonerate them in a spectator's mind, or rather deny them one element of the causal control model, remains to be seen.

Just-World:

The Just World Hypothesis, also called the Just World Fallacy, is the assumption that the consequences of a person's actions are deserved (Lerner, 1980). According to this theory, people believe that good actions are rewarded and bad actions are punished (Kushner, 1981). The result of this concept is the idea that if something bad happened to a person, then that person must have somehow deserved it. This idea plays into attribution in the form of victim-blaming, or crediting

higher fault to the injured party than the situation calls for. This relies on the fallacious assumption that the injured person must be at least partly to fault.

Theory of Social Conduct:

Weiner (1995) notes that there is a difference between actions caused by one's lack of effort, versus those from a lack of ability. That is, a person might be faulted for a collision that came about as a result of their own laziness or unwillingness to act appropriately (such as maintaining proper driving etiquette). The perpetrator in these cases, at some point, made a decision to not check their mirror or to engage in whatever action lead to the incident. This is a different situation from those who have caused collisions due to a lack of ability (such as not being able to react quickly enough in order to stop before rear-ending another car). If an individual lacks an inherent ability, they are perceived to have no control over the situation and thus cannot be blamed for the outcome.

How People Fault Machines

The ways in which people attribute fault to non-human entities may or may not necessarily be influenced by the same inherent cognitive biases that they exhibit in relation to human attribution. There are many factors involved in interaction with machines which are not present in similar interactions between people. For instance, intelligent machines possess different levels of autonomy. Some machines may be entirely independent, while others rely on input from their operators (see Table 1). Thus, a non-autonomous machine may simply represent a blameless tool and any responsibility for an associated event would lie entirely with its human operator. A fully autonomous machine, however, can be seen as separate from its operator, and

in acting alone, can cause damage without any human input. In this case, responsibility lies with the machine to the extent that that is legally feasible. In the case of incidents involving machines with middling autonomy, any attribution of fault can be shared between the machine and its operator.

In support of this proposition, Kim and Hinds (2006) found that when working with a robot, participants were less likely to attribute either credit or fault to themselves and the other humans involved, and more likely to attribute those elements to the robot, when the robot was highly autonomous. However, knowing that an autonomous agent may be faulted for its actions, it is important to consider the differences between blaming a human and blaming an autonomous machine. For instance, even the most highly autonomous technology is, to some extent, dependent on its original designer and programmer for guidelines as to how to respond in new situations. Harkening back to Weiner's (1995) theory of social conduct, if people believe that the robot does not have control over its own actions, then the consequences of such actions are due to a lack of ability, not a lack of effort, and thus not faultworthy. Van der Woerdt and Haselager (2017) demonstrated that when a robot failed at the task of putting away a toy, viewers attributed both greater agency and more responsibility to the robot whose failure was a lack of effort (throwing the toy on the floor) than to the robot whose failure resulted from lack of ability (being physically unable to pick up the toy). Bigman, Waytz, Alterovitz and Gray (2019) have also observed that people see robots as capable of having intentions but not desires. Robots are viewed as lacking free will, which makes it harder to attribute fault to them. The latter authors note that most people do not *want* machines to be responsible for making moral decisions, particularly not those that lead to life and death consequences.

Moral Decisions in Autonomous Driving

Notwithstanding the above, autonomous vehicles *must* make decisions, and due to the nature of driving some of those decisions must eventually be life or death matters. Those situations have been likened to an applied trolley problem (Nyholm & Smids, 2016). There are times when collisions prove unavoidable, and consequently one must contend with the issue that a self-driving algorithm must bear both the benefit and the burden of precognition. Programmers do not know that a collision will occur, but they know that it might, and thus they have time to prepare for many eventualities (including, but not limited to, the vehicle's choice in which hypothetical trolley track to take). Unlike the driver in the trolley problem, or drivers in real collision situations, there is no element of time constraint in making these decisions. In essence, an algorithm has already decided which direction the vehicle will take, long before the situation occurs. Thus, its response is more choice than chance. Automated vehicles must be equipped to make these sorts of decisions but in what fashion will they be blamed for any negative outcomes? The problem becomes progressively more complex as ever greater momentary information is loaded into that decision.

In a study where automated vehicles were faced with an unavoidable collision and could hit one of two virtual pedestrians, participants were not happy to learn the inner workings of the algorithm used to determine which of these two would be hit, particularly as the algorithm's utilitarian programming determined the worth of each human's life based on their profession and gender (Wilson, Theodorou & Bryson, 2019). This is somewhat in contrast to a study which found participants blamed robots, more than people, for failing to take the "best" but utilitarian option in an applied trolley problem (Malle et al., 2015). These opposing findings beg the

question as to what degree do people want their automation to be pragmatic? For safety's sake, this remains a problem only in controlled laboratory studies, and not in the real world. A national ethics committee has already decreed that, in the instance of unavoidable collision where the car must strike a person, the decision cannot be based on any attributes of the person such as age, gender, or physical status. It is legislatively forbidden (Ethics Commission Report, 2017). Whether this moral design imperative is ubiquitously adhered to remains to be seen (Hancock, 2009).

Knowing that drivers do not necessarily want to be privy to the processes behind the pragmatic decisions which must be taken by autonomous vehicles in incipient collision situations, and knowing also that, at level 2 or 3 automation, such vehicles share responsibility with the human operator (see Table 1), it is currently unclear how fault will be attributed in collisions involving self-driving vehicles. How will people attribute fault when there is both a driver and an algorithm controlling the steering wheel? And what biases will lead them to alter their attributions when they are either a witness to, or participant in, such events?

Rationale for the Present Research

In light of the foregoing observations, the present work has examined the ways in which people attribute fault in cases of collisions involving self-driving vehicles. As automated vehicles become more common, so, too, will their involvement in collisions. It is important to consider exactly how people will fault both the driver, and the car itself, when such collisions do occur. Much theoretical work has been written about the morality of creating algorithms that induce life threatening events. Relatively little empirical work has examined the ways in which people

respond to such collisions, both the real collisions which have occurred as well as the hypothetical ones which may be an issue in the future. It is also important to understand the ways in which people attribute fault to the vehicle and to themselves when they are the operator. With most people believing themselves to be better drivers than average, it is possible that, as a whole, they will fault other operators who have accidents involving automated vehicles, and yet blame the vehicle when they themselves are the ones at the wheel.

To evaluate such propositions, Experiment 1 was designed to compare the ways in which people attribute fault to the driver of a regular vehicle involved in a collision with the ways they attribute fault to the driver of an automated vehicle, as well as the vehicle itself, involved in that same situation. Participants viewed collision scenarios with the information that one of the cars involved was either an automated vehicle, or a human-driven vehicle. They then viewed the collision scenario and judged the extent to which each player involved was at fault for said collision. Within the condition where participants viewed an automated vehicle, fault was subdivided into that shared between the driver and the vehicle's algorithms. Additionally, participants were asked to what extent the collision was unavoidable. If the collision was indeed unavoidable, then no one was truly at fault.

Experiment 2 placed the participants in the position of the controller of an autonomous vehicle. They were required to navigate complex driving situations with the help of a Safety Suggestion Algorithm, which gave them advice on what actions to take at each opportunity. Mainly, the algorithm made the correct recommendation, but occasionally, it was wrong. Participants had to correctly choose whether to agree or disagree with the algorithm at each step,

with the goal of always making the right choice. If they made the wrong choice, they were told that they their drive has ended in a collision. They were also asked to rate their own driving abilities. It was anticipated that their confidence in their own abilities would be correlated with their attribution of fault to the vehicle rather than themselves.

Experiment 3 had participants manually controlling an on-screen vehicle, navigating through a crowded scene full of other, moving vehicles. They had to maneuver their vehicle across the screen safely, without colliding with any of the other vehicles. They were told that the other vehicles moved only in relation to their own; that is, they were not playing against any other people. All of the obstacles which they might encounter were automated. Here, participants played the role of a driver of a manually controlled car, but ran the risk of being involved in a collision with a vehicle which was *not* controlled by another human. It was hypothesized that driver confidence in their own abilities would lead to placing higher fault on the other, non-human controlled vehicles, and lower fault to themselves. The specific hypotheses for the overall study are shown below.

Hypotheses

H1: When collisions are perceived to be at medium or low avoidability levels, participants will attribute greater fault to manually-operated vehicles than to autonomous vehicles involved in identical collisions.

Rationale: The reasoning for this hypothesis is founded on the Just-World Fallacy, which states that if something bad happens to someone, they must have done something to deserve it. Additionally, the idea of defensive attribution, where people assign blame because they want to

believe that an event is preventable, also supports this hypothesis. Both of these biases center around the idea of events with negative outcomes being controllable, and their results being deserved. These biases might lead participants to believe that a person who was currently in control of their vehicle is more at fault than someone in an identical situation who was not in control at the time of the collision.

H2: Participants will attribute greater fault to a self-driving vehicle than to a human-driven vehicle when they believe that the accident was highly avoidable.

Rationale: This hypothesis states that the previous hypothesized relationship will be moderated by perceived avoidability. In support of this hypothesis is the defensive attribution theory, which again focuses on the belief that events are controllable and preventable. Here, blame falls on whoever fails to control or prevent these outcomes. Also in support is the culpable control model, which states that people take motivations into account when apportioning out blame. Participants can likely judge the extent to which the actions of each vehicle led to the outcome, but with humans the additional "mind to behavior" link exists which, in most cases, will help an individual be exonerated of some blame as it is unlikely that they were motivated to cause a collision. With automated vehicles, only the behavior to consequence link exists, which leaves less room for doubt. Also in support of this hypothesis is the theory of social conduct which states that individuals differentiate between bad outcomes caused by one's lack of effort, and those caused by their lack of ability. This is similar to the Van der Woerdt and Haselager (2017) findings where people judged a robot as being more responsible when the robot was meant to put away a toy and it demonstrated a lack of effort by throwing the toy, compared to

judging it less responsible when it demonstrated a lack of ability by being unable to pick up the toy. In both of these latter cases, the findings showed that people find more fault if one is capable of doing the right thing, but does not do so. An avoidable collision will appear to be a larger failure of effort, not of ability, and here people might find the automated vehicle at fault moreso than the human driver.

H3: Participants higher in confidence in their own driving abilities will attribute greater fault towards human drivers – of both autonomous and regular vehicles - than those with low confidence in their driving abilities.

Rationale: Part of the rationale for this is the just-world hypothesis, which states that people deserve what happens to them. There is no way for algorithms to deserve consequences. Secondly, the above average effect also supports this hypothesis. If people believe themselves to be safer drivers than another person, then it stands to reason that they might think that if they themselves were in the identical situation, they could have avoided the collision.

H4: Participants higher in confidence in their own driving abilities will attribute greater fault to a self-driving vehicle, compared to people with low confidence in their own driving abilities.

Rationale: This is hypothesis is supported by the above average effect. If someone believes themselves to be a better driver than most, but is still involved in a collision, then they most likely believe that almost any driver in that situation would experience the same outcome. Thus, it was not avoidable by human actions and therefore the one to blame would have to be the vehicle. Additionally, the causal attribution theory, where people tend to believe that actions are

driven by internal or external factors, supports this hypothesis. People tend to judge another by internal factors, but judge themselves by external factors such as luck. They are more likely to assign external factors to their own failures, and here the vehicle algorithms are one case of an external factor they can blame.

H5: Participants with low confidence in their own driving abilities will attribute greater fault to themselves than people with high confidence in their own abilities, even when the collision was unavoidable.

Rationale: See rationale for hypothesis four.

H6: The higher a participant rates themselves in their driving abilities compared to their peers, the higher their score will be on the Personal Fable Scale (Lapseley et al., 1989).

Rationale: The more that one believes themselves to be special, and unique from their peers, the more likely it is that these same beliefs about themselves will carry over into their driving abilities.

H7: Those with higher automation complacency will have a higher blame for the driver of an automated vehicle than for that vehicle's algorithms.

Rationale: If someone trust technology, they are more likely to believe that the technology is a safe alternative to human users, and will mitigate human error. Therefore, someone who finds automation in general to be trustworthy will place more blame on the human user than on the automation itself.

CHAPTER 2: EXPERIMENT ONE

Experiment 1 was designed to evaluate Hypotheses 1, 2, 3, 6, and 7. Participants viewed several scenarios of cars involved in collisions and were asked to rate to what extent each actor in the scenario was at fault for the events presented. Scenarios all involved minor collisions, but other than that were dissimilar with some showing collisions while moving (such as lane-changing) and some showing collisions while one vehicle was stopped (such as a rear-ending at a traffic signal). In each case there was one car (Car 1) which was of interest. Half of the participants in each scenario were told that Car 1 was an autonomous vehicle, and that its driver was responsible for monitoring the situation and taking over when they deemed it necessary (i.e., levels 2 or 3 automation in the SAE hierarchy). In these cases, the drivers did not re-establish control. Participants judged the fault of the car in the collision (including its driver and programming, separately, in the autonomous condition). Each participant's belief about their own driving skill, as well as their opinion on the extent to which the collision was avoidable, acted as covariates.

Method

Participants:

Two hundred and sixty-eight participants were drawn from the University of Central Florida (UCF) undergraduate student body. Two were rejected for failing to answer key questions, so a total of 266 responses were analyzed. Of these, 158 were females and 108 males. The average age was 20.79 years (*sd* = 5.70). They were rewarded with SONA credit for their participation. According to statistical power analysis program G*Power (Faul, Erdfelder, Lang & Buchner,

2007), for a multiple regression with seven predictors to reach a power level of .95, with alpha of .05 and an expected effect size of $f^2 = .15$ (which is a medium effect size) a total sample size of 153 participants was required. However due to the online format, the number of participants was much larger to account for potential lack of engagement during online studies. All participants were required to possess a valid driver's license and had to be over the age of 18. Additionally, since the experiment involves situations including car collisions, participants were excluded from the survey if they did not wish to view the driving scenarios or believed that doing so may cause them any degree of distress.

Design

The experiment used five predictor variables. The type of car was a categorical variable which possessed two levels, i.e., whether participants were told Car 1 was a regular car, or a self-driving vehicle. Participants were randomly assigned to one of the two conditions in each driving scenario. The other variables were continuous variables and were determined by the participant's response to specific questions. One continuous ratio variable was the avoidability of the collision, which was determined by the participant for each driving scenario. Another continuous variable was the participant's confidence in their own driving skills. This was measured once at the beginning of the experiment. Additionally, participant's responses to a scale measuring automation complacency, and their response to a scale measuring their belief in themselves as being special or unique, were analyzed as predictors.

Materials

Surveys and Scales:

Driving Experience Survey: Participants were asked several questions about their experience driving, such as how long they had a valid driver's license, how often they drove in a week, and how far they tended to drive in a day, as well as whether they had experience with minor car collisions or self-driving vehicles.

Driving Confidence Survey: This survey measured a participant's confidence in their driving skills and was adapted from Matthews and Moran (1986). Participants were asked to rate their vehicle handling skills, defined as their control over the car; their driving judgement, or their ability to make good decisions while driving, and their driving reflexes, or reaction speed. They rated these skills, as compared to their peers, on a scale of 1 (much worse than my peers) to 9 (much better than my peers). They were also asked about their overall belief regarding their driving skills on a 9-point Likert scale (1 being very poor and 9 being excellent; and see Figure 2). Their self-rated score on the four questions were averaged in order to determine one numerical value for their driving confidence. Here a score of 5 was average, values less than 5 were below average, and anything greater than 5 was above average.

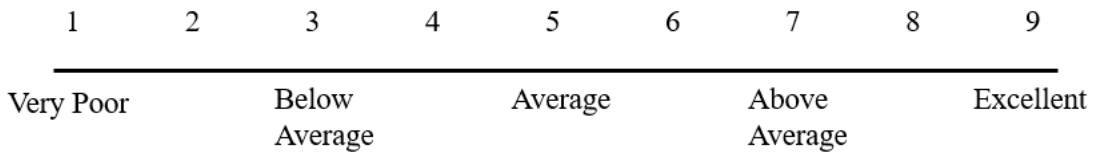


Figure 2: Likert scale for rating driving abilities (Matthews & Moran, 1986).

Personal Fable Scale: Participant evaluated themselves using Lapsley et al.'s (1989) New Personal Fable Scale. This scale measured participants' beliefs about their own omnipotence,

invulnerability, and uniqueness. Participants were asked to rate their agreement with statements such as “I believe that I am unique” and “no one sees the world the way that I do.” Five-point Likert scales were used, ranging from 1(Strongly Disagree) to 5 (Strongly Agree).

Automation Complacency: Since one’s pre-conceived notions about automation may influence how they attribute blame to vehicles they believe to be automated, a scale was used to assess participants’ complacency towards automation (Singh, Molloy, & Parasuraman, 1993). The scale measures confidence, trust, perceived safety, and reliance on automation by asking participants to agree or disagree with statements such as “Automated devices in medicine save time and money in the diagnosis and treatment of disease” and “ I would rather purchase an item using a computer than deal with a sales representative on the phone because my order is more likely to be correct using the computer.” Five-point Likert scales were used, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Avoidability Survey: The perceived avoidability of each incident was determined by the participant in every driving scenario, with regards to each separate car involved. They were asked how avoidable the collision was, if each car had aken different actions. This variable was measured on a sliding scale of 1-5 (one end, 1, being impossible to avoid, and the other, 5, labeled as being very easy to avoid), but participants could not see the numerical value and only the textual descriptor. In a case where two cars were involved in the collision, the participants would answer the question twice, once referring to the avoidability had Car 1 taken different actions, and once referring to avoidability in the case that Car 2 had behaved differently. The

actual objective avoidability of the collision (if such a thing exists) was not assessed, only the participant's perceptions thereof.

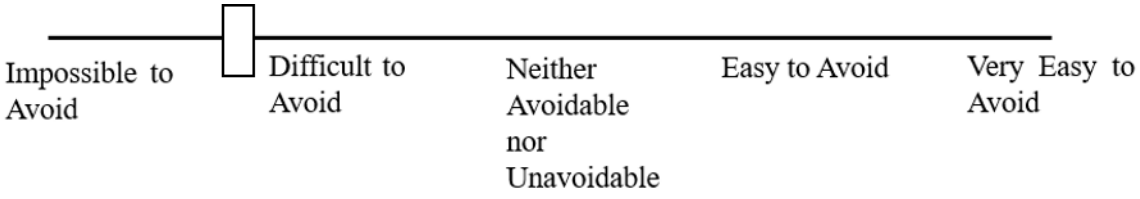


Figure 3: Likert-style question measuring perceived avoidability of the accident.

Fault Surveys: Participants were asked to what extent each individual in the collision scenario was at fault. This included any pedestrians involved, any other cars involved, and Car 1. In the condition where participants were told that Car 1 was an autonomous vehicle, they were be asked to what extent the driver was at fault, and to what extent the car's self-driving algorithm was at fault. They gave their answer on a sliding scale for each actor in the scenario, as seen in Figure 4. Scales ranged from 1, not at all at fault, to 5, fully at fault. However, participants could see only the descriptors and not the numerical value associated with each location on the sliding scale. For both fault and avoidability, participants responded to questions about all vehicles involved. However, the results that were of interest were only those relating to the vehicle that was called autonomous in certain conditions. Responses to questions relating to the other vehicles, were simply used as distractors so that participants could not as easily guess the purpose of the study.

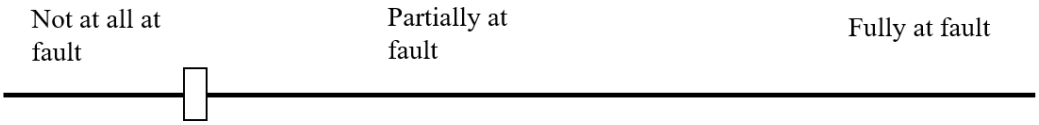


Figure 4: Sliding scale assessing perceived culpability of Car 1.

Driving Scenarios:

Driving scenarios were selected via a pilot study, in which 28 people viewed multiple videos of vehicular collisions, and rated the avoidability of each on a scale ranging from 0 (not avoidable at all) to 100 (very easily avoidable). An average rating below 30 was considered to be a very difficult collision to avoid, a rating from 30- 69 was considered to be of medium avoidability, and any rating of 70 or higher was considered to be an easily avoidable collision. Two scenarios of each avoidability level were selected for inclusion in the study. The six selected scenarios came from either dashboard cameras which show a view of the road from a windshield, as a driver would see; or from traffic cameras which show the road from a bird's eye view. Additionally, participants were shown diagrams which included additional information not always visible in the video, such as construction cones, stop signs, road structure, pedestrians,

and other vehicles, and included any turn signals or pertinent facts. See Figure 5. Each participant viewed all of six different driving scenarios which were presented in random order.

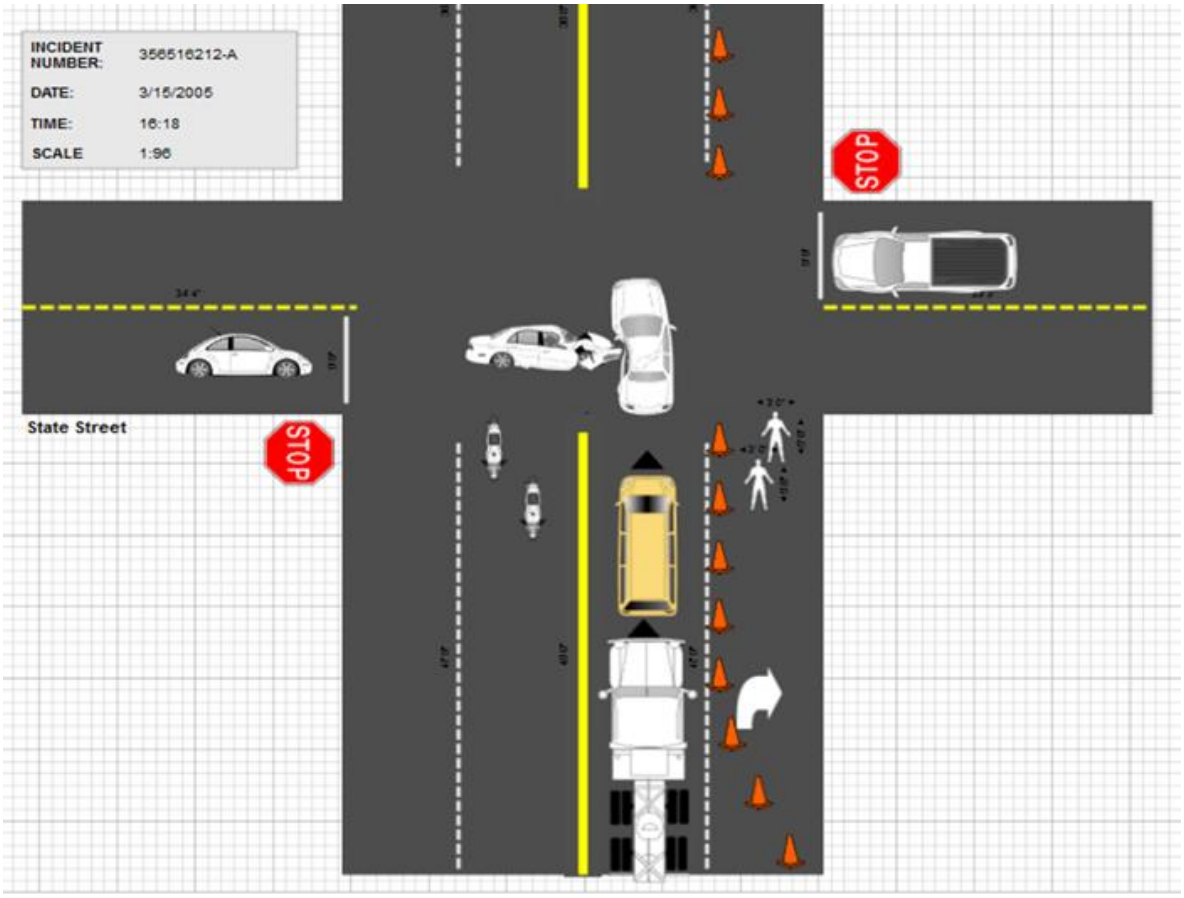


Figure 5: A diagram showing various factors of a car collision including other vehicles, pedestrians, signage, and intersection structure

Procedure

The experiment took place via a computer screen using the Qualtrics survey engine. Participants were first required to list their age and gender and confirm that they possessed a

valid driver's license. They then completed the driving experience, driving confidence, personal fable, and automation complacency surveys. Following the completion of the surveys, they were informed that they were about to witness six different driving scenarios, all ending in car collisions. Participants were also informed that some of the driving scenarios would feature collisions involving self-driving vehicles. They were given information about level two automation, in which a human driver must monitor the vehicle and take over as necessary. Following this, participants were presented with one of the six driving scenarios, and randomly assigned to the group which was told that the car in question was self-driving, or the group which was told that all cars in the scenario were regular manually-operated vehicles. After viewing the scenario, participants rated the avoidability of the collision, and completed the sliding scales attributing fault. There was no maximum or minimum value set for overall fault: participants could say that no one was to fault at all, or that all parties were highly at fault. Then, they proceeded to the next scenario, which repeated the process.

Results

Descriptive Statistics

Fault Attribution: The participants each rated fault in the six different scenarios. In each scenario, they were randomly and evenly split between the condition where they were told that the car in question was autonomous, or the condition where they were told that the car was a regular vehicle. Since most of the participants viewed 6 scenarios each, they collectively viewed 791 scenes in which they were told the vehicles were autonomous, and 799 scenes in which they were told that all vehicles involved were regular cars. This difference was due to some participants failing to answer questions. Overall, the average rating of fault was 3.60, with a

standard deviation of 1.59. For autonomous vehicles, the average rating was 3.67 (1.57) and for regular vehicles the average rating was 3.54 (1.61). The difference between the two groups was not statistically significant $t(1588) = -1.70$, $p = .09$, but it was a level that encouraged further investigation.

Avoidability: When participants rated the avoidability of each collision, based on the actions taken by each vehicle, they gave an average rating of 3.86 (1.49). For autonomous vehicles, the average rating of avoidability was 3.88 (1.46) and for regular vehicles, the average avoidability rating was 3.84 (1.52). The difference was not significant $t(1588) = -0.57$, $p = .57$.

Driving Abilities Overall: Participants were asked to rate their driving abilities. The average rating for a participant's vehicle handling skills was 7.57 (1.26). Their driving judgement was scored at 7.52 (1.46) and their driving reflexes were rated 7.55 (1.52). Overall, their average driving skills were perceived to be 7.55 (1.26). As this was a scale of 1-9, participants in general perceived themselves to be good drivers.

Driving Abilities Compared to Peers: Participants were asked to rate their driving skills compared to their peers. Their vehicle handling skills were rated at 7.05 (1.56). Their driving judgement was 7.06 (1.68), and their driving reflexes were 7.07 (1.67). Overall, compared to their peers, participants rated their driving abilities at 7.06 (1.52), again considering themselves better than the average driver in their age group.

Personal Fable: The Personal Fable Scale measured participants' beliefs about their own omnipotence, invulnerability, and uniqueness on a scale of 1-5. For omnipotence, participants

rated themselves an average of 3.03 (0.51). On the invulnerability subscale, participants rated themselves at a 2.89 (0.50), and on the uniqueness subscale the average score was 3.40 (0.49).

Automation Complacency: Automation complacency was measured on the subscales of confidence in automation, reliance on automation, trust in automation, and belief in the safety of automation. On the confidence subscale, participants scored an average 3.59 (1.06). For reliance, the average score was 3.32 (0.84) and for trust the average was 3.18 (0.77). For belief in the safety of automation, the average score was 3.15 (1.01). Overall, the average score for automation complacency was 3.31 (0.92).

Overall Fault Attribution Model

In order to determine the overall fault attribution model, a hierarchical regression was conducted. This analytic style was based somewhat on Wohleber and Matthews (2016). In the first step, automation condition (Auto) was considered. This variable measured whether the participant was told that the vehicle in question was autonomous or not. It was dummy coded with a score of 0 meaning a regular vehicle and 1 indicating an autonomous vehicle. This was the first step as it was the primary focus of the experiment, and was an integral part of the most hypotheses. However, in this step, the value of R^2 was only .002. In the second step, perceived avoidability of the collision (Avoidability) was entered. This was the logical next step as the extent to which a collision is avoidable directly impacts whether or not any fault can be attributed at all, and the scenarios were chosen specifically to have a range of potential perceived avoidability scores. This produced an R^2 value of .584, accounting for around 58% of the variance in fault attributed to the vehicle of interest. For the third step, driver confidence was entered as well. This included

both driver confidence in general (Driving-General), and driver confidence compared to their peers (Driving-Peers). This was done because driver confidence, in terms of the Above Average Effect, has been examined previously in driving literature (Wohleber & Matthews, 2016) and because it was hypothesized to have a relationship with fault attribution. This produced a R^2 value of .585, only slightly increasing R^2. In the last step, gender (dummy coded with 0=males and 1=females), scores on the Personal Fable Scale (PF) and Automation Complacency (Comp) were added but had no effect. See Table 2. A regression table including all main effects and two-way interactions is included in Appendix D.

Table 2: *The regression model predicting fault attribution*

	Step 1			**Step 2**			**Step 3**			**Step 4**		
Variable	b	se	B	b	se	B	b	se	B	b	se	B
Auto	.130	.080	.041	.234	.144	.074	.231	.144	.073	.229	.144	.072
Avoidability				.834	.024	.779*	.834	.024	.779*	.834	.024	.780*
Auto*Avoidability				-.036	.035	-.050	-.036	.035	-.049	-.035	.035	-.048
Driving-Peers							-.044	.023	-.042^{+}	-.036	.024	-.034
Driving-General							.013	.028	.010	.015	.028	.012
PF										-.032	.024	-.023
Comp										-.001	.009	-.002
Gender										-.031	.054	-.010
R^2	.002			.584			.585			.585		
R^2 change	.002			.582			.001			.000		

*indicates significance at $p < .05$
$^{+}$indicates significance at $p < .10$

Avoidability was the single largest predictor of fault attribution. The more avoidable a participant perceived a collision to be, the higher their fault attribution to the vehicle involved. Automation condition alone was not significant, with slightly higher fault attribution when vehicles were autonomous. This is the opposite of what was expected in Hypothesis 1, which

stated that participants would attribute higher fault to drivers of regular vehicles than to those of autonomous vehicles. The interaction between automation condition and avoidability was non-significant. This is not what was predicted by Hypothesis 2, which predicted that higher avoidability would lead to higher fault in autonomous vehicles than in regular vehicles. Thus, the results failed to support Hypotheses 1 and 2. Driver confidence compared to one's peers was a marginally significant predictor of fault attribution, but was actually a negative predictor and thus Hypothesis 3 was not supported. Driver confidence in general did not predict a significant amount of variance in fault attribution after the model had taken into account the effects of driver confidence compared to one's peers. Additionally, there was multicollinearity between those two measures of driver confidence.

Correlations

Correlations between the variables were examined in order to find support for Hypotheses 6 and 7. Table 3 shows the correlations between overall variables of fault attributed to the vehicle in question (Fault), perceived avoidability of the collision (Avoidability), whether or not the vehicle was automated (Autonomous), fault in the algorithms (Algorithms Fault), confidence in driving abilities compared to peers (Driving Peers), confidence in driving abilities in general (Driving General), score on the Automation Complacency Scale (Complacency), and score on the Personal Fable (Personal Fable). Whether or not the vehicle was autonomous was dummy coded with 0 indicating that the vehicle was a regular, manually operated car, and 1 indicating that it was autonomous. Gender was dummy-coded, with 1 indicating a female participant and 0 indicating a male participant. Correlations that show the relationship between all of the subscales are found in Appendix A.

***Table 3*:** Correlations between the overall variables in Study 1.

	1	2	3	4	5	6	7	8
1. Fault	1							
2. Avoidability	.763*	1						
3. Autonomous	.043	.014	1					
4. Algorithms Fault	-.134**	-.210**	---	1				
5. Driving Peers	-.017	.029	-.002	-.101**	1			
6. Driving General	.024	.056*	.014	-.127**	.665**	1		
7. Complacency	.023	.032	.010	-.059	.052*	.099**	1	
8. Personal Fable	-.001	.042	-.017	-.074	.366**	.286**	.114*	1
9. Gender	-.023	-.026	.030	-.101**	.008	.015	-.117**	-.218**

*indicates significance at $p < .05$ level, **indicates significance at the $p < .01$ level, a dash indicates a correlation that could not be determined because one of the variables is constant.

Hypothesis 6 stated that participants who scored highly on the Personal Fable Scale (Lapsley et al., 1989) would also have a higher rating of their own driving skills compared to their peers. This was supported by a positive correlation ($r = .366$). While the correlation is not particularly large, it is statistically significant when taking into account that the majority of participants believed themselves to be much better drivers than their peers (with a mean score of 7.06 out of 9), and shows that the scores on the Personal Fable Scale were related to participant's confidence in their driving abilities compared to their peers. Additionally, participants who scored highly on Personal Fable also had confidence in their driving skills in general ($r = .286$).

Hypothesis 7 stated that participants who scored highly on the Automation Complacency Scale would be more likely to fault the human driver of an automated vehicle than to fault the algorithms. While the correlation between fault in algorithms and complacency with automation was in the expected, negative direction, it was not statistically significant (r = -.059, p > .05). Therefore, Hypothesis 7 was not supported by the data.

The attribution of fault was positively and strongly correlated with perceived avoidability (r = .763). The more avoidable a collision appeared, the more fault was attributed to the car involved. Fault in algorithms was negatively related to perceived avoidability, indicating that participants blamed the human driver more than the algorithms, for an avoidable collision (r = -.210). Fault in algorithms was also negatively related to fault in general (r = -.134), again suggesting that participants placed more blame on the human driver. A participant's confidence in their driving abilities compared to their peers, and their driving skills in general, were both negatively correlated with their fault attributed to the algorithms (r = -.101 and r = -.127). This indicated that those with high confidence in their driving abilities placed more blame on the human driver than on the algorithms.

There were small but significant correlations between automation complacency and driver confidence both compared to their peers, and in general (r = .052, and r = .099). There was also a relationship between scores on the Personal Fable Scale and one's complacency with automation (r = .114). Those with higher automation complacency also had higher scores on Personal Fable. Women overall scored lower on the Personal Fable (r = -.218). They also had

lower automation complacency scores (r = -.117) but, paradoxically, attributed less fault to the algorithms of an automated vehicle (r = -.101).

Fault of Autonomous Vehicles

The regression model for fault of the drivers of autonomous vehicles focused only on data from those participants who were told that Car 1 in the given driving scenario was autonomous. Again, a hierarchical regression was performed. In this case, automation condition was not entered as a variable as all vehicles in this sub-analysis were the ones that participants had been told were autonomous. The variables were entered in the same order as the previous analysis. In the first step, perceived avoidability accounted for an R^2 of .550. In the next step, both types of driver confidence only accounted for a change in R^2 of .006. Again, gender, scores on the Personal Fable, and scores on the Automation Complacency scales had little effect on R^2, bringing it to .557. See Table 4.

Table 4: Regression model predicting fault attribution in only the algorithms of autonomous vehicles

	Step 1			**Step 2**			**Step 3**		
Variable	b	se	B	b	se	B	b	se	B
Avoidability	.797	.026	.742*	.798	.026	.742*	.798	.026	.742*
Driving-Peers				-.102	.034	-.098*	-.105	.035	-.101*
Driving-General				.054	.041	.043	.057	.041	.046
PF							.007	.036	.006
Comp							-.015	.013	-.028
Gender							.004	.080	.001
R^2	.550			.556			.557		
R^2 change	.550			.006			.001		

*indicates significance at $p < .05$
$^{+}$indicates significance at $p < .10$

Overall, the same factors that influenced fault attribution overall had effects of a similar magnitude, and in the same direction, as those exerting influence on fault attribution in autonomous vehicles. Again, the more avoidable a collision seemed to be, the higher the fault attributed. Additionally, higher driver confidence compared to one's peers actually contributed to *lower* attributions of fault. It is possible that one's confidence in their own driving abilities contributed to their belief that other drivers were not as competent, and thus at less fault.

Fault of Algorithms

In cases where the participants were told that the vehicle of interest was autonomous, they were given the opportunity to divide fault between both the driver, and the algorithm, which shared control of the vehicle. This was evaluated on a bipolar sliding scale, so the lower the fault is to the algorithm, the higher it is to the human driver and vice versa. Variables were entered in the same order. In the first step, avoidability had the largest influence on fault attribution with R^2 of .044, which only accounted for 4.4% of the variance in fault attribution. In the next step, driver confidence increased R^2 to .058, and finally in the last step, the scores on both the Personal Fable Scale and the scale measuring automation complacency, increased R^2 to .071. Overall, the model did not predict much of the variation in fault attributed to the algorithms specifically.

Table 5: Fault attribution of algorithms

	Step 1			Step 2			Step 3		
Variable	b	Se	B	b	se	B	b	se	B
Avoidability	-.219	.039	-.210*	-.214	.039	-.205*	-.211	.039	-.202*
Driving-Peers				-.041	.046	-.045	-.023	.047	-.025
Driving-General				-.095	.057	-.083^{+}	-.086	.057	-.075
PF							-.081	.048	-.070^{+}
Comp							-.025	.018	-.052
Gender							-.294	.107	-.107*
R^2	.044			.058			.071		
R^2 change	.044			.014			.003		

*indicates significance at $p < .05$
$^{+}$indicates significance at $p < .10$

The results here showed that avoidability negatively predicted the fault attributed to the algorithms. That is, the more avoidable a collision seemed, the *less* fault was placed on the algorithm and the more fault was placed on the human driver. Driving confidence was also negatively related to fault in the algorithms. The more confidence a driver had in their own ability, the more fault they placed on the human and the less on the algorithm, which partially supports Hypothesis 3. Gender was also a significant predictor of fault attribution, with males more likely to attribute fault to the algorithms, and females more likely to attribute fault to the human drivers.

Gender Differences

Some of the variables were compared between genders. The fault attributed to the vehicle was similar, regardless of gender, with males giving an average rating of 3.65 (1.53) and females rating it at 3.57 (1.63). There was no significant difference in a *t*-test ($p > .1$). Similarly, there was no significant difference between scores on the driving confidence scales, neither the general

scale (males = 7.57 (1.32) and females = 7.53 (1.20)), nor the scale measuring confidence compared to their peers (males = 7.05 (1.60) and females = 7.07 (1.47)). However, there was a significant difference in the fault attributed to algorithms only, with males more likely to blame the algorithm (m = 2.44, sd = 1.85) than females (m = 2.21, sd = 1.06). This difference was significant ($p < .05$).

Discussion of Experiment One

Experiment One showed that perceived avoidability of a collision had the largest impact on fault attribution both in collisions involving autonomous vehicles, and overall. The second largest predictor of fault attribution was a participant's confidence in their own driving abilities compared to their peers. Whether or not a vehicle was autonomous did not have a large effect on the factors that affected fault attribution. Overall, the two primary predictors of fault attribution accounted for approximately 50% of the variance.

There are several reasons the vehicle's status as either a regular or autonomous car may not have been a significant predictor of fault. One reason is that the participants were largely unfamiliar with autonomous vehicles and their capabilities. In a pre-survey regarding their experience with autonomous vehicles, many reported no such prior experience. A few had been passengers in a Tesla, but none of the present sample owned an autonomous vehicle or had any extensive experience using one. The majority who did claim some knowledge of autonomous vehicles largely cited experience with lane assist. Although autonomous vehicles were described in the study, it is possible that some participants had pre-conceived notions of autonomous vehicles as simply regular vehicles with higher-quality automated technologies. Additionally, all

of the collisions illustrated were minor. This was done intentionally to prevent any unnecessary trauma to the participants. However, the minor collisions may not have warranted in-depth examination of who is to blame in the same way that a deadly accident may have done.

Based on both the similarities between the overall fault attribution model and the fault model for autonomous vehicles, and the non-significant effect of automation condition on fault attribution, it appears that people have similar attributions for both autonomous and regular vehicles. This may indicate that there will be little or no change in the way fault is attributed, in both legal and moral settings, even as the world begins to shift in the direction of autonomous vehicles. However, the present findings may only be relevant in this current, brief moment in time where participants are familiar with the concept of autonomous vehicles but do not yet have any extensive prior experience with them.

There was also a significant, positive relationship between a participant's confidence in their driving skills compared to their peers, and their score on the Personal Fable scale. However, there was a negative, but non-significant relationship between their trust in automation and fault attributed to specifically the algorithms of an autonomous vehicle. These findings supported Hypothesis 6 and partially supported Hypothesis 7. It is possible that one's complacency with automation in general, does not extend to the high-risk and new technology of autonomous vehicles.

CHAPTER 3: EXPERIMENT TWO

Having examined the factors that affect fault attribution when participants are witnesses to collisions involving autonomous vehicles, the next step was to examine whether those findings remained constant when participants played the role of the driver. The second experiment placed participants in the position of the operator/driver of an autonomous vehicle, and addressed Hypotheses 4 and 5. Participants completed a task that simulated using an automated decision aid in driving. In this task, they were faced with different driving situations and had to make the correct decisions with the help of a Safety Suggestion Algorithm that mimicked what some automated decision aids are capable of. Participants could choose to agree or disagree with the algorithm at each point. Again, participant confidence in their driving abilities, and their opinion about the avoidability of the collision, were examined as predictors of their fault in both themselves, and the automation.

Method

Participants

Participants were undergraduates at the University of Central Florida, and were given SONA credits in exchange for their participation. All were over the age of 18 and could not have any objection to being involved in simulated car collisions. A total of 188 participants were involved in the study. Of these, 123 were female, 63 were male, and 3 declined to state a gender. Their average age was 20.05 years with a standard deviation of 4.35 years. Again, statistical power analysis program G*Power (Faul, Erdfelder, Lang & Buchner, 2007) determined that for a multiple regression with nine predictors to reach a power level of .95, with alpha of .05 and an

expected effect size of f^2= .15, a total sample size of 166 participants was required. This number was again inflated due to the online nature of the experiment. However, since engagement proved to be relatively high in the first experiment, *n* was not as drastically increased as in the prior study.

Materials

Surveys and Scales:

The same scales were given prior to this task, as were given in Experiment One. The Driver Confidence and Driver Experience Scales, as well as the Personal Fable and Automation Complacency scales were the same as the previous study. A similar fault survey was used, however in Experiment Two, rather than attributing fault to the various players in the accident scenario, the participant attributed fault to a) the vehicle that they were operating and b) themselves, and was measured on a scale of 1-10 (with 1 indicating not at all at fault, and 10 indicating highly at fault). This was done to allow for greater variation in responses, and to give participants an option to not place any blame at all on themselves or the SSA. The avoidability scale was also different here, asking participants how easily the incident could have been avoided on a scale of 1 (not easily at all) to 10 (very easily).

The Scenarios:

Scenarios consisted of driving situations which a driver might experience in their commute. For example, coming across an obstacle in the road and having to decide between severe braking or swerving around the obstacle (and see Hancock & de Ridder, 2003). In each

case, participants were given two options (such as swerving, or using the brakes). The Safety Suggestion Algorithm indicated which option it recommended, and participants were free to choose to agree or disagree with it. They were given the option to choose between agreeing or disagreeing with the suggestions from the SSA, rather than given the choice between the two driving options, to ensure that they attended to the SSA's suggestion. This task measured the cognitive aspects of driving, such as decision making and knowing the rules of the road. However, it did not examine the physical aspects of driving such as reaction time.

The Images:

Scenarios were accompanied by images and videos which helped to show the situation more clearly. In every case, the participant's car was yellow, and the other vehicles in the scene were white. Participants were informed that theirs was the yellow vehicle, and told to use the images to help guide them in their decision making if they were confused by the textual description of the scenario.

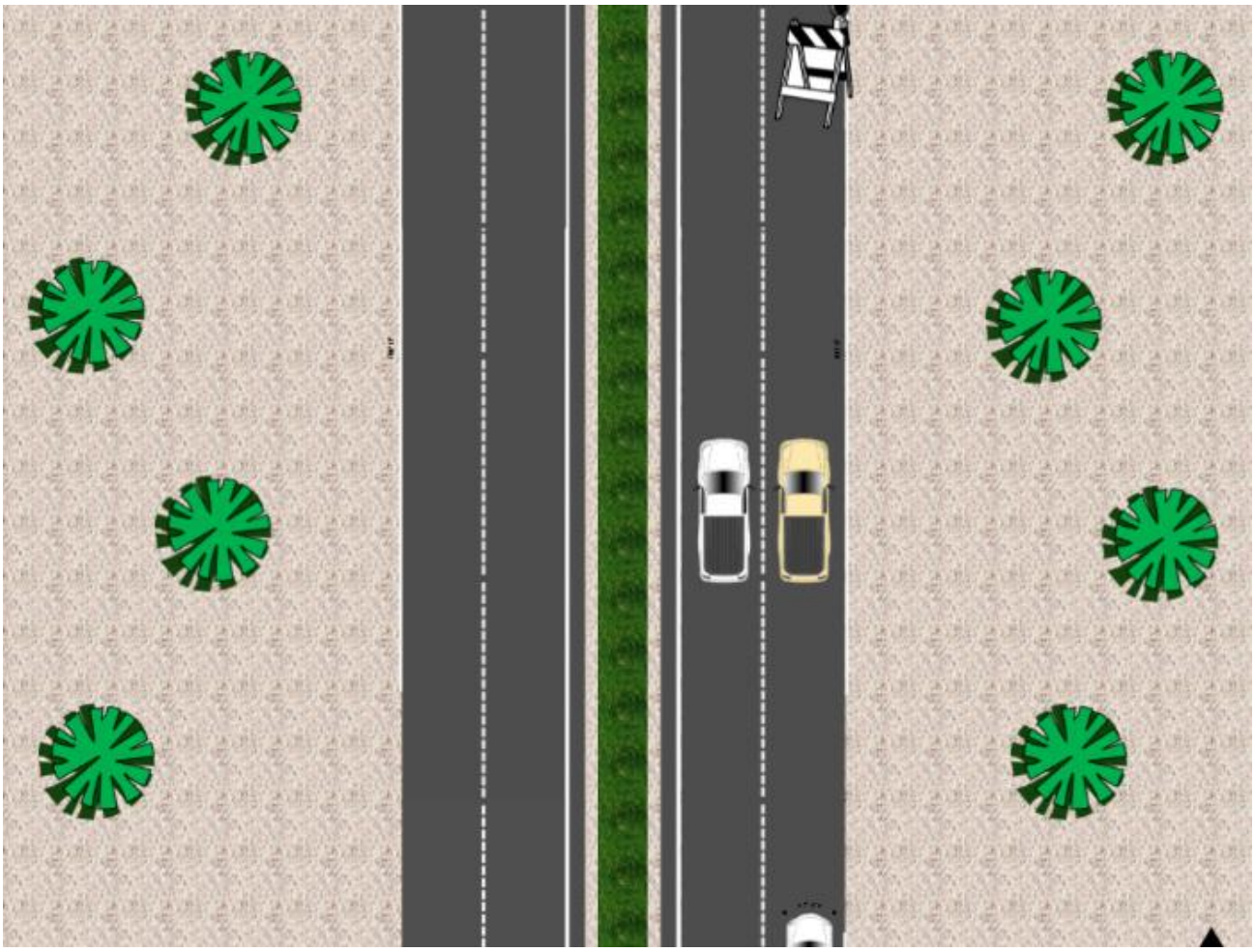


Figure 6: A diagram of a driving situation in which a yellow vehicle (the participant's vehicle) drives in the far right lane alongside a white vehicle in the left lane. Ahead of the two vehicles, the right lane is blocked.

Procedure

Participants read the informed consent, and agreed to take place in the experiment. Then, they filled out the driver experience, driver confidence, personal fable and automation complacency surveys which were hosted via Qualtrics. Once those surveys were complete, they began the task. In the task, participants were given up to nine driving scenarios. The scenarios all occurred in the same order, and participants advanced to the next scenario by choosing the

correct course of action. In each scenario, a driving safety issue came up and the participants were given two options. The Safety Suggestion Algorithm would make a recommendation, and participants could choose to either agree, or disagree with the algorithm. In most cases, the algorithm was correct and agreeing with it was the appropriate course of action. In the third, sixth, and eighth scenarios, the algorithm was incorrect and disagreeing with it was the way to advance to the next step. When a participant made an incorrect choice, they were informed that their drive had ended and they were given the collision avoidability scale and the fault scales to fill out regarding the extent to which they faulted themselves, and the Safety Suggestion Algorithm, for any failure which occurred. They were also given scales to measure perceived avoidability of the failure, on both their part and the part of the algorithm. Those who successfully navigated the driving task were not given these additional surveys.

Results

The Driving Task

Participants, overall, did not complete the driving task successfully. Of the 188 total participants, only 5 got through all scenarios successfully. Most were thus incorrect in one of the preceding scenarios. Figure 7 shows the breakdown of which rounds were most difficult for the participants. While some were easy with very few failures, some rounds were more difficult. In 97 cases, the participants agreed with the SSA's incorrect advice, and failed at the task. In 86 cases, the SSA was correct, but the participants disagreed with its suggestion and failed.

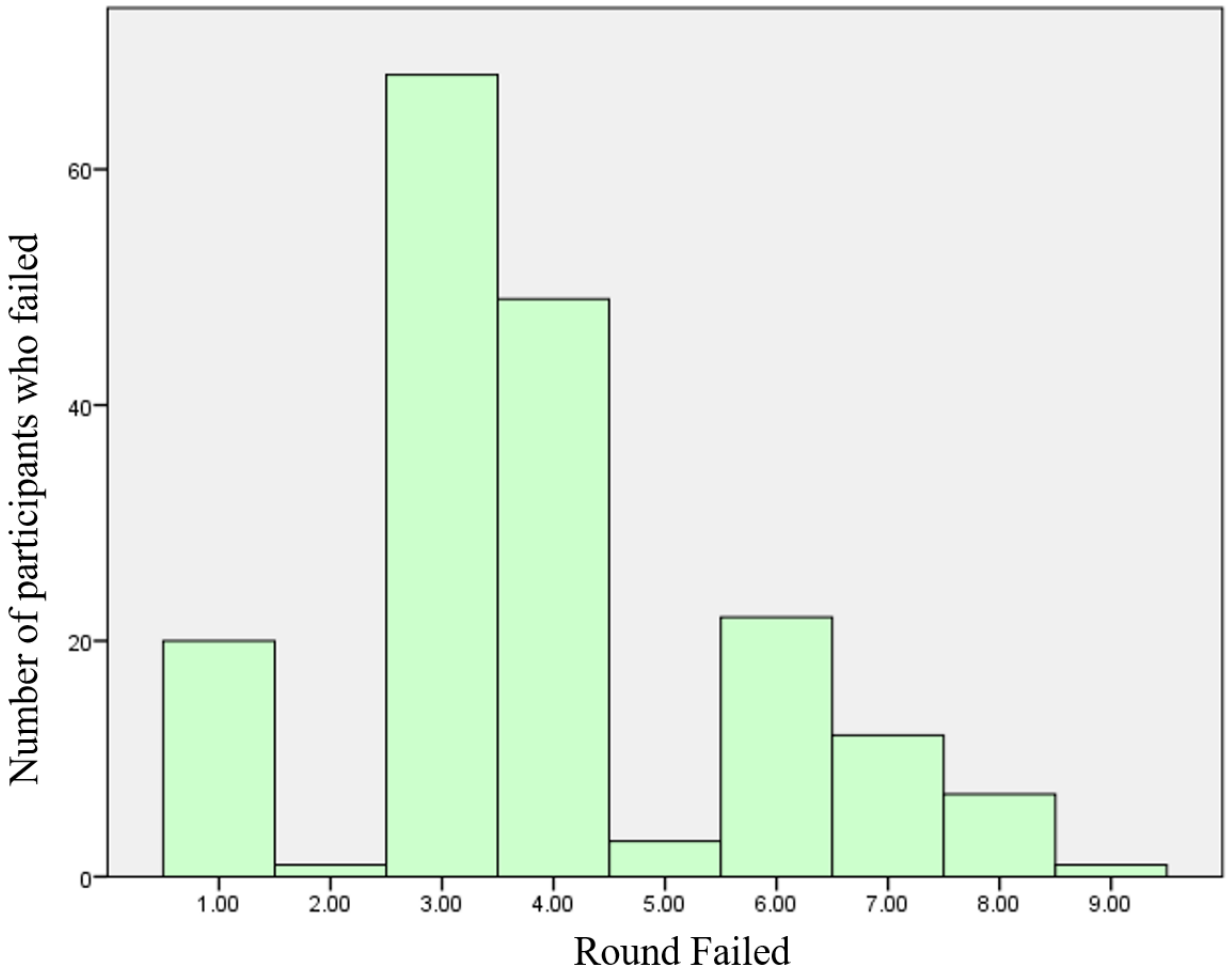


Figure 7: Graphical representation of the rounds in which participants failed, and the number who failed in each round.

Of the 5 participants who successfully managed to navigate the task, three were males and two were females. These participants did not attribute fault or avoidability, as there was no failure in regards to which they needed to attribute fault.

Descriptive Statistics

Fault Attribution: Those 183 participants who did not successfully complete the task attributed fault to both themselves, and the Safety Suggestion Algorithm. When asked to attribute fault to the SSA, participants gave an average fault attribution of 3.75 (2.57). Surprisingly, a higher fault was attributed to the SSA when the driver's mistake had been to disregard its advice (m = 4.07, sd = 2.61), compared to when the driver had followed incorrect advice (m = 3.66, sd = 2.45). However, the difference between these two groups was not statistically significant $t(181) = 1.095$, $p > .05$.

Participants tended to place higher fault on themselves than on the SSA. The average fault that the participant attributed to themself was 5.21 (3.32). Again, the difference between fault attribution when the algorithm was wrong (m = 5.67, sd = 3.21), was not significantly different from when the algorithm made the right choice and the participant disagreed (m = 4.99, sd = 3.26; $t(181) = -1.42$, $p > .05$). This indicates that, regardless of whether the vehicle or the participant made the incorrect choice, fault was attributed similarly.

Avoidability: Participants rated the avoidability of the incident which made them fail the task, by expressing the extent to which the incident was avoidable if different actions had been taken by themselves as the driver, and by the SSA. Overall, participants thought that the incident was more avoidable had different actions been taken by themselves as the driver (m = 7.04, sd = 2.77) compared to how easily it could have been avoided if the SSA had taken different actions (m = 6.47, sd = 2.85)

Driving Abilities Overall: Participants were asked to rate their driving abilities. The average rating for a participant's vehicle handling skills was 7.86 (1.35). Their driving judgement was scored at 7.70 (1.41) and their driving reflexes were rated 7.69 (1.38). Overall, their average driving skills were perceived to be 7.75 (1.26). These findings were similar to Study 1, in that participants overall believed themselves to be good drivers. Those 5 who successfully completed the driving task had similar perceptions regarding their driving ability, with an overall score of 7.27.

Driving Abilities Compared to Peers: Participants were asked to rate their driving skills compared to their peers. Their vehicle handling skills were rated at 7.23 (1.53). Their driving judgement was 7.32 (1.60), and their driving reflexes were 7.21 (1.64). Overall, compared to their peers, participants rated their driving abilities at 7.25 (1.47), again considering themselves better than the average driver in their age group. Interestingly, the five who successfully completed the driving task gave themselves a lower rating of 6.53.

Personal Fable: The Personal Fable Scale measured participants' beliefs about their own omnipotence, invulnerability, and uniqueness on a scale of 1-5. For omnipotence, participants rated themselves an average of 3.01 (0.47). On the invulnerability subscale, participants rated themselves at a 2.90 (0.58), and on the uniqueness subscale the average score was 3.34 (0.45). The average on the scale overall was 3.08 (0.33). Scores on all these subscales were very similar to those on the same scales as in Experiment 1.

Automation Complacency: Automation complacency was measured on the subscales of confidence in automation, reliance on automation, trust in automation, and belief in the safety of

automation. On the confidence subscale, participants scored an average 3.33 (1.02). For reliance, the average score was 3.12 (0.79) and for trust the average was 3.03 (0.68). For belief in the safety of automation, the average score was 2.79 (1.14). Overall, the average score for automation complacency was 3.07 (0.61).

Fault Attributed to the Vehicle

A hierarchical regression was conducted to determine the predictors of fault attributed to the vehicle—specifically the SSA. Based on the results of Experiment 1, here the first step included perceived avoidability due to actions of the SSA (SSA Avoidability) and the driver (Driver Avoidability). This accounted for 10.1% of the variance in fault attribution. In the second step, the round in which participants had failed was added as a variable. Additionally, the dummy-coded variable of whether the algorithm had made the wrong suggestion prior to the incident (SSA Wrong; dummy coded with 0 = the SSA was correct in the last round played, and 1 = the SSA was wrong), was included in this step. This was because the algorithm's suggestion being incorrect should have an effect on its perceived role in causing a collision. This increased predicted variance to 11 percent. In the third step, both types of driver confidence were entered, and in the last step, again, gender, scores from the Personal Fable, and scores from the Automation Complacency scales were included as predictors. See Table 6. Overall, this model predicted only 16.7% of the variance in fault attribution.

Table 6: Regression for fault attributed to the SSA.

	Step 1			Step 2			Step 3			Step 4		
Variable	b	se	B	b	se	B	b	se	B	b	se	B
Driver Avoidability	-.263	.066	-.296*	-.266	.067	-.299*	-.274	.067	-.308*	-.281	.066	-.381*
SSA Avoidability	.210	.068	.230*	.216	.068	.236*	.214	.068	.235*	.228	.068	.250*
SSA Wrong				-.434	.364	-.086	-.459	.364	-.091	-.261	.367	-.052
Round Failed				.071	.102	.050	.084	.102	.059	.102	.101	.072
Driving-Peers							.306	.168	.172^{+}	.265	.169	.155
Driving-General							-.290	.196	-.146	-.289	.193	-.146
PF										.837	.576	.109
Comp										-.061	.298	-.015
Gender										-.818	.400	-.152*
R^2	.101			.110			.127			.167		
R^2 change	.101			.009			.017			.040		

*indicates significance at $p < .05$
$^{+}$indicates significance at $p < .10$

As shown in the results of Study 1, perceived avoidability was the largest predictor of fault attribution. Here, the less the participant felt that they as the driver could have avoided the incident, the more they blamed the SSA for the undesirable outcome. The opposite was true when it came to perceived avoidability by actions of the SSA, where participants felt that the more easily the SSA could have avoided the incident, the more the SSA was at fault. Participants' confidence in their driving skills compared to their peers was significant in the third step, with those who had higher confidence finding higher fault in the algorithm. These results partially support Hypothesis 4. In the last step, gender was a significant predictor, with women faulting the vehicle less than men.

Fault Attributed to Self

Participants were also asked to attribute fault to themselves for the collision which occurred. Again, a hierarchical regression was conducted with the same predictor variables as the model for fault in the SSA. However, this time the outcome variable was the fault that the

participant attributed to themselves, as the driver. This time, the first step resulted in an R^2 of .041, predicting 4.1% of the variance in fault attribution. In the next step, R^2 increased to .057. In the next step, when driver confidence was taken into account as a predictor, the R^2 value increased to .084. Finally, when the Personal Fable Scale, Automation Complacency Scale, and gender were entered as predictors, R^2 increased to .087. See Table 7.

Table 7: Model for fault that participants attributed to themselves.

	Step 1			**Step 2**			**Step 3**			**Step 4**		
Variable	b	se	B	b	se	B	b	se	B	b	se	B
Driver Avoidability	.099	.088	.087	.102	.088	.089	.109	.088	.096	.108	.089	.094
SSA Avoidability	-.248	.090	-.211*	-.257	.090	-.220*	-.237	.090	-.202*	-.243	.091	-.207*
SSA Wrong				.764	.481	.118	.669	.480	.103	.648	.494	.100
Round Failed				-.094	.135	-.052	-.074	.134	-.041	-.080	.136	-.044
Driving-Peers							-.168	.222	-.076	-.139	.227	-.064
Driving-General							-.264	.258	-.104	-.263	.260	-.103
PF										-.295	.775	-.030
Comp										-.208	.401	-.039
Gender										.102	.538	.015
R^2	.041			.057			.084			.087		
R^2 change	.041			.016			.027			.003		

*indicates significance at $p < .05$
$^{+}$indicates significance at $p < .10$

Results showed that, when the participant perceived the incident to be more avoidable by the actions of the SSA, they blamed themselves less than if the incident did not seem avoidable. That is, if the SSA could have, but failed to, prevent an incident then participants placed less fault in themselves as the driver.

Correlations

Table 8 shows the correlations between the variables of fault attributed to the vehicle (SSA Fault), fault attributed to the driver (Driver Fault), avoidability by the actions of the vehicle

(SSA Avoidability), and by the actions of the driver (Driver Avoidability), whether the SSA was wrong last (SSA Wrong; dummy coded with 0 = the SSA was correct in the last round played, and 1 = the SSA was wrong), the participant's perceived driving abilities compared to their peers (Driving-Peers), and in general (Driving-General), as well as score on the Automation Complacency Scale (Complacency), and the Personal Fable Scale (Fable). Gender was dummy coded with a 0 indicating a male participant and a 1 indicating a female participant. Further correlations are found in the Appendix.

Table 8: Correlations between overall variables in Study 2.

	1[+]	2[+]	3[+]	4[+]	5[+]	6	7	8	9
1. SSA Fault[+]	1								
2. Driver Fault[+]	-.124	1							
3. SSA Avoidability[+]	.086	-.233**	1						
4. Driver Avoidability[+]	-.274**	-.011	.314**	1					
5. SSA Wrong[+]	-.036	.144*	.033	.009	1				
6. Driving-Peers	.101	-.142	.084	.069	-.038	1			
7. Driving-General	.015	-.163*	.084	.038	-.084	.689**	1		
8. Complacency	.004	-.068	-.050	-.081	.030	.108	.054	1	
9. Fable	.147*	-.034	-.038	.030	-.067	.245**	.196**	.052	1
10. Gender	-.144	.057	.062	.005	.218**	.001	.014	-.034	-.231**

*indicates significance at $p < .05$ level, **indicates significance at the $p < .01$ level.
$n = 188$
[+]indicates that $n = 183$

A negative correlation was found between the fault attributed to the SSA, and the perceived avoidability of the incident based on actions by the driver ($r = -.273$). Thus, the less the driver

felt that the incident was avoidable based on their own actions, the more they blamed the SSA for the outcome. Additionally, participants who scored higher on the Personal Fable Scale blamed the SSA more for the negative outcome (r = .147). When it came to attributing fault to themselves, participants tended to place higher blame on themselves the less they felt that the SSA could have prevented the incident. This is shown by the negative correlation (r = -.233). Participants tended to blame themselves more when the SSA had been given wrong information prior to the collision, and they had fallen for the bad information (r = .144). The correlations showed a significant, negative relationship between the fault a driver attributed to themselves, and their confidence in their driving skills in general (r = -.163). This indicated that the higher their confidence in their driving skills, the less fault they attributed to themselves for the outcome. There was a positive correlation between avoidability due to the actions of the driver, and due to the actions of the SSA (r = .314), indicating that participants felt that some incidents were avoidable or inevitable regardless of who took action. Not surprisingly, those who had a higher level of confidence in their driving skills in general also felt that they drove better than their peers (r = .689). Those who scored highly on the Personal Fable Scale also felt that they were good drivers in general, as well as when compared to their peers (r = .245 and r = .196, respectively).

Females scored lower on the Personal Fable (r = -.231). Additionally, there was a positive correlation between the SSA being wrong in the last round played, and gender (r = .218). This indicates that female participants were more likely to agree with the algorithm, even when it was incorrect.

Gender Differences

Women placed less fault on the SSA than men did, with an average fault score of 3.48 (2.52) compared to men's 4.26 (2.59). The difference was approaching significance ($p = .051$). There was no significant difference between the fault they placed on themselves, though women ($m = 5.33$, $sd = 3.46$) did have a higher score, on average, than men ($m = 4.94$, $sd = 3.05$). Scores on both measures of driver confidence were similar. There was, however, a significant difference in whether males or females were more likely to agree with the SSA's wrong information. Women were more likely to fail by agreeing with the SSA even when it was wrong, whereas men were less likely to do so ($p < .05$).

Discussion of Experiment Two

The hypotheses examined in this study were Hypothesis 4, which stated that those higher in their driving confidence would attribute more fault to the vehicle, and Hypothesis 5, which stated that participants with low driving confidence would attribute more fault to themselves. The correlations supported Hypothesis 5, but the regression models did not support either hypothesis, and no evidence was found to support Hypothesis 4. It is possible that the method involved in the driving task, which required decision making but no reflexes or physical driving skill, may have been different enough from actual driving that participant's confidence in their driving abilities had little effect. It is possible they did not feel as though this exercise was applicable to their real-world driving skills.

Additionally, Hypothesis 6 was examined which stated that participants would rate their driving abilities higher, the higher they scored on the Personal Fable Scale. This hypothesis was

supported by the correlations. Hypothesis 7 proposed that those with higher automation complacency scores would attribute less fault to an autonomous vehicle. However, the results did not support this hypothesis. Similar to Study 1, it is possible that complacency and trust in automation in general does not yet (and perhaps never will) extend to autonomous vehicles.

CHAPTER 4: EXPERIMENT THREE

The previous experiment examined the predictors of fault attribution when participants were the drivers of vehicles with autonomous qualities. The next, examined those predictors in a situation where participant was still a driver, but was controlling their own vehicle, and who had to encounter other autonomous vehicles and navigate a driving path without collision. The third experiment had participants play the role of the driver of a manually-operated vehicle, sharing the road with vehicles which were automated rather than controlled by other people. This experiment addressed Hypotheses Three and Four. Participants manipulated an on-screen vehicle by controlling it with the arrow keys on their computer keyboard. They had to drive their car from one side of the screen to the other, while avoiding any obstacles. If their vehicle collided with any others, or went offscreen, they failed at the task. The other vehicles onscreen moved only in relation to the participant's car, and were not controlled by human players. If participants failed at the driving task, they were given a survey to complete in which they indicated who they felt was to blame for the collision. Additionally, participant personality, automation complacency, and confidence in their driving abilities were measured as covariates.

Method

Participants:

Participants were drawn from the UCF undergraduate student body, and from volunteers. If participants were from UCF, they were rewarded with SONA credit for their participation. All participants were required to possess a valid driver's license and had to be over the age of 18.

Additionally, they had to use a computer with a keyboard, rather than a Smartphone. A total of 137 participants took part in the study. Of those, data from 14 were rejected for failing to answer key questions. The remaining 123 participants were composed 75 females, 47 males, and one individual who declined to state their sex. The average age was 24.91 (*sd* = 11.57). G*Power (Faul, Erdfelder, Lang & Buchner, 2007) indicated that for a multiple regression with five predictors to reach a power level of .95, with alpha of .05 and an expected effect size of f^2= .179 (which was calculated from the R^2 value of Experiment 2, where the same outcome variable was measured) a total sample size of 117 participants was required. Again, this number was slightly increased to account for potential lack of engagement during the online study.

Materials

Surveys and Scales:

The same scales were included in this task, as were given prior to Experiment One and Experiment Two. The Driver Confidence and Driver Experience Scales, as well as the Personal Fable and Automation Complacency scales were the same as the previous study. Here, the Fault scale was different in that participants attributed fault to themselves as the driver of the manually operated vehicle, and the other non-manually controlled vehicles and, similarly to study 2, was on a scale of 1-10. The Avoidability scale, also, involved judging the avoidability if different actions had been taken by the participant, or by the other automated vehicles, and was the same as in Study 2.

The Driving Task:

The driving task required participants to steer a vehicle across the screen. There were other moving vehicles and obstacles shown onscreen, and the participant had to reach the other side of the screen without having a collision. The participant was informed that the other vehicles were not being controlled by anyone. That is, they were not playing against another person, and the vehicles were not being controlled by someone who wanted to cause or avoid a collision. The vehicles automatically moved in relation to the participant's vehicle, and so it was up to the participant to avoid crashing.

The participant's vehicle was controlled by arrow keys. The motion was set to be as intuitive as possible, with the car moving upwards when the up key was pressed, downwards when the down arrow key was pressed, and left and right when the left or right arrow keys were pressed. The other vehicles each moved when an arrow key was pressed, but not in the same direction as the user-controlled vehicle. For instance, when the user pressed the up arrow key, their vehicle would move upwards, but an adjacent vehicle might move to the left, or move downwards and to the right diagonally. Figure 8 shows the screen that the participant viewed, and the directions in which each obstacle moved when each key was pressed. It was possible, but difficult, to complete the task successfully, and approximately 44.72 percent of participants were able to complete the driving task without a collision. This task measured some of the physical skills inherent to driving, such as reflex time and judgement of space. However, it did not examine the cognitive aspects of driving as some of the vehicles did not follow the rules of the road.

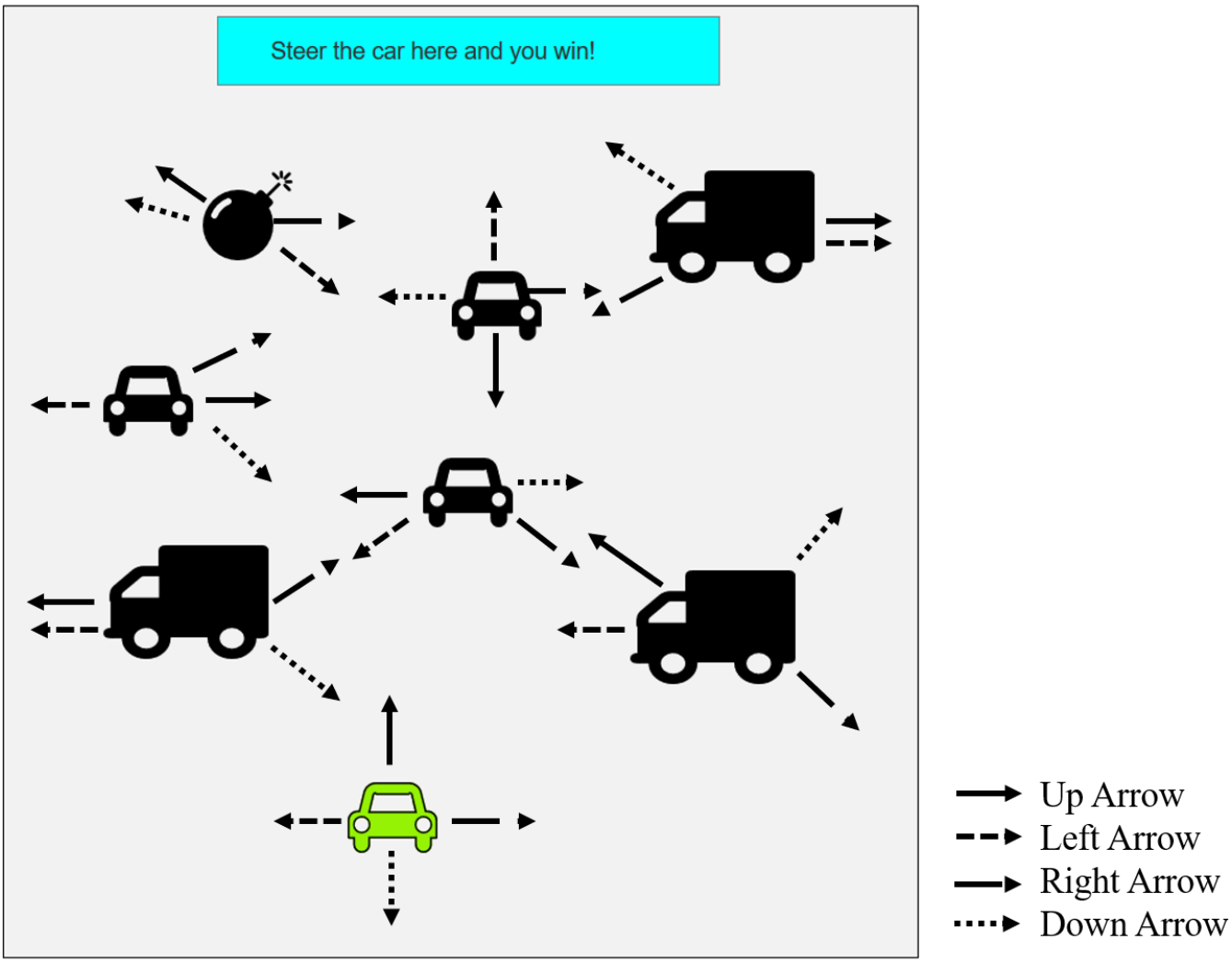


Figure 8: The driving task, and the directions in which each obstacle-vehicle moved when an arrow key was pressed.

If participants were successful, they were taken to a screen which told them they had completed the driving task without a collision. If they were not successful, i.e., if their car went offscreen, or collided with another vehicle, they were taken to a page where they were informed that there had been a collision. The driving game was created using the Axure software, and was hosted via the Axure cloud. Participants reached the game via a link in the initial online survey. Once they

had completed the task, whether successful or not, they were redirected back to the survey on Qualtrics.

Procedure

Participants read the informed consent, and agreed to take part in the experiment. Following this, they were directed to the instructions for the driving game, which explained how to steer the vehicle, and informed participants that the other vehicles moved automatically in relation to their own car, and that the latter were not actively being controlled by any other individual. When they had read the instructions, they followed a link to the driving task, which was hosted via the Axure cloud. They then completed the task by controlling the vehicle with the arrow keys, and attempting not to come into contact with any of the other moving vehicles. Following their completion of the driving task, they were redirected back to Qualtrics to complete the rest of the study. Participants who had been involved in a collision were directed to the Fault and Avoidability Surveys, where they attributed fault between both themselves, and the other vehicles, and rated the avoidability of the collision both based on actions they could have taken as the driver, and actions the other vehicles might have taken. Then, they completed the driver experience, driver confidence, personal fable and automation complacency surveys. Participants who had avoided a crash were directed straight to the follow-up surveys, skipping over the measures of Fault and Avoidability.

Results

The Driving Task

Participants had to steer a virtual car by using the arrow keys, while avoiding collisions with other, non-human-controlled vehicles, which moved only in relation to their own. Of the 123 participants who completed the entire study a total of 55, or 44.72 percent of the participants, managed to successfully complete the task. Sixty-eight were involved in a collision. Data from those 68 were used in the fault attribution models and descriptive statistics for fault and avoidability, while data from the entire set of 123 was used in the correlations and other descriptive statistics. There was no significant difference in any of the driving confidence measures between those who succeeded or failed on the task.

Descriptive Statistics

Fault Attribution: The 68 participants who did not successfully complete the driving task without collision attributed fault to both themselves, and the other vehicles. They provided a self-fault rating of 5.01 (2.62) on the 1-10 scale, and gave similar fault attributions to the other vehicles involved in the collision (m = 5.04, sd = 3.14).

Avoidability: Participants who had been in a collision, rated the avoidability of that collision based on actions they could have taken, and actions the other vehicles could have taken. They rated the avoidability of the collisions based on their own actions a 4.94 (2.85), and believed the other vehicles had slightly more ability to avoid the collision (m = 5.82, sd = 3.07).

Driving Abilities Overall: The means for driving abilities included all 123 participants. When asked to rate their vehicle handling skills, participants gave themselves a score of 7.62 (1.38).

Their driving judgement was scored at 7.46 (1.37) and their driving reflexes were rated 7.51 (1.46). Overall, their average driving skills were perceived to be 7.53 (1.72).

Driving Abilities Compared to Peers: Participants were asked to rate their driving skills compared to their peers. Their vehicle handling skills were rated at 7.11 (1.55). Their driving judgement was 7.02 (1.58), and their driving reflexes were 7.03 (1.68). Overall, compared to their peers, participants rated their driving abilities at 7.05 (1.49), considering themselves better than the average driver in their age group. This was, again, a common finding across all three experiments.

Personal Fable: The Personal Fable Scale measured participants' beliefs about their own omnipotence, invulnerability, and uniqueness on a scale of 1-5. For omnipotence, participants rated themselves an average of 2.95 (0.59). On the invulnerability subscale, participants rated themselves at a 2.92 (0.55), and on the uniqueness subscale the average score was 3.34 (0.49). The average on the scale overall was 3.07 (0.37). Scores on all these subscales were very similar to those on the same scales as reported in Experiment 1.

Automation Complacency: Automation complacency was measured on the subscales of confidence in automation, reliance on automation, trust in automation, and belief in the safety of automation. On the confidence subscale, participants scored an average 3.53 (1.82). For reliance, the average score was 3.23 (1.99) and for trust the average was 2.88 (0.86). For belief in the safety of automation, the average score was 2.93 (1.74). Overall, the average score for automation complacency was 3.16 (1.30).

Fault Attributed to the Vehicles

When participants were asked to attribute fault to the vehicles for the collision, the same variables were considered as in previous studies. Again, a hierarchical regression was conducted. In the first step, avoidability due to the actions of the vehicles (Vehicle Avoidability) and the participant (Driver Avoidability) accounted for a R^2 value of .111. When driver confidence was considered, that value increased to .129. Finally, scores on the Personal Fable and Automation Complacency scales, along with gender, only increased R^2 to .137. See Table 9.

Table 9: Participants' fault attribution towards the vehicles

	Step 1			**Step 2**			**Step 3**		
Variable	b	se	B	b	se	B	b	se	B
Vehicle Avoidability	-.036	.120	-.036	-.020	.121	-.019	-.033	.130	-.033
Driver Avoidability	-.351	.129	-.326*	-.366	.130	-.339*	-.383	.140	-.355*
Driving-Peers				-.337	.377	-.171	-.305	.396	-.154
Driving-General				.487	.433	.214	.507	.466	.223
PF							-.108	1.45	-.011
Comp							.076	.601	.017
Gender							.588	.863	.092
R^2	.111			.129			.137		
R^2 change	.111			.018			.008		

*indicates significance at $p < .05$
$^{+}$indicates significance at $p < .10$

There was a negative relationship between perceived avoidability due to actions of the driver, and the fault attributed to the vehicles. That is, the less the participant felt that they could have had any control over the outcome, the more they blamed the vehicles for that outcome. While this is not a surprising finding, what is surprising is the fact that avoidability due to actions of the vehicles was not a significant predictor, and nor was driver confidence.

Fault Attributed to Self

A hierarchical regression was conducted to determine what variables predicted fault that participants attributed to themselves. In the first step, perceived avoidability due to actions of the participant (Driver Avoidability), and the vehicles (Vehicle Avoidability) were entered, based on results from the previous experiments. With these predictor variables, R^2 was .083. In the next step, both types of driver confidence changed R^2 by .054, increasing it to a value of .137. In the third and final step, gender, as well as the scores from the Personal Fable and the Automation Complacency scale were added, bringing R^2 to .158. See Table 10.

Table 10: Regression predicting fault in self.

	Step 1			**Step 2**			**Step 3**		
Variable	b	se	B	b	se	B	b	Se	B
Vehicle Avoidability	.115	.104	.134	.139	.103	.162	.138	.109	.161
Driver Avoidability	.217	.110	.236^{+}	.190	.111	.207^{+}	.150	.118	.163
Driving-Peers				-.591	.320	-.351^{+}	-.536	.333	-.318
Driving-General				.685	.367	.354^{+}	.643	.375	.332^{+}
PF							-.573	1.22	-.069
Comp							.468	.506	.125
Gender							.016	.726	.003
R^2	.083			.137			.158		
R^2 change	.083			.054			.021		

*indicates significance at $p < .05$
$^{+}$indicates significance at $p < .10$

There was a positive relationship between fault that participants attributed to themselves, and perceived avoidability due to actions of the participant. That is, participants felt that the

more power *they* personally had to avoid the collision, the more fault rested with themself. Surprisingly, avoidability due to actions of the vehicles did not prove to be a significant predictor. In the final step, driver confidence in their abilities in general, had a positive relationship with fault. Here, participants who felt that their driving skills were very good, tended to place higher blame on themselves for a collision. This is surprising, but in line with the idea that their own avoidability predicting the fault they attributed to themselves.

Correlations

Correlations between the overall variables in this experiment are presented in Table 11. Variables include of fault attributed to the vehicles (Vehicle Fault), fault attributed to the driver (Driver Fault), avoidability by the actions of the vehicle (Vehicle Avoidability), and by the actions of the driver (Driver Avoidability), the participant's perceived driving abilities compared to their peers (Driving-Peers), and in general (Driving-General), as well as score on the Automation Complacency Scale (Complacency), and the Personal Fable Scale (Fable). Gender was dummy coded with a 0 indicating a participant was male, and a 1 indicating they were a female. Further correlations are found in Appendix A.

Table 11: Correlations between overall variables in Study 3

	1[+]	2[+]	3[+]	4[+]	5	6	7	8
1. Vehicle Fault[+]	1							
2. Driver Fault[+]	-.314**	1						
3. Vehicle Avoidability[+]	-.100	.173	1					
4. Driver Avoidability[+]	-.333*	.257*	.152	1				
5. Driving-Peers	.053	-.060	-.049	-.122	1			
6. Driving-General	.120	.055	-.086	-.035	.528**	1		
7. Complacency	-.058	.193	-.089	.251*	.165	-.472**	1	
8. Fable	.056	-.189	-.251*	-.211	.392**	.446**	-.245**	1
9. Gender	.020	.047	.142	.169	-.202*	-.063	-.106	-.213*

*indicates significance at $p < .05$ level, **indicates significance at the $p < .01$ level
$n = 123$
[+]indicates $n = 68$

There was a significant, negative correlation between fault attributed to the driver and fault attributed to the vehicles ($r = -.314$). This outcome indicated that participants tended to blame either themselves or the vehicles for the collision but not both. Additionally, there was a negative correlation between the perception that the collision was avoidable by the driver, and the fault attributed to the vehicle ($r = -.333$), which means, quite understandably, that the less the

driver felt they could have prevented the collision, the more they faulted the other vehicles. There was a logical and positive relationship between driver fault in themselves, and their perception that they could have avoided the collision (r = .257). The more the driver felt they had the ability to effect the outcome, the more they blamed themselves when that outcome proved to be aversive.

The higher participants scored on the Personal Fable Scale, the less they felt that the other vehicles in the situation had any control over the collision outcome (r = -.251). The higher their complacency with automation, the more they felt that they themselves had the ability to control the collision (r = .251). Their confidence in their driving skills compared to their peers was, again, highly correlated with their confidence in their driving skills in general (r = .528) and their scores on the Personal Fable Scale (r = .392). Their confidence in their driving skills, in general, was *negatively* correlated with their automation complacency (r = -.472) but positively correlated with their Personal Fable score (r = .446). Their complacency with automation was also negatively correlated with their scores on the Personal Fable (r = -.245). Females scored lower on the Personal Fable (r = -.213), and were less likely to consider themselves better drivers than their peers (r = -.202).

Gender Differences

There were no significant differences between the level of fault that the different genders attributed either to themselves, or to the vehicles. Additionally, there was no significant difference between the scores the participants gave themselves for their driving abilities in general. However, when asked to rate their driving abilities compared to their peers, there was a

significant difference ($p < .05$) between how males and females rated themselves. Males gave themselves an average rating of 7.44 (1.27), while females rated themselves at 6.82 (1.58).

Participants who Failed at the Driving Task Compared to those who Succeeded

The 68 participants who failed at the driving task were compared to the 55 who succeeded, in terms of their driver confidence. When ranking their driving skills in general, those who succeeded gave themselves a 7.19 (2.09) while those who failed rated their driving skills at 7.37 (1.37). There was no significant difference. Compared to their peers, those who succeeded rated themselves 7.18 (1.54), while those who failed rated themselves slightly lower at 6.95 (1.45) but not significantly so.

Discussion of Experiment Three

The findings in the correlational examination supported Hypothesis 6, which stated that participants who scored highly on the Personal Fable Scale would have a higher confidence in their own driving abilities. The only significant predictor of either fault attributed to the vehicles or to the driver, was perceived avoidability due to actions of the driver and, in the third step of the model predicting fault in the self, driver confidence, which was positively related to fault in the self.

In general, the more participants believed that they, *themselves*, had the ability to avoid collision, the more fault they attributed to themselves and the less they attributed to the other vehicles. Rationally, this makes sense, as if they could have avoided a collision but failed to do so, they were aware that the blame lay largely with themselves. If they felt their actions could not

have prevented a collision, and yet a collision still occurred, then it stands to reason that they would attribute fault to the other vehicles involved.

Perceived avoidability due to actions of the vehicle was not a significant predictor of fault. This may be due to the fact that participants were aware that the non-controlled vehicles moved only in relation to their own. Additionally, it was anticipated that complacency with automation may have exerted an influence on fault attribution, but this was not the case here. Perhaps participants' feelings towards automation, in general, do not extend to automated vehicles because of the familiarity factor at this time. Perhaps also, the stakes of an online driving game were not of sufficient importance to elicit a response, and perhaps the task was dissimilar enough from actual driving.

The correlational results showed that, again, those high on the Personal Fable Scale also thought highly of their own driving abilities. This is unsurprising, as the Personal Fable Scale measures one's feelings that they are special in some way, and it is reasonable that they would feel their abilities in many fields-including driving-may be above average (Lapsley et al., 1989). It was fairly unexpected that participants' confidence in their driving skills in general was negatively related to their complacency with automation. It is possible that, knowing the study involved autonomous vehicles, they were primed prior to completing the scale of automation complacency. This awaits further investigation.

CHAPTER 5: GENERAL DISCUSSION

As self-driving cars become more common and begin to share the road with regular drivers not as rare novelties, but everyday vehicles, the number of collisions in which they are involved will only increase. Therefore, it is necessary to understand the differences between how people attribute fault to these vehicles compared to non-autonomous, manually-operated cars. A thorough identification of the way different factors affect fault in autonomous vehicles will have implications for the operators of those vehicles, the victims of collisions, and any jury involved in legal cases concerning such events.

Perhaps of interest is that the predictor variables mainly examined here (perceived avoidability of the accident and driver confidence, among other personality factors) are not easily manipulated. That is, unlike anthropomorphism, transparency, or other factors relating to the automation, they cannot be tweaked for the direct purpose of mitigating fault or altering perceptions. This is intentional in that the rationale of the present work was to make an observation, and not a suggestion. The work is not directed as to examine how a manufacturer might be able to avoid fault for any collisions their product is involved in, but merely to help explain the ways in which a person may attribute fault differently when a vehicle is powered by an algorithm rather than controlled by a human.

Results of from the first experiment examined the extent to which the fault attributed to autonomous vehicles, as well as their operators, was influenced by driver confidence and collision avoidability. Additionally, results of that study examined whether autonomous vehicles were judged more harshly than regular cars, for their involvement in identical accidents. The

second experiment showed how people react when they are in the driver's seat of an autonomous vehicle which is involved in a collision, and the ways in which their own confidence and perception of collision avoidability influenced the fault they placed on both themselves, and the vehicle. Experiment Three approximated the experience of being a regular driver of a manually controlled vehicle, while other vehicles on the road were not directly controlled by a human driver and move only in relation to one's own vehicle.

The Hypotheses

Hypothesis 1: The first hypothesis stated that "when accidents are perceived to be at medium or low avoidability, participants will attribute greater fault to manually-operated vehicles than to autonomous vehicles involved in identical collisions." This hypothesis asserted that a more avoidable collision would lead to higher fault attributed to the autonomous vehicles, or autonomous aspects of the vehicles in question. This proposition was not supported by the results of Experiment 1, where the fault model showed no significant interaction between perceived avoidability and automation condition. However, avoidability did predict fault attribution, so Hypothesis 1 was partially supported in the sense that the more avoidable the collisions, the greater the fault attribution. Experiment 2 showed that the less avoidable a participant thought the collision was, based on their own actions, the more fault they attributed to the SSA. So a collision with low avoidability did contribute to fault attribution in the SSA. However the opposite was also true, when a collision appeared to be difficult to avoid due to the actions of the SSA, the driver attributed more fault to themselves. Findings from Experiment 3 were similar with higher fault attributed to the vehicles when the driver did not believe they could have taken action to avoid the collision, and higher fault attributed to the driver when they

believed that they *could* have avoided the collision. Overall, Hypothesis 1 was only partially supported.

Hypothesis 2: Hypothesis 2 stated "participants will attribute greater fault to a self-driving vehicle than to a regular vehicle when they believe that the accident was highly avoidable." This hypothesis was not supported in Experiment 1. While there was evidence of a significant interaction between avoidability and automation condition, it was not in the expected direction. However, avoidability did predict fault in general. The correlations showed a negative relationship between avoidability and fault attributed specifically to the algorithms, meaning that the more avoidable a collision seemed, the less fault was attributed to the algorithms. This was the opposite of what was predicted in Hypothesis 2. In Experiment 2, fault in the SSA was positively predicted by perceived avoidability on the part of the vehicle, and negatively predicted by the driver's ability to avoid the collision. So participants here believed that the more the algorithms had a chance to avoid the collision, the more at fault it was for any resulting incidents. Experiment 3 showed that fault in the vehicles was negatively predicted by the driver's ability to avoid a collision, meaning that when involved in a highly avoidable collision, the algorithms received *less* blame than a human driver. Overall, Hypothesis 2 was not supported.

Hypothesis 3: Hypothesis 3 proposed that "participants higher in confidence in their own driving abilities will attribute greater fault towards human drivers – of both autonomous and regular vehicles - than those with low confidence in their driving abilities." This hypothesis asserted that driver confidence would be a predictor of fault attribution. Experiment 1 actually found the opposite in the overall fault model, showing that high driver confidence *negatively*

predicted fault attribution overall. However, when it came to just the drivers of autonomous vehicles, driver confidence predicted lower fault attributed to the algorithms and more attributed to the human driver. Hypothesis 3 was also partially supported.

Hypothesis 4: This hypothesis stated that "participants higher in confidence in their own driving abilities will attribute greater fault to a self-driving vehicle, compared to people with low confidence in their own driving abilities." This was not supported by the results of Experiment 1. Results from Experiment 3 also failed to support this hypothesis. However, in Experiment 2, participants with higher levels of confidence in their driving abilities proved more likely to attribute fault to the SSA. Hypothesis 4 was partially supported.

Hypothesis 5: This hypothesis stated that "participants with low confidence in their own driving abilities will attribute greater fault to themselves than people with high confidence in their own abilities, even when the collision was unavoidable." The proposition was examined in Experiments 2 and 3, where the participant was a driver. In Experiment 2, confidence was negatively correlated with fault that the participants attributed to themselves as a driver. In Experiment 3, driver confidence compared to peers had a negative relationship with fault in the self, meaning that those with high driver confidence attributed less fault to themselves, however driving confidence in general had a positive relationship where those with higher confidence in their driving skills, in general, attributed more fault to themselves. These findings partially support Hypothesis 5.

Hypothesis 6: This hypothesis predicted "the higher a participant rates themselves in their driving abilities compared to their peers, the higher their score will be on the Personal Fable

Scale (Lapseley et al., 1989).” This was supported by the correlational results in all three experiments. There were positive relationships between the Personal Fable score and driver confidence both overall, and compared to one’s peers.

Hypothesis 7: Hypothesis 7 stated “those with higher automation complacency will have a higher blame for the driver of an automated vehicle than for that vehicle’s algorithms.” The only significant finding relating to automation complacency was in Experiment 3, where it was positively related to the extent to which the *driver*, rather than the vehicles, could have avoided the collision. While this may indirectly support Hypothesis 7, overall the hypothesis was not supported by the data presented here.

Predictors of Fault

The variables that were significant predictors of fault attribution varied slightly between experiments. In Experiment 1, where participants were witnesses to collisions, perceived avoidability was a positive predictor, and driver confidence was a negative predictor. Those who had high confidence in their own driving abilities actually attributed *less* fault to others, compared to those who did not have high confidence in that regard. This may be a form of empathetic pity. Additionally, the more avoidable a collision appeared, the more fault was attributed to both autonomous and regular vehicles.

In Experiment 2, participants adopted the role of the driver of a vehicle with an autonomous algorithm that made safety suggestions. Here, if they were involved in a collision, they attributed fault to both themselves and the algorithm. For the model predicting fault in the algorithm, perceived avoidability based on actions of the driver negatively predicted fault. This

meant that the less the driver could have avoided the collision, the more fault was placed on the algorithms. Avoidability by the algorithms was a positive predictor, meaning that if the algorithm could have easily prevented the collision, it was perceived as more at fault for any collisions that did occur. One aspect of driving confidence was a positive predictor, with participants who had high confidence in their driving skills compared to their peers, placing more blame on the algorithm. Additionally, participants who scored higher on the Personal Fable Scale attributed more fault to the vehicle, meaning that the more unique and invulnerable they thought they were, the more fault for a collision (even one caused by a shared decision) was attributed to the vehicle. When participants judged the fault that they themselves deserved as the driver, fault was negatively predicted by avoidability based on the algorithm, meaning that the less the algorithm could have helped prevent the collision, the more the driver blamed themself for the outcome. Additionally, attributed fault was higher when the SSA had made a wrong suggestion prior to the incident. The participants faulted themselves higher for failing to spot bad information. This pattern may result from the way that agency is still not granted to non-living beings.

Experiment 3 had the participant driving a vehicle across a screen where the other vehicles were not manually controlled, mimicking the situation of sharing the road with autonomous vehicles. Here, fault in the other vehicles was negatively predicted by the extent to which the participant believed they could have avoided the collision, meaning that if they could have easily avoided the collision themselves, they attributed less fault to the automated vehicles involved. Fault in themselves was positively predicted by the same variable, where they blamed themselves more when they believed they could have easily avoided the collision. Additionally,

fault that participants attributed to themselves was positively predicted by their driving confidence, where those with a higher opinion of their own driving skills accepted more blame.

Overall, the results from the three experiments showed that the level of perceived avoidability was a consistent predictor of fault attribution. Patterns were similar, whether the fault was being attributed to autonomous or human-controlled vehicles. While it may be surprising that fault attribution towards both human drivers and algorithms were similar, there are many reasons why this may be so. It is possible that participants attribute fault abstractly, without knowing anything about other party involved in the collision. They may attribute fault to the other party similarly regardless of whether that party is a fellow student, a neighbor, a stranger, or an autonomous vehicle. It is possible that participants viewed strangers and autonomous objects fairly similarly, since neither one was a part of their daily experience.

Real-World Implications

There are many ways in which fault attribution plays an important role in modern society. Perceived fault for any negative outcomes may serve to influence trust. While trust in some forms of automation, such as robotics, has many antecedents, fault has not been explored as a precursor to trust (Hancock et al., 2020). However, it is quite possible that fault is one of the many trust predictors that determine how much someone will subsequently trust, and ultimately use, a technological advancement.

Autonomous vehicles are one such advancement. If they are not deemed to be trustworthy, then there may be relatively few who will risk the very-real consequences of relying on such technology. However, if they are deemed trustworthy, and do become more common,

the potential for a collision-and thus a decrease in trust-only increases. This may be thought of as a ‘trust conundrum.’ For this reason it is necessary to understand where one will place fault when a collision occurs with an autonomous vehicle. The findings presented here indicate that fault attribution is similar between human drivers and automation. The factors that made participants more likely to fault drivers also applied to their fault attributed to autonomous vehicles, and largely, the difference in the magnitude of fault attributed was not significant. In short, fault is perceived as similar regardless of whether or not a vehicle was autonomous. This may be because people tend to be unable to distinguish between autonomous and regular vehicles based solely on the vehicle’s maneuvers while driving (Stanton et al., 2020).

I several ways, this is an encouraging finding. Collisions-even deadly ones-have been an accepted part of the transportation industry for a long time. Though people may fault each other in collisions, there has not yet been a time when groups of people have been banned from driving, or when roads have closed due to the inherent risk of their utilization. The potential for a collision has been considered an acceptable risk in transportation, and though strategies exist to mitigate potential dangers, every driver gets behind the wheel with the knowledge that one inherent aspect of driving is the chance of collision. Since this has not stopped people from driving before, and the factors affecting fault attribution for autonomous vehicles are similar to those affecting the blame of other drivers, it stands to reason that the risk of collision with an autonomous vehicle will be considered another such acceptable risk. Of course, the objective risk and the perception of that risk will remain influential.

Though the experiments presented here focused on autonomous vehicles, the findings may well apply to other technological domains such as aviation, boating, or non-transportation areas such as medicine. Every instance where a human may be replaced by automation carries with it some of the risk of failure that existed when a human was involved and may involve new sorts of risk also. Whereas before, human error was a known component of these processes, automation error is less commonly understood. If the findings here apply to other operational realms, it is possible that *any* scenario where autonomous technology causes an incident will be perceived as being similar to, and as acceptable as, the same outcome when caused by a human.

Even if the findings apply only to the domain of road transportation, the implications will have effects in several ways. For instance, legal defenses will necessarily be different when it is someone's technology, and not their own client's error, that causes a collision. Whether the driver of an autonomous car can be blamed for failing to appropriately monitor their vehicle, remains to be seen. However, the results of Experiment 1 imply that defense attorneys will not have to alter their tactics as technology becomes more advanced. The same precursors of fault that affected their case when human drivers cause collisions, will still be relevant when the incident is the result of a poor choice by automation. The fact that individuals blamed human drivers more than the automation with which they shared control, indicates that the future of legal defenses in this field will remain very similar to its past.

Policy-making, too, will be most likely be affected by any increase in autonomous vehicles. However, policy-makers will most likely base their decisions upon their own attributions relating to new technology. If they blame automation more than humans for an

incident, then associated laws will reflect that fact. However, if the findings expressed here can be extended, it is likely that future blame for collisions will not be far different from that which exists currently, and whether one's vehicle is autonomous or manual will not have a large impact on any potential repercussions of a collision, if the present attribution profile holds.

Driver training currently helps novice drivers to become aware of the potential future actions of others on the road. By anticipating where another driver may turn, one can potentially avoid a collision. With an autonomous vehicle, such anticipation may be difficult. When predicting whether a human driver will turn left, one can see that driver's eyes and body language, and so be able to determine in which direction they are looking. This is not possible with autonomous vehicles, which have sensor cameras surrounding them (Hancock, 2018). However, the principal issue in predicting a driver's actions based on their gaze is not where the driver is looking, but actually where they are *looking away from.* Autonomous vehicles do not have this problem, as they can 'look' in more than one direction at a time. When viewed from a bird's eye view, the perceived avoidability of a collision where a driver pulls out and hits a pedestrian may seem extremely high, as the pedestrian is obvious to witnesses. What witnesses do not know, however, is in which direction the driver was looking. For a high-avoidability collision, fault attribution will likewise be high. Autonomous vehicles can avoid this sort of high-avoidability collision through the use of their sensors. Therefore, any collision in which automation is involved is more likely to have low avoidability. As perceived avoidability is the main predictor of fault attribution, it is likely that in general there will be little blame attributed to autonomous vehicles not for identical collisions, but because the type of collision in which they are involved will appear to witnesses to be less avoidable.

Consistent Findings

In all three studies, females scored lower on the Personal Fable Scale (Lapsley et al., 1989). This was shown in the correlations. As this scale measured one's feelings of invulnerability and uniqueness, it is possible that the male participants did indeed feel that they were more special and unique, but it is also possible that the difference lay in how participants chose to answer the questions, and not in their true opinions of themselves.

Another strong, consistent finding was that people considered themselves better drivers than average. In all three experiments, participants gave high ratings of their driving abilities both overall, and compared to their peers. This is not surprising as it is in line with previous research (for example, see Wohleber & Matthews, 2016), but serves as additional support for the theory that most people rate their driving skills as being above average.

Limitations

There were several limitations which could affect the experiments reported here. Due to a global pandemic of the novel coronavirus 2019 (COVID-19), planned experiments in a driving simulator had to be moved to an online format for safety reasons. This led to several potential concerns. First, the online format may have led participants to take the study less seriously. While there were questions intended to determine whether participants were paying attention and reading each question carefully, there was no way to determine the extent to which each participant put thought into their answer. Being in an online format, removed from any researcher, it is possible that they did not put sufficient effort into the survey. However, results in

general were in-line with *a priori* expectations, and results were fairly consistent between studies. Thus it appears likely that participants did attempt to fill out all surveys accurately.

Additionally, the driving tasks were somewhat removed from the everyday reality of driving that the participants were used to. Perhaps a participant's confidence in their driving ability does not translate to their confidence in their abilities to control a vehicle via keyboard, as in Experiment 3, or to make good decisions with the help of the SSA, as in Experiment 2. These factors may have affected the hypothesized relationships between driving confidence and fault attribution.

One more aspect of autonomous vehicles that was difficult to examine empirically was the fact that a single autonomous vehicle does not act alone, in the same way that a single driver does. The algorithms that power these vehicles are constantly interacting with other information, be it stored in a cloud or in one's personal cell phone. To judge a single vehicle is ignoring a large part of the network of algorithms that all interact in order to make autonomous driving possible. However, the experiments presented here only examine the fault attributed to individual vehicles, and not to the conglomerate. This was necessary for comparative purposes, as an individual is generally judged in a more favorable light than is the aggregate (see Giladi & Klar, 2002). That is to say, asking participants to judge the fault of *an individual* autonomous vehicle is fundamentally different than asking them to judge the fault of the *aggregate* of autonomous vehicles, even though the vehicles do not behave or make decisions on an individual basis. The results of the scale measuring fault attribution were meant to be compared to those same ratings for *individual* human drivers, and thus the only direct comparison could be a rating of *individual*

autonomous vehicles. This is a potential flaw of the data, as it is not an entirely realistic view of autonomous vehicles.

Future Work

Future work will include a driving-simulator study in order to confirm the results of Experiments 2 and 3. Since both experiments took place on a computer, a simulator will lend a stronger degree of real-world credibility to the experience and will determine whether the experiments did in fact capture the experience of driving an autonomous vehicles or sharing the road with such vehicles. Additionally, further examination may include a wider variety of personality measures such as the Big Five personality traits of Openness, Extroversion, Conscientiousness, Agreeableness, and Neuroticism (Goldberg, 1993). It has already been shown that extroversion increases one's tendency to anthropomorphize, and to like, a robot (Kaplan, Sanders, & Hancock, 2019). It is possible that this same effect extends to autonomous vehicles. Future work will tell what other factors may influence acceptance of, and fault attribution to, autonomous vehicles.

Conclusion

In the early stages of development of self-driving vehicles, there was great hope that such automation would eliminate human error and thereby greatly reduce the number of traffic accidents. This will certainly be closer to the truth at level five automation, where vehicles can monitor themselves in nearly all situations and can indeed be said to be 'driverless." However, most automated vehicles on the road currently contain level two automation, requiring a great deal of human intervention and tasking the operator with remaining vigilant during the drive. If

such self-driving vehicles become more common before they become more autonomous, there may even be an increase in the number of collisions as humans struggle to adapt. Perhaps in a more distant future, there will be less opportunity for fault in general. In the immediate coming interval, however, fault attribution in vehicular collisions remains of critical concern and importance.

APPENDIX A: CORRELATIONS

Correlations for Experiment One

Variables are fault attributed to the vehicle in question (Fault), perceived avoidability of the collision (PA), whether or not the vehicle was automated (Auto; dummy coded as 0=regular, 1= autonomous), fault in the algorithms (Alg), confidence in driving abilities compared to peers (Peers), vehicle handling skills compared to peers (VHP), driving judgement compared to peers (DJP), driving reflexes compared to peers (DRP), confidence in driving abilities in general (Driving), vehicle handling skills in general (VHG), driving judgement in general (DJG), driving reflexes in general (DRG), score on the Automation Complacency Scale (AC), with the subscales on that Scale of confidence, reliance, trust, and safety (ACC; ACR; ACT; and ACS), and score on the Personal Fable (PF) with subscales of Omnipotence, Invulnerability, and Uniqueness (PFO; PFI; and PFU).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Fault	1																				
2. PA	.76**	1																			
3. Auto	.04	.01	1																		
4. Alg	-.13**	-.21**	X	1																	
5. Peers	-.02	.03	-.00	-.10*	1																
6. VHP	-.02	.03	.00	-.11*	.94**	1															
7. DJP	-.01	.03	-.02	-.09*	.93**	.83**	1														
8. DRP	-.02	.02	.01	-.08*	.92**	.80**	.77**	1													
9. Driving	.02	.06*	.01	-.13*	.67**	.61**	.63**	.62**	1												
10. VHG	.02	.04	.00	-.11*	.56**	.57**	.54**	.47**	.86**	1											
11. DJG	.03	.06*	.02	-.09*	.58**	.52**	.60**	.50**	.90**	.67**	1										
12. DRG	.01	.05*	.02	-.14*	.62**	.54**	.54**	.66**	.90**	.65**	.71**	1									
13. AC	.02	.03	.01	-.06	.05*	.05*	.07**	.02	.10**	.10**	.10**	.07**	1								
14. ACC	.03	.04	.01	-.07	.08**	.07**	.10**	.04	.12**	.11**	.09**	.09**	.85**	1							
15. ACR	.01	.02	.02	-.08*	-.04	-.04	-.02	-.04	.04	.04	.08**	-.01	.83**	.69**	1						
16. ACT	.01	-.01	.01	-.05	.05*	.05	.05	.05	.09**	.06*	.10**	.07**	.74**	.50**	.56**	1					
17. ACS	.02	.03	-.01	.01	.06*	.08**	.07**	.02	.07**	.10**	.05*	.06*	.71**	.42**	.39**	.36**	1				
18. PF	-.00	.04	-.02	-.07	.37**	.35**	.34**	.34**	.29**	.22**	.25**	.29**	.11**	.13**	.07**	.07**	.08**	1			
19. PFO	-.01	.02	.00	-.08*	.39**	.36**	.34**	.38**	.32**	.23**	.28**	.32**	.10**	.08**	. 06*	.05*	.12**	.83**	1		
20. PFI	.01	.03	-.01	-.05	.27**	.26**	.23**	.26**	.26**	.23**	.20**	.26**	.05*	.09**	.013	.014	.04	.81**	.56**	1	
21. PFU	.00	.05*	-.03	-.04	.20**	.19**	.22**	.16**	.10**	.06*	.11**	.08**	.11**	.13**	.10**	.09**	.03	.72**	.38**	.34**	1

*indicates significance at p < .05 level, **indicates significance at the p < .01 level, X indicates a correlation that could not be determined because one of the variables is constant.

Correlations for Experiment Two

Variables are fault attributed to the SSA in the vehicle (FV), fault that the participant attributed to themselves as the driver (FD), perceived avoidability of the collision due to actions of the vehicle (PAV), and the driver (PAD), whether or not the algorithm was wrong last (wrong; dummy coded as 0=the driver was wrong last, 1= the SSA was wrong last), confidence in driving abilities compared to peers (Peers), vehicle handling skills compared to peers (VHP), driving judgement compared to peers (DJP), driving reflexes compared to peers (DRP), confidence in driving abilities in general (Driving), vehicle handling skills in general (VHG), driving judgement in general (DJG), driving reflexes in general (DRG), score on the Automation Complacency Scale (AC), with the subscales on that Scale of confidence, reliance, trust, and safety (ACC; ACR; ACT; and ACS), and score on the Personal Fable (PF) with subscales of Omnipotence, Invulnerability, and Uniqueness (PFO; PFI; and PFU).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. FV	1																					
2. FD	-.12	1																				
3. PAV	.09		1																			
4. PAD	-.27**	-.23**	.31**	1																		
5. Wrong	-.04	-.01	.03	.01	1																	
6. Peers	.10	-.14	.08	.07	-.04	1																
7. VHP	.10	-.20**	.113	.11	-.07	.92**	1															
8. DJP	.09	-.06	.05	.06	.03	.92**	.77**	1														
9. DRP	.09	-.14	.06	.03	-.07	.93**	.78**	.77**	1													
10. Driving	.02	-.16*	.08	.04	-.08	.69**	.61**	.67**	.63**	1												
11. VHG	-.08	-.10	.08	.11	-.11	.58**	.55**	.56**	.50**	.91**	1											
12. DJG	.12	-.18*	.10	-.03	-.05	.64**	.56**	.67**	.54**	.92**	.75**	1										
13. DRG	-.00	-.16*	.04	.02	-.07	.66**	.55**	.60**	.67**	.91**	.72**	.75**	1									
14. AC	.00	-.07	-.05	-.08	.03	.11	.09	.15*	.06	.05	.04	.05	.05	1								
15. ACC	-.02	-.11	.07	.01	-.11	.06	.05	.08	.03	.12	.13	.07	.13	.68**	1							
16. ACR	-.05	-.06	.05	-.02	.02	-.01	.01	.03	-.06	-.01	.01	-.01	-.02	.75**	.60**	1						
17. ACT	.06	-.06	-.00	-.13	.01	.11	.09	.16*	.07	.01	-.01	.04	-.01	.72**	.32**	.46**	1					
18. ACS	.02	.03	-.20**	-.09	.15*	.12	.08	.14	.10	.01	-.02	.03	.01	.57**	-.05	.09	.33**	1				
19. PF	.15*	-.03	-.04	.03	-.07	.25**	.20**	.23**	.25**	.20**	.12	.16*	.25**	.05	.03	-.01	.07	.05	1			
20. PFO	.17*	.00	-.00	.06	-.05	.20**	.14	.19*	.21**	.17*	.08	.16*	.23**	.11	.08	.00	.08	.10	.82**	1		
21. PFI	.09	.02	-.13	-.01	-.11	.14	.11	.10	.17*	.12	.10	.05	.19**	.01	-.01	-.04	.02	.05	.74**	.55**	1	
22. PFU	.04	-.09	.06	.02	.01	.18*	.18*	.20**	.13	.12	.07	.14	.11	-.01	-.01	.02	.06	-.05	.53**	.17*	-.02	1

n = 188

*indicates significance at p < .05 level, **indicates significance at the p < .01 level

Correlations for Experiment Three

Variables are fault attributed to the vehicles (FV), fault that the participant attributed to themselves as the driver (FD), perceived avoidability of the collision due to actions of the vehicle (PAV), and the driver (PAD), confidence in driving abilities compared to peers (Peers), vehicle handling skills compared to peers (VHP), driving judgement compared to peers (DJP), driving reflexes compared to peers (DRP), confidence in driving abilities in general (Driving), vehicle handling skills in general (VHG), driving judgement in general (DJG), driving reflexes in general (DRG), score on the Automation Complacency Scale (AC), with the subscales on that Scale of confidence, reliance, trust, and safety (ACC; ACR; ACT; and ACS), and score on the Personal Fable (PF) with subscales of Omnipotence, Invulnerability, and Uniqueness (PFO; PFI; and PFU)

	1[+]	2[+]	3[+]	4[+]	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. FV[+]	1																				
2. FD[+]	-.31**	1																			
3. PAV[+]	-.10	.17	1																		
4. PAD[+]	-.33**	.26*	.15	1																	
5. Peers	.05	-.06	-.05	-.12	1																
6. VHP	.05	-.00	-.04	-.08	.94**	1															
7. DJP	.04	-.02	-.07	-.10	.93**	.83**	1														
8. DRP	.06	-.13	-.03	-.16	.93**	.81**	.77**	1													
9. Driving	.12	.06	-.10	-.04	.53**	.50**	.55**	.48**	1												
10. VHG	.11	.07	-.10	.05	.71**	.74**	.60**	.65**	.55**	1											
11. DJG	.05	.19	-.13	-.05	.75**	.71**	.72**	.67**	.54**	.78**	1										
12. DRG	.19	.00	-.09	-.12	.73**	.70**	.58**	.75**	.55**	.77**	.76**	1									
13. AC	-.06	.19	-.09	.25*	.17	.16	.08	.22*	-.47**	.21*	.25**	.18	1								
14. ACC	-.08	.20	-.17	.12	.10	.09	.06	.13	-.40**	.17	.21*	.11	.84**	1							
15. ACR	-.15	.11	-.15	.13	.13	.12	.05	.20*	-.61**	.14	.18*	.14	.89**	.70**	1						
16. ACT	.11	-.02	.17	.15	.25**	.23*	.22*	.24**	-.02	.14	.16	.14	.53**	20*	.40**	1					
17. ACS	-.03	.18	-.05	.26*	.11	.14	.01	.17	-.29**	.21*	.24**	.18*	.83**	.57**	.59**	.42**	1				
18. PF	.06	-.19	-.25*	-.21	.39**	.29**	.43**	.38**	.45**	.18*	.32**	.34**	-.25**	-.20*	-.26**	-.07	-.20*	1			
19. PFO	-.07	-.16	-.19	-.15	.28**	.17	.34**	.25**	.48**	.00	.12	.14	-.45**	-.45**	-.42**	-.04	-.38**	.79**	1		
20. PFI	-.03	-.20	-.25*	-.25*	.28**	.23*	.33**	.23*	.44**	.15	.26**	.23*	-.23**	-.10	-.38**	-.18*	-.06	.73**	.45**	1	
21. PFU	.258*	-.01	-.08	-.03	.24**	.20*	.18	.29**	-.06	.24**	.28**	.33**	.26**	.22*	.35**	.10	.09	.48**	.076	-.01	1

n = 123

*indicates significance at p < .05 level, **indicates significance at the p < .01 level [+]indicates *n* = 68

APPENDIX B: SURVEYS AND SCALES

New Personal Fable Scale (Lapsley et al., 1989)

How well do the following statements describe you?

1 I believe I can do anything I set my mind to.
2 Nothing seems to really bother me.
3 No one has the same thoughts and feelings I have.
4 I think that I am more persuasive than my friends.
5 I believe that no one can stop me if I really want to do something.
6 I'm somehow different from everyone else.
7 It often seems like everything I do turns out great.
8 I don't think anything will stand in the way of my goals.
9 I'm the only one than can really understand me.
10 I believe that other people control my life.
11 I don't believe in taking chances.
12 I believe that I am unique.
13 I think that I can be anything I want to be.
14 I'm a fragile person.
15 I think that deep down everybody is the same.
16 I believe that everything I do is important.
17 I believe in knowing how something will turn out before I try it.
18 I'm just like everyone else.
19 I think I'm a powerful person.
20 I believe in taking risks.
21 Everybody goes through the same things that I am going through.
22 I think that I am better than my friends at just about anything.
23 I tend to doubt myself a lot.
24 It's hard for me to tell if I am different from my friends.
25 I often feel that I am insignificant and that I don't really matter.
26 Other people have no influence on me.
27 There isn't anything special about me.
28 I often think that people don't listen to what I have to say.
29 There are times when I think that I am indestructible.
30 I honestly think I can do things that no one else can.
31 I can get away with things that other people can't.
32 Everyone knows that I am a leader.
33 Nobody will ever know what I am really like.
34 No one sees the world the way that I do.
35 It is impossible for people to hurt my feelings.
36 People always do what I tell them to do.
37 People usually wait to hear my opinion before making a decision.

38 I usually let my friends decide what we are going to do.
39 My feelings are easily hurt.
40 The problems that some people get into could never happen to me.
41 I enjoy taking risks.
42 It is easy for me to take risks because I never get hurt or caught.
43 I don't take chances because I usually get in trouble.
44 I am always in control.
45 I am not afraid to do dangerous things.
46 Sometimes I think that no one really understands me.

Reverse-score: 10 23 28 38 11 14 17 39 43 15 18 21 24 25 27
Omnipotence: 1, 4 5 7 8 10 13 16 19 22 23 26 28 30 32 36 37 38 44
Invulnerability: 2 11 14 17 20 29 31 35 39 40 41 42 43 45
Personal Uniqueness: 3 6 9 12 15 18 21 24 25 27 33 34 46

Automation Complacency Scale

Complacency- trust in automation scale (Singh, Molloy, & Parasuraman, 1993)

Confidence:
1. I think that automated devices used in medicine, such as CT scans and ultrasound, provide very reliable medical diagnosis.
2. Automated devices in medicine save time and money in the diagnosis and treatment of disease.
3. If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because it is more reliable and safer than manual surgery.
4. Automated systems used in modern aircraft, such as the automatic landing system, have made air journeys safer.

Reliance:
1. ATMs provide a safeguard against the inappropriate use of an individual's bank account by dishonest people.
2. Automated devices used in aviation and banking have made work easier for both employees and customers.
3. Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass a police radar speed trap in case the automatic control is not working properly.

Trust:
1. Manually sorting through card catalogues is more reliable than computer-aided searches for finding items in a library.

2. I would rather purchase an item using a computer than have to deal with a sales representative on the phone because my order is more likely to be correct using the computer.
3. Bank transactions have become safer with the introduction of computer technology for the transfer of funds.

Safety:
1. I feel safer depositing my money at an ATM than with a human teller.

Confidence in Driving Skills

Driver abilities scale (Matthews & Moran, 1986)

Rate your:

Vehicle handling skills (the ability to maneuver a vehicle and control its path)
Driving judgement (the ability to make safe vehicle-handling decisions)
Driving reflexes (the speed at which you can react to important driving events)

1=very poor, 9 = excellent

Rate in comparison to your peers (people in your age group)

Vehicle handling skills (the ability to maneuver a vehicle and control its path)
Driving judgement (the ability to make safe vehicle-handling decisions)
Driving reflexes (the speed at which you can react to important driving events)

1= much worse than my peers, 9= much better than my peers

Fault Attribution and Perceived Avoidability Scales

To what extent was [Car 1/The Safety Suggestion Algorithm/The non-human controlled, driverless vehicles] at fault for making the incorrect driving choice?

To what extent were you, as the driver, at fault for making the incorrect driving choice?

How easily could the collision have been prevented, if different actions were taken by [Car 1/The Safety Suggestion Algorithm in your car/ The non-human controlled, driverless vehicles]?

How easily could the collision have been prevented, if different actions were taken by you, as the driver?

Script for Experiment Two

You are going to be playing a game called a "Choose your own Adventure." At each page, you will be given two options. You should try to pick the correct option to advance in the game. If you pick the wrong option, the game is over and you will be given a survey to fill out.

In this scenario, you are getting ready to drive to work at an office building in Orlando, Florida. Your car has a new technology called the Safety Suggestion Algorithm. This algorithm uses sensors to take in information about the surrounding road conditions, traffic, and weather, and makes a suggestion based on factors such as obstacles in the road or other vehicles. However, sometimes the Safety Suggestion Algorithm is incorrect. In these cases, it will make the wrong suggestion. Your job now is to get to work quickly and safely by agreeing or disagreeing with each of the Safety Suggestion Algorithm's recommendations. If you get all the way to work on time and safely, you win! If not, you lose and the game ends.

Let's try an example! Pick whichever answer you want, because this one doesn't count.

***

Practice round: You are getting ready to begin your commute to work. It is still dark outside. Will you turn your headlights on?

The Safety Suggestion Algorithm says: Yes, turn your headlights on.

Do you….. Agree/Disagree

Agree: Great job! Let's move to the next scenario. It's another practice round!

Disagree: Oops! Wrong answer. In this case, the Safety Suggestion Algorithm was correct. You should use your headlights when it's dark outside. Let's move on to another practice round!

***

Practice round: You are stopped at a red light, but no other cars are around. The coast seems clear and you don't want to be late. Do you run the red light?

The Safety Suggestion Algorithm says: Yes, run the light.

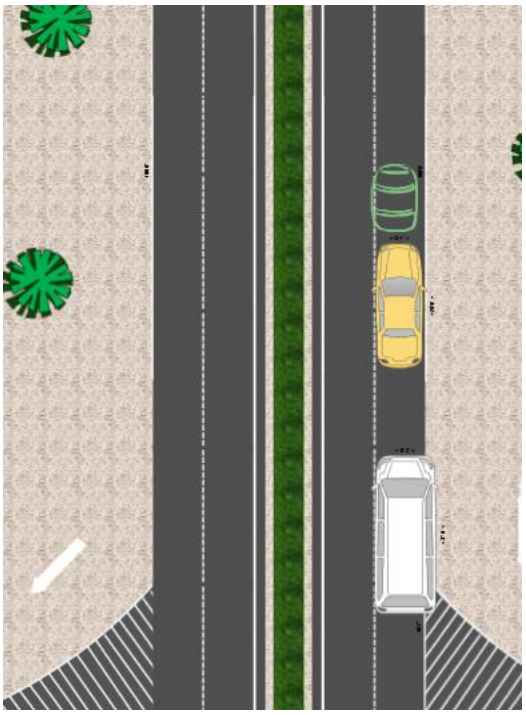
Do you….. Agree/Disagree

Agree: Oops! The Safety Suggestion Algorithm was wrong here. You should not run red lights. Let's move on now to the real game!

Disagree: Great job! The Safety Suggestion Algorithm was wrong here. You should not run red lights. Let's move on now to the real game!

***

Round 1: You are driving down the road when you come across a barrel in the road! You cannot safely continue your course. There is a large vehicle following you very closely at a high speed. You can either slam on the brakes, or swerve to avoid the obstacle. See the image and video below. In this image, your car is the *yellow* car.



The Safety Suggestion Algorithm says: Swerve to avoid the obstacle.

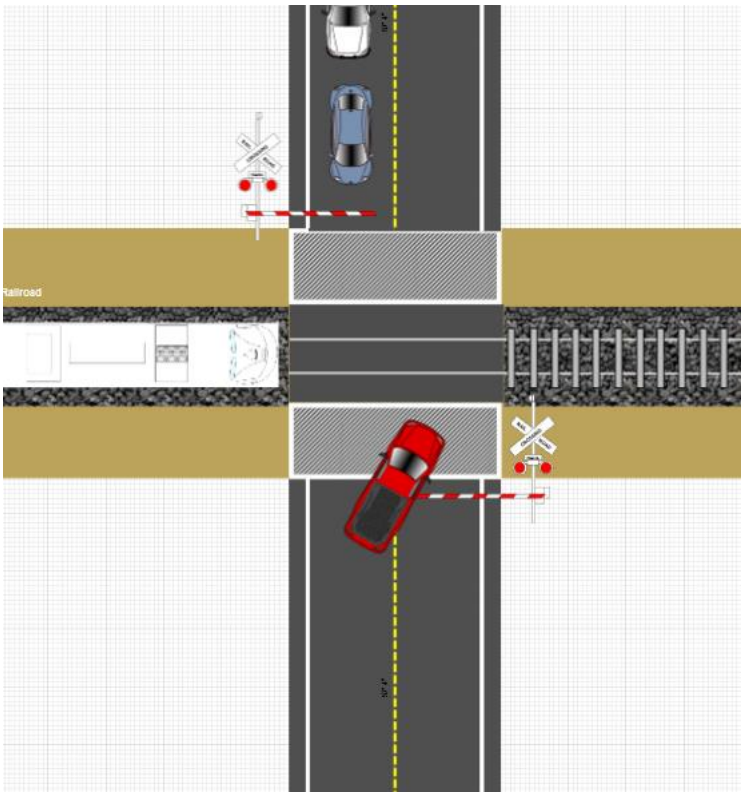
Do you….. Agree/Disagree

Agree: Great job! The vehicle behind you would not have stopped in time. Move on to the next round.

Disagree: Oops! The vehicle behind you could not stop in time. You have lost. Move on to the final survey.

***

Round 2: You come to the train tracks. The crossing arm is lowered, but the train is not there yet. You might be late if you wait for the train to pass. You can wait for the train, or you can try to cross the tracks before it arrives.



The Safety Suggestion Algorithm says: Wait for the train to pass

Do you….. Agree/Disagree

Agree: Great job! You should not cross train tracks while the train is approaching. Move on to the next round.

Disagree: Oops! You should not cross train tracks while the train is approaching. You have lost. Move on to the final survey.

***

Round 3: You are on a two lane road, and there is a bicycle traveling slowly in your lane. You are separated from oncoming traffic by a dashed line in the road, not a solid line, so you know it is legal to overtake by moving into the lane of oncoming traffic. You do not see any other cars coming. Do you overtake the bicycle, or wait? See the video below.

The Safety Suggestion Algorithm says: Wait. Do not pass a bicycle.

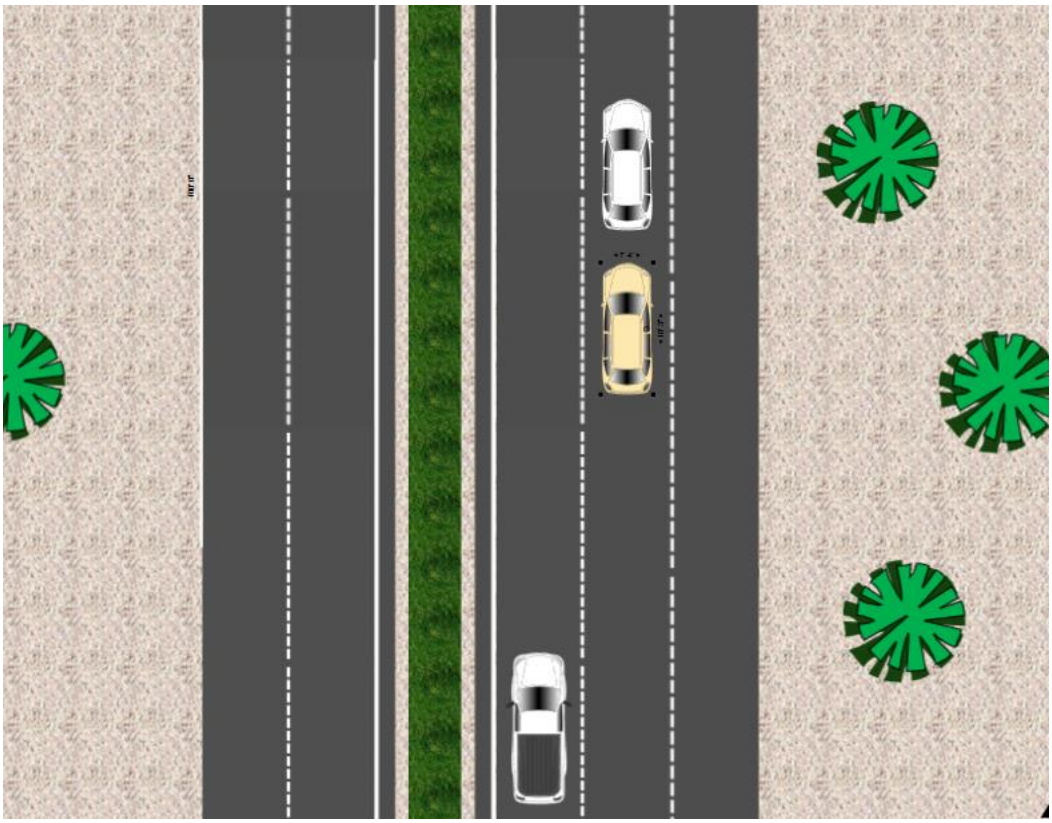
Do you….. Agree/Disagree

Agree: Oops! You are allowed to pass a bicycle if it is safe. You have lost. Move on to the final survey.

Disagree: Great job! You are allowed to pass a bicycle if it is safe. Move on to the next round.

***

Round 4: You are on the highway, and are driving in the middle lane behind a very slow car. You know you need to overtake or you will be late to work. Do you move into the left-hand lane and overtake the slow car, or move into the right-hand lane and overtake the car? See the image below. In this image, you are in the *yellow* car.



The Safety Suggestion Algorithm says: Move into the left land.

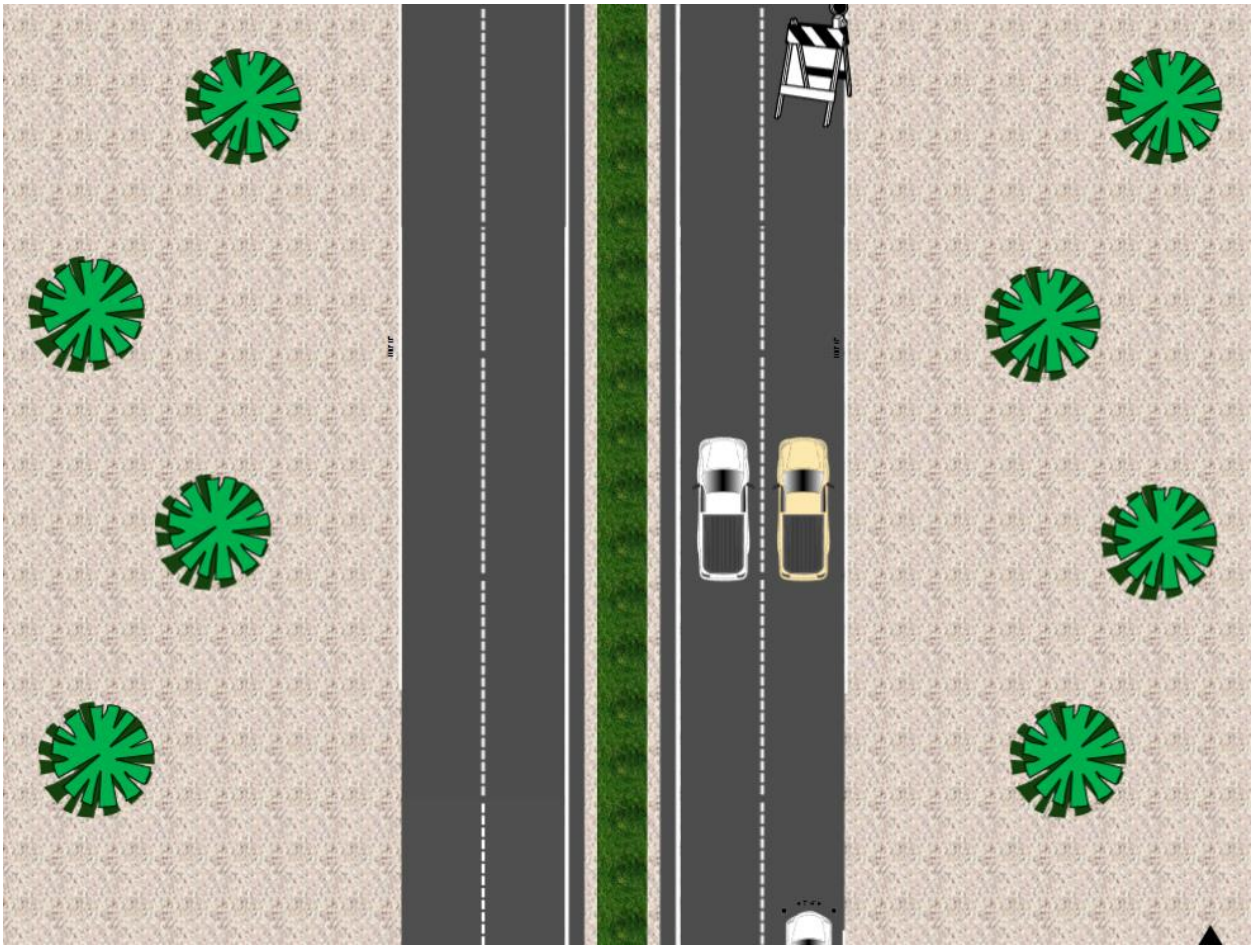
Do you….. Agree/Disagree

Agree: Great job! You should always pass on the left. Move on to the next round.

Disagree: Oops! You should always pass on the left. You have lost. Move on to the final survey.

***

Round 5: You are still driving on the highway. Now you are in the right lane. You notice that up ahead, your lane ends. You will need to move into the left lane. However, there is a car directly next to you, blocking you from moving. Do you speed up to merge in front of them, or slow down to merge behind them? See the image below. In this image you are driving the *yellow* car.



The Safety Suggestion Algorithm says: Slow down and merge behind the other car

Do you….. Agree/Disagree

Agree: Great job! People in the right-hand lane should allow those in the left-hand lane to pass. Move on to the next round.

Disagree: Oops! People in the right-hand lane should allow those in the left-hand lane to pass. You have lost. Move on to the final survey.

***

Round 6: It is raining very hard. You know that in the heavy rain, your car might not be visible to other drivers. See the video below. Will you put on your hazard lights to be more visible, or will you just use the headlights and windscreen wipers?

The Safety Suggestion Algorithm says: Put on your hazard lights.

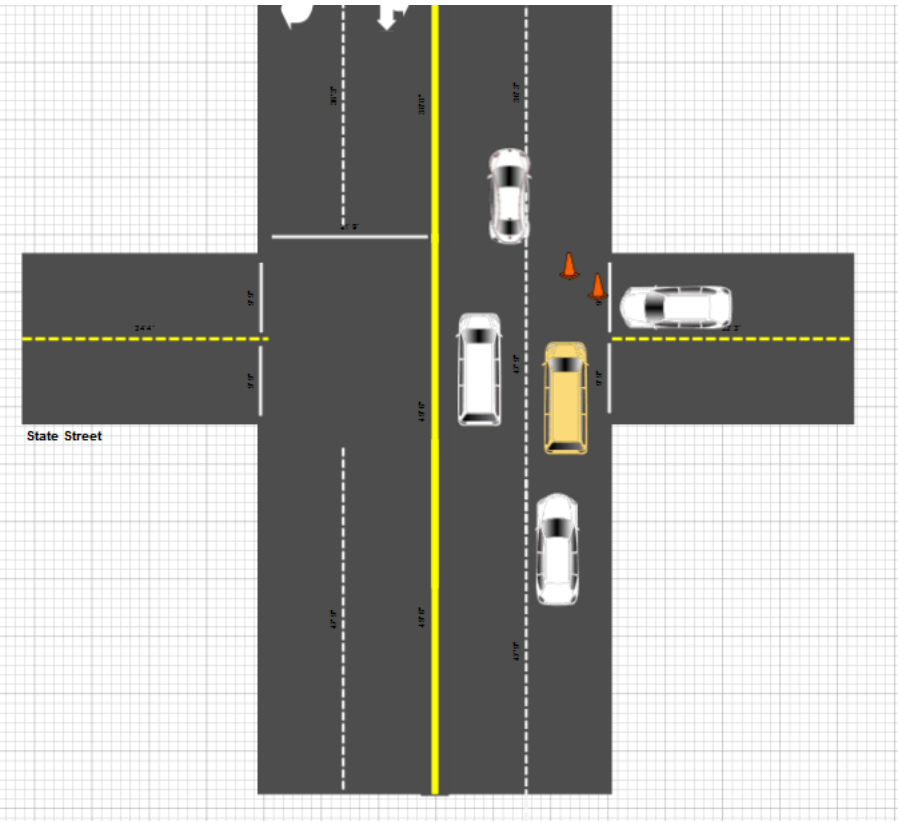
Do you….. Agree/Disagree

Agree: Oops! Even though it makes you easier to see, using hazard lights while driving is actually illegal in Florida. You have lost. Move on to the final survey.

Disagree: Great job! Even though it makes you easier to see, using hazard lights while driving is actually illegal in Florida. Move on to the next round.

***

Round 7: You are driving along a crowded stretch of the highway when you notice that two traffic cones have been left behind by construction workers who had finished their work. There are cars in the other lane and a car following closely behind you. You can either slam on the brakes, or run over the traffic cone. See image below. In this image, you are in the *yellow* car.



The Safety Suggestion Algorithm says: Run over the traffic cone.

Do you….. Agree/Disagree

Agree: Great job! Traffic cones can be run over, and doing so is safer than causing a collision. Move on to the next round.

Disagree: Oops! Traffic cones can be run over, and doing so is safer than causing a collision. You have lost. Move on to the final survey.

***

Round 8: You come to a construction zone and a sign that tells you the speed limit is 45 miles per hour when workers are present. See the image below. The normal speed limit on this stretch of road is 65. You do not see any workers. Do you keep going 65 miles per hour, or do you slow to 45 mph?

The Safety Suggestion Algorithm says: Slow down to 45.

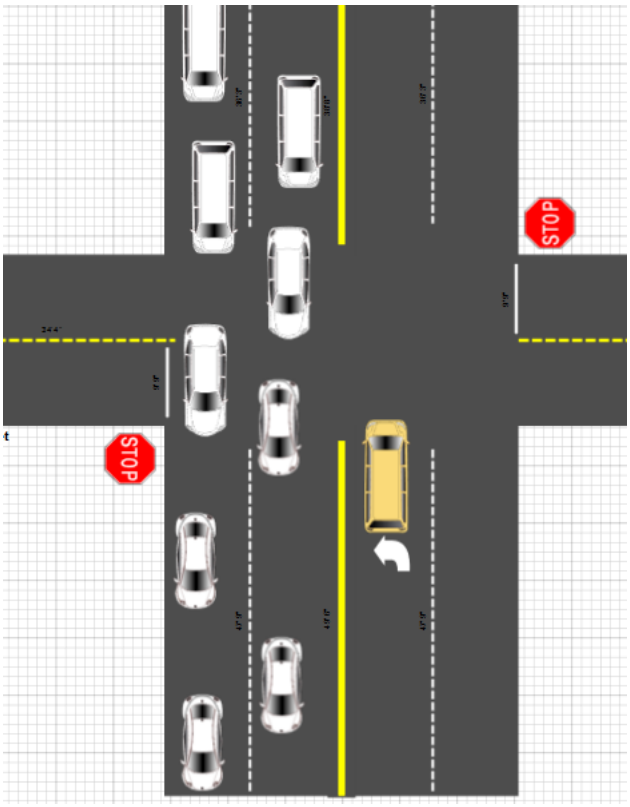
Do you….. Agree/Disagree

Agree: Oops! The reduced speed limit only applies when workers are present. You should always follow the correct speed limit- don't go to fast *or* too slow. You have lost. Move on to the final survey.

Disagree: Great job! The reduced speed limit only applies when workers are present. You should always follow the correct speed limit- don't go to fast *or* too slow. Move on to the next round.

***

Round 9: You are almost at work. You need to turn left into the parking lot, but in order to do so, you must cut across traffic. There are a lot of cars present. See image below. In this image, you are in the *yellow* car. You can either wait for a gap and make an unprotected left turn into the parking lot, or drive further down the road and use a traffic light to make a legal U-turn.



The Safety Suggestion Algorithm says: Go to the light and make a U-tun.

Do you….. Agree/Disagree

Agree: Great job! It is always safer to use a traffic light to turn, when you have the option. Move on.

Disagree: Oops! It is always safer to use a traffic light to turn, when you have the option. You have lost. Move on to the final survey.

***

Congratulations! You have safely made it to work.

APPENDIX C: INSTITUTIONAL REVIEW BOARD APPROVAL

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

EXEMPTION DETERMINATION

June 11, 2020

Dear Alexandra Kaplan:

On 6/11/2020, the IRB determined the following submission to be human subjects research that is exempt from regulation:

Type of Review:	Initial Study, Category 3(i)(A)
Title:	Driver Decision Making
Investigator:	Alexandra Kaplan
IRB ID:	STUDY00001869
Funding:	None
Grant ID:	None
Documents Reviewed:	• faculty advisor review, Category: Faculty Research Approval; • blame survey, Category: Survey / Questionnaire; • demographics and driving, Category: Survey / Questionnaire; • Explanation of Research, Category: Consent Form; • personal fable scale, Category: Survey / Questionnaire; • protocol, Category: IRB Protocol; • recruitment info, Category: Recruitment Materials; • rules and narrative, Category: Test Instruments; • Trust in automation scale, Category: Survey / Questionnaire;

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made, and there are questions about whether these changes affect the exempt status of the human research, please submit a modification request to the IRB. Guidance on submitting Modifications and Administrative Check-in are detailed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system. When you have completed your research, please submit a Study Closure request so that IRB records will be accurate.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Racine Jacques, Ph.D.
Designated Reviewer

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

EXEMPTION DETERMINATION

May 20, 2020

Dear Alexandra Kaplan:

On 5/20/2020, the IRB determined the following submission to be human subjects research that is exempt from regulation:

Type of Review:	Initial Study
Title:	Fault Attribution in Vehicular Collisions
Investigator:	Alexandra Kaplan
IRB ID:	STUDY00001771
Funding:	None
Grant ID:	None
Documents Reviewed:	• advisor form, Category: Faculty Research Approval; • debrief, Category: Consent Form; • demographics scale, Category: Survey / Questionnaire; • driving skills scale, Category: Survey / Questionnaire; • Explanation of Research, Category: Consent Form; • personal fable scale, Category: Survey / Questionnaire; • Protocol May 4 HRP 255 5/20, Category: IRB Protocol; • questions in each scenario, Category: Survey / Questionnaire; • recruitment , Category: Recruitment Materials; • rules, Category: Survey / Questionnaire; • Trust in automation scale, Category: Survey / Questionnaire; • Warning before the videos, Category: Test Instruments;

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made, and there are questions about whether these changes affect the exempt status of the human research, please submit a modification request to the IRB. Guidance on submitting Modifications and Administrative Check-in are detailed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system. When you have completed your research, please submit a Study Closure request so that IRB records will be accurate.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Due to current COVID-19 restrictions, in-person research is not permitted to begin until you receive further correspondence from the Office of Research stating that the restrictions have been lifted.

Sincerely,

Kamille C. Birkbeck

Kamille Birkbeck
Designated Reviewer

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

APPROVAL

June 10, 2020

Dear Alexandra Kaplan:

On 6/10/2020, the IRB reviewed the following submission:

Type of Review:	Initial Study
Title:	Driving with Semi-Autonomous Vehicles
Investigator:	Alexandra Kaplan
IRB ID:	STUDY00001846
Funding:	None
Grant ID:	None
IND, IDE, or HDE:	None
Documents Reviewed:	• HRP-251 - FORM - Faculty Advisor Review.pdf, Category: Faculty Research Approval; • blame survey in case of collision, Category: Survey / Questionnaire; • complacency trust in automation , Category: Survey / Questionnaire; • consent, Category: Consent Form; • demographics and driving, Category: Survey / Questionnaire; • driving skills compared to peers, Category: Survey / Questionnaire; • new personal fable scale, Category: Survey / Questionnaire; • protocol, Category: IRB Protocol; • recruitment, Category: Recruitment Materials;

The IRB approved the protocol on 6/10/2020.

Due to current COVID-19 restrictions, in-person research is not permitted to begin until you receive further correspondence from the Office of Research stating that the restrictions have been lifted.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system. Guidance on submitting Modifications and a Continuing Review or Administrative Check-in are detailed in the manual. When

you have completed your research, please submit a Study Closure request so that IRB records will be accurate.
If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Racine Jacques, Ph.D.
Designated Reviewer

APPENDIX D: ALL MAIN EFFECTS AND TWO-WAY INTERACTIONS IN THE REGRESSION MODELS

Experiment One: Overall Fault Attribution Model

Variable	b	se	B
Auto	.31	.51	.10
Avoidability	1.13	.16	1.05**
Driving-Peers	-.25	.24	-.23
Driving-General	.24	.27	.19
Comp	.14	.08	.25^{+}
PF	.35	.17	.26*
Auto*Avoidability	-.04	.04	-.05
Auto* Driving-Peers	-.13	.05	-.30**
Auto*Driving-General	.08	.06	.18
Auto*PF	.08	.05	.22
Auto*Comp	-.03	.02	-.14^{+}
Avoidability*Driving-Peers	-.02	.02	-.12
Avoidability*Driving-General	.01	.02	.09
Avoidability*PF	-.02	.02	-.14
Avoidability*Comp	-.01	.01	-.14
Driving-Peers*General	.02	.01	.25*
Driving-Peers*PF	.02	.02	.22
Driving-Peers*Comp	-.00	.01	-.01
Driving-General*PF	-.05	.02	-.50*
Driving-General*Comp	.00	.01	.01
PF*Comp	-.01	.01	-.19

R^2 = .592; Adjusted R^2 = .586

**indicates significance at $p < .01$; *indicates significance at $p < .05$; $^{+}$indicates significance at $p < .10$

Experiment One: Fault of Autonomous Vehicles

Variable	b	se	B
Avoidability	1.35	.24	1.26**
Driving-Peers	-.15	.36	-.14
Driving-General	.26	.39	.21
Comp	.02	.12	.04
PF	.44	.26	.32^{+}
Avoidability*Driving-Peers	-.03	.03	-.23
Avoidability*Driving-General	.02	.03	.16

Avoidability*PF	-.04	.02	-.40^{+}
Avoidability*Comp	-.01	.01	-.12
Driving-Peers*General	.01	.02	.06
Driving-Peers*PF	.01	.03	.16
Driving-Peers*Comp	-.00	.01	-.02
Driving-General*PF	-.04	.04	-.42
Driving-General*Comp	.00	.02	.08
PF*Comp	-.00	.01	-.08

R^2 = .562; Adjusted R^2 = .554

**indicates significance at $p < .01$; *indicates significance at $p < .05$; $^{+}$indicates significance at $p < .10$

Experiment One: Fault in Regular Vehicles

Variable	b	se	B
Avoidability	.89	.22	.84**
Driving-Peers	-.54	.32	-.51^{+}
Driving-General	.34	.36	.26
Comp	.26	.11	.48*
PF	.34	.24	.25
Avoidability*Driving-Peers	-.00	.02	-.01
Avoidability*Driving-General	.00	.03	.03
Avoidability*PF	.01	.02	.07
Avoidability*Comp	-.01	.01	-.17
Driving-Peers*General	.04	.02	.47**
Driving-Peers*PF	.03	.03	.31
Driving-Peers*Comp	.00	.01	.00
Driving-General*PF	-.06	.03	-.58^{+}
Driving-General*Comp	-.01	.02	-.15
PF*Comp	-.02	.01	-.32

R^2 = .624; Adjusted R^2 = .617

**indicates significance at $p < .01$; *indicates significance at $p < .05$; $^{+}$indicates significance at $p < .10$

Experiment One: Fault in Algorithms

Variable	b	se	B
Avoidability	-.30	.39	-.29
Driving-Peers	-.85	.50	-.92^{+}

Driving-General	.87	.57	.76
Comp	-.06	.17	-.12
PF	.35	.37	.30
Avoidability*Driving-Peers	-.03	.04	-.27
Avoidability*Driving-General	.08	.04	.70^{+}
Avoidability*PF	-.05	.04	-.51
Avoidability*Comp	.01	.01	.20
Driving-Peers*General	-.02	.02	-.24
Driving-Peers*PF	.04	.04	.47
Driving-Peers*Comp	.06	.02	1.26**
Driving-General*PF	-.05	.05	-.62
Driving-General*Comp	-.05	.02	-1.12*
PF*Comp	-.00	.01	-.06

R^2 = .083; Adjusted R^2 = .063

**indicates significance at $p < .01$; *indicates significance at $p < .05$; $^{+}$indicates significance at $p < .10$

Experiment Two: Fault in Vehicle

Variable	b	se	B
Avoidability-SSA	-1.08	1.01	-1.16
Avoidability-Driver	1.14	.83	1.26
SSA Wrong	10.88	4.49	2.11*
Driving-General	-4.97	3.05	-2.42
Driving-Peers	5.24	2.63	2.98*
Comp	-2.67	4.40	-.63
PF	-10.14	5.71	-1.29^{+}
Avoidability-SSA*Driver	-.01	.02	-.12
Avoidability-SSA*SSA Wrong	-.08	.14	-.12
Avoidability-SSA*Driving-General	.04	.08	.37
Avoidability-SSA*Driving-Peers	-.07	.07	-.58
Avoidability-SSA*PF	.39	.28	1.3
Avoidability-SSA*Comp	.10	.13	.38
Avoidability-Driver*SSA Wrong	.10	.14	.16
Avoidability-Driver*Driving-General	-.03	.08	-.32
Avoidability-Driver*Driving-Peers	-.06	.06	-.51
Avoidability-Driver*PF	-.20	.22	-.69
Avoidability-Driver*Comp	-.06	.12	-.22
SSA Wrong*Driving-General	-.59	.45	-.91
SSA Wrong*Driving-Peers	-.37	.38	-.54

SSA Wrong*PF	-.65	1.19	-.39
SSA Wrong*Comp	-.56	.66	-.35
Driving-General*Peers	-.00	.07	-.02
Driving-General*PF	1.44	.81	2.82^{+}
Driving-General*Comp	.26	.42	.62
Driving-Peers*Comp	-.51	.35	-1.33
Driving-Peers*PF	-.80	.72	-1.75
Comp*PF	1.49	1.08	1.26

R^2 = .274; Adjusted R^2 = .145

**indicates significance at $p < .01$; *indicates significance at $p < .05$; $^{+}$indicates significance at $p < .10$

Experiment Two: Fault in Self

Variable	b	se	B
Avoidability-SSA	1.30	1.30	1.09
Avoidability-Driver	-1.31	1.07	-1.13
SSA Wrong	1.13	5.79	.171
Driving-General	5.04	3.93	1.91
Driving-Peers	-2.40	3.39	1.06
Comp	-4.93	5.67	-.90
PF	17.52	7.35	1.73*
Avoidability-SSA*Driver	-.07	.03	-.62*
Avoidability-SSA*SSA Wrong	.34	.18	.39^{+}
Avoidability-SSA*Driving-General	-.06	.10	-.43
Avoidability-SSA*Driving-Peers	.04	.09	.27
Avoidability-SSA*PF	-.09	.37	-.25
Avoidability-SSA*Comp	-.27	.16	-.76
Avoidability-Driver*SSA Wrong	-.16	.18	-.19
Avoidability-Driver*Driving-General	-.11	.11	-.77
Avoidability-Driver*Driving-Peers	.06	.08	.44
Avoidability-Driver*PF	.22	.29	.61
Avoidability-Driver*Comp	.50	.16	1.48**
SSA Wrong*Driving-General	.21	.58	.26
SSA Wrong*Driving-Peers	.80	.49	.91
SSA Wrong*PF	-3.13	1.53	-1.47*
SSA Wrong*Comp	.25	.84	.122
Driving-General*Peers	-.05	.09	-.26
Driving-General*PF	-1.71	1.04	-2.61

Driving-General*Comp	.35	.53	.65
Driving-Peers*Comp	.44	.45	.88
Driving-Peers*PF	.11	.92	1.90
Comp*PF	-1.15	1.39	-.75

R^2 = .270; Adjusted R^2 = .140

**indicates significance at $p < .01$; *indicates significance at $p < .05$; $^{+}$indicates significance at $p < .10$

Experiment Three: Fault in Vehicles

Variable	b	se	B
Avoidability-Vehicles	.89	1.54	.87
Avoidability-Driver	-2.56	1.96	-2.33
Driving-General	-2.57	5.19	-1.12
Driving-Peers	1.91	4.75	.95
PF	11.24	12.14	1.13
Comp	14.92	7.23	3.34*
Avoidability-Vehicles*Driver	-.01	.06	-.10
Avoidability-Vehicles*Driving-General	.27	.18	2.11
Avoidability-Vehicles*Driving-Peers	.10	.15	.72
Avoidability-Vehicles*PF	-.38	.37	-1.13
Avoidability-Vehicles*Comp	-.78	.19	-2.55**
Avoidability- Driver*Driving-General	-.07	.16	-.52
Avoidability- Driver*Driving-Peers	.01	.15	.05
Avoidability- Driver*PF	.90	.46	2.46^{+}
Avoidability- Driver*Comp	.045	.28	.15
Driving-General*Peers	-.38	.22	-2.29^{+}
Driving-General*PF	1.22	1.38	2.11
Driving-General*Comp	.12	.71	.26
Driving-Peers*PF	-.68	1.3	-1.29
Driving-Peers*Comp	.71	.48	1.65
Comp*PF	-5.52	2.35	-3.85*

R^2 = .599; Adjusted R^2 = .415

**indicates significance at $p < .01$; *indicates significance at $p < .05$; $^{+}$indicates significance at $p < .10$

Experiment Three: Fault in Self

Variable	b	se	B
Avoidability-Vehicles	-.32	1.39	-.38
Avoidability-Driver	-.30	1.78	-.32
Driving-General	.78	4.70	.41
Driving-Peers	1.80	4.30	1.07
PF	4.05	10.99	.49
Comp	-7.44	6.54	-1.99
Avoidability-Vehicles*Driver	-.06	.05	-.63
Avoidability-Vehicles*Driving-General	.04	.16	.35
Avoidability-Vehicles*Driving-Peers	-.01	.13	-.13
Avoidability-Vehicles*PF	-.09	.33	-.32
Avoidability-Vehicles*Comp	.26	.17	1.00
Avoidability- Driver*Driving-General	.45	.14	3.83**
Avoidability- Driver*Driving-Peers	-.21	.13	-1.68
Avoidability- Driver*PF	.33	.41	-1.07
Avoidability- Driver*Comp	-.02	.25	-.08
Driving-General*Peers	.45	.20	3.22*
Driving-General*PF	-1.16	1.25	-2.39
Driving-General*Comp	-.68	.64	-1.84
Driving-Peers*PF	-1.07	1.19	-2.42
Driving-Peers*Comp	-.37	.44	-1.01
Comp*PF	4.74	2.13	3.96*

R^2 = .530; Adjusted R^2 = .315

**indicates significance at $p < .01$; *indicates significance at $p < .05$; $^{+}$indicates significance at $p < .10$

APPENDIX E: TABLE OF DESCRIPTIVE STATISTICS

Experiment	*n*	Variable	Mean	*sd*
1	266	Fault Attribution	3.60	1.59
1	266	Fault in Autonomous Vehicles	3.67	1.57
1	266	Fault in Regular Vehicles	3.54	1.61
1	266	Collision Avoidability	3.86	1.49
1	266	Collision Avoidability for Autonomous Vehicles	3.88	1.46
1	266	Collision Avoidability for Regular Vehicles	3.84	1.52
1	266	Driving Abilities Overall	7.55	1.26
1	266	Vehicle Handling Skills Overall	7.57	1.26
1	266	Driving Judgement Overall	7.52	1.46
1	266	Driving Reflexes Overall	7.55	1.52
1	266	Driving Abilities Compared to Peers	7.06	1.22
1	266	Vehicle Handling Skills Compared to Peers	7.05	1.56
1	266	Driving Judgement Compared to Peers	7.06	1.68
1	266	Driving Reflexes Compared to Peers	7.07	1.67
1	266	Omnipotence	3.03	0.51
1	266	Invulnerability	2.89	0.50
1	266	Uniqueness	3.40	0.49
1	266	Automation Complacency Overall	3.31	0.92
1	266	Confidence in Automation	3.59	1.06
1	266	Reliance on Automation	3.32	0.84
1	266	Trust in Automation	3.18	0.77
1	266	Perceived Safety of Automation	3.15	1.01
2	183	Fault in the SSA	3.75	2.57
2	183	Fault in Self	5.21	3.32
2	183	Avoidability by SSA	7.04	2.77
2	183	Avoidability by Self	6.47	2.85
2	188	Driving Abilities Overall	7.75	1.26
2	188	Vehicle Handling Skills Overall	7.86	1.35
2	188	Driving Judgement Overall	7.70	1.41
2	188	Driving Reflexes Overall	7.69	1.38
2	188	Driving Abilities Compared to Peers	7.25	1.47
2	188	Vehicle Handling Skills Compared to Peers	7.23	1.53
2	188	Driving Judgement Compared to Peers	7.32	1.60
2	188	Driving Reflexes Compared to Peers	7.21	1.64
2	188	Omnipotence	3.01	0.47
2	188	Invulnerability	2.90	0.58
2	188	Uniqueness	3.34	0.45
2	188	Automation Complacency Overall	3.07	0.61
2	188	Confidence in Automation	3.33	1.02
2	188	Reliance on Automation	3.12	0.79
2	188	Trust in Automation	3.03	0.68

2	188	Perceived Safety of Automation	2.79	1.14
3	68	Fault in Self	5.01	2.62
3	68	Fault in Vehicles	5.04	3.14
3	68	Avoidability by Self	4.94	2.85
3	68	Avoidability by Other Vehicles	5.82	3.07
3	123	Driving Abilities Overall	7.53	1.72
3	123	Vehicle Handling Skills Overall	7.62	1.38
3	123	Driving Judgement Overall	7.46	1.37
3	123	Driving Reflexes Overall	7.51	1.46
3	123	Driving Abilities Compared to Peers	7.05	1.49
3	123	Vehicle Handling Skills Compared to Peers	7.11	1.55
3	123	Driving Judgement Compared to Peers	7.02	1.58
3	123	Driving Reflexes Compared to Peers	7.03	1.68
3	123	Omnipotence	2.95	0.59
3	123	Invulnerability	2.92	0.55
3	123	Uniqueness	3.34	0.49
3	123	Automation Complacency Overall	3.16	1.30
3	123	Confidence in Automation	3.53	1.82
3	123	Reliance on Automation	3.23	1.99
3	123	Trust in Automation	2.88	0.86
3	123	Perceived Safety of Automation	2.93	1.74

LIST OF REFERENCES

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49* (6), 1621-1630.

Alicke, M. D. (2000). Culpable control and the psychology of fault. *Psychological Bulletin, 126* (4), 556-574.

Alicke, M., & Govorlin, O. (2005). The better-than-average effect. In M.D. Alicke, D.A. Dunning, and J. Kreuger (Eds): *The Self in Social Judgement* (pp. 85-106). New York: Psychology Press.

Bigman, Y. E., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in Cognitive Sciences, 23* (5), 365-368.

Bucher, R. (1957). Fault and hostility in disaster. *American Journal of Sociology, 62* (5), 467-475.

Burger, J. M. (1981). Motivational biases in the attribution of responsibility for an accident: A meta-analysis of the defensive-attribution hypothesis. *Psychological Bulletin, 90* (3), 496-512.

Chodoff, P., Friedman, S. B., & Hamburg, D. A. (1964). Stress, defenses and coping behavior: Observations in parents of children with malignant disease. *American Journal of Psychiatry, 120* (8), 743-749.

Elkind D (1967) Egocentrism in adolescence. *Child Development, 38* (4),1025-1034.

Ethics Commission Report. (2017). Ethics Commission: Automated and Connected Driving. https://www.bmvi.de/SharedDocs/EN/Documents/G/ethic-commission-report.pdf?__blob=publicationFile

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39* (2), 175–191.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, 5, 5-15.

Giladi, E. E., & Klar, Y. (2002). When standards are wide of the mark: Nonselective superiority and inferiority biases in comparative judgments of objects and concepts. *Journal of Experimental Psychology: General, 131* (4), 538-551.

Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26-34.

Green, J. (2018). "Tesla: Autopilot was on during deadly Mountain View crash". *The Mercury News.* Palo Alto. ISSN 0747-2099. OCLC 145122249. Retrieved November 13, 2019.

Hancock, P. A. (2009). *Mind, machine and morality: Toward a philosophy of human-technology symbiosis.* Boke Raton, FL: CRC Press.

Hancock, P. A. (2013). In search of vigilance: the problem of iatrogenically created psychological phenomena. *American Psychologist, 68* (2), 97-109.

Hancock, P. A. (2018). Are autonomous cars really safer than human drivers. *The Conversation*.

Hancock, P.A. (2019) Some pitfalls in the promises of automated and autonomous vehicles, *Ergonomics*, *62* (4), 479-495, DOI: 10.1080/00140139.2018.1498136.

Hancock, P. A. (2019). On the dynamics of conspicuity. *Human Factors*, *61* (6), 857-865.

Hancock, P. A. (2019). Science in court. *Theoretical Issues in Ergonomics Science*, 1-19.

Hancock, P. A., & De Ridder, S. N. (2003). Behavioural accident avoidance science: understanding response in collision incipient conditions. *Ergonomics*, *46* (12), 1111-1135.

Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2020). Evolving trust in robots: Specification through sequential and comparative meta-analyses. *Human Factors*, 0018720820922080.

Hawkins, J. (2019). Tesla's Smart Summon feature is already causing chaos in parking lots across America. *The Verge*. New York, NY. Retrieved February 19, 2020 from https://www.theverge.com/2019/9/30/20891343/tesla-smart-summon-feature-videos-parking-accidents.

Heider, F. (1958). *The Psychology of Interpersonal Relations.* New York: Wiley.

Horwitz, J., & Timmons, H. (2016). *The scary similarities between Tesla's (TSLA) deadly autopilot crashes*. Quartz. Atlantic Media. Retrieved November 13, 2019.

Kaplan, A. D., Sanders, T., & Hancock, P. A. (2019). The Relationship Between Extroversion and the Tendency to Anthropomorphize Robots: A Bayesian Analysis. *Frontiers in Robotics and AI, 5*, 135-145.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28* (2), 107-128.

Kim, T., & Hinds, P. (2006, September). Who should I fault? Effects of autonomy and transparency on attributions in human-robot interaction. In ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication (pp. 80-85). IEEE.

Kushner, H. S. (1981). *When bad things happen to good people*. New York: Schocken.

Lapsley, D. K., FitzGerald, D. P., Rice, K. G., & Jackson, S. (1989). Separation-individuation and the" new look" at the imaginary audience and personal fable: A test of an integrative model. *Journal of Adolescent Research, 4* (4), 483-505.

Lerner, M. (1980). *The belief in a just world: A fundamental delusion.* New York: Plenum Press.

Lewin, K. (1936). A dynamic theory of personality: Selected papers. *The Journal of Nervous and Mental Disease*, *84* (5), 612-613.

Lubben, A. (2018). "Self-driving Uber killed a pedestrian as human safety driver watched". Vice News. Vice Media. Retrieved November 13, 2019.

MacRae, F. (2016). Motorist becomes first to die in driverless car crash as he watches Harry Potter film. *Daily Mail*, 29.

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015, March). Sacrifice one for the good of many?: People apply different moral norms to human and robot agents. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 117-124).* ACM.

Matthews, M. L., & Moran, A. R. (1986). Age differences in male drivers' perception of accident risk: The role of perceived driving ability. *Accident Analysis & Prevention, 18* (4), 299-313.

National Highway Traffic Safety Administration (2013). *Preliminary Statement of Policy Concerning Automated Vehicles.* NHTSA.gov. Retrieved 11/13/2019.

Nyholm, S., & Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem?. *Ethical Theory and Moral Practice*, *19* (5), 1275-1289.

O'Kane, S. (2019). Tesla hit with another lawsuit over a fatal Autopilot crash. *The Verge.* New York, NY. Retrieved February 19, 2019 from https://www.theverge.com/2019/8/1/20750715/tesla-autopilot-crash-lawsuit-wrongful-death.

Roy, M. M., & Liersch, M. J. (2013). I am a better driver than you think: examining self-enhancement for driving ability. *Journal of Applied Social Psychology, 43* (8), 1648-1659.

SAE International. (2016). *Taxonomy and definitions for terms related to driving automation systems for On-Road motor vehicles*. Warrendale, PA: SAE International.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced" complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology, 3* (2), 111-122.

Stanton, N.A., Eriksson, A., Banks, V.A., & Hancock, P.A. (2020). Turing in the driver's seat: Can people distinguish between automated and manually driven vehicles? *Human Factors and Ergonomics in Manufacturing & Service Industries,* https://doi.org/10.1002/hfm.20864

Thomson, J.J. (1985) The trolley problem. *Yale Law Journal, 94* (5), 1395–1415

van der Woerdt, S., & Haselager, P. (2017). When robots appear to have a mind: the human perception of machine agency and responsibility. *New Ideas in Psychology, 54,* 93-100.

Veltfort, H. R., & Lee, G. E. (1943). The Cocoanut Grove fire: a study in scapegoating. *The Journal of Abnormal and Social Psychology, 38* (2S), 138-154.

Walster, E. (1966). Assignment of responsibility for an accident. *Journal of Personality and Social Psychology, 3 (*1), 73-79.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct.* New York: Guilford Press.

Wilson, H., Theodorou, A., & Bryson, J. J. (2019, August). Slam the brakes: perceptions of moral decisions in driving dilemmas. *International Workshop in Artificial Intelligence Safety (AISafety),* (Macau).

Wohleber, R. W., & Matthews, G. (2016). Multiple facets of overconfidence: Implications for driving safety. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 265-278.