Article:

Madigan, R orcid.org/0000-0002-9737-8012, Lee, YM orcid.org/0000-0003-3601-4191 and Merat, N orcid.org/0000-0003-4140-9948 (2021) Validating a methodology for understanding pedestrian – vehicle interactions: A comparison of video and field observations. Transportation Research Part F: Traffic Psychology and Behaviour, 81. pp. 101-114. ISSN 1369-8478

https://doi.org/10.1016/j.trf.2021.05.006

Validating a Methodology for Understanding Pedestrian – Vehicle Interactions: A Comparison of Video and Field Observations

Ruth Madigan[a], Yee Mun Lee[a], Natasha Merat[a]

[a] *Institute for Transport Studies, University of Leeds, LS2 9JT, United Kingdom*

Abstract

The successful deployment of automated vehicles (AVs) will depend on their capacity to travel within a mixed traffic environment, adopting appropriate interaction strategies across different scenarios. Thus, it is important to gain a detailed understanding of the specific types of interactions that are most likely to arise. The overall purpose of this paper was to present a methodology designed to facilitate the systematic observation of pedestrian-vehicle interactions, and to validate its use for both onsite and video based observations. A detailed observation protocol was developed to capture pedestrian and vehicle movement and communication patterns across four interaction phases. Onsite coders completed field observations of 50 pedestrian-vehicle interactions at a UK intersection, while video coders observed the same interactions recorded through a wireless camera mounted on a nearby rooftop. Results show that the observation protocol provides a reliable methodology for capturing patterns of pedestrian-vehicle interactions, with high levels of inter-coder consistency emerging across all categories of codes. A detailed examination of the specific descriptors selected suggests that onsite coding may be particularly beneficial in situations where the aim is to capture any explicit, and perhaps subtle, communication cues, whereas video based coding may be more appropriate in situations where exact sequences of behaviours or measurements of timings are desired. It is anticipated that this type of observation tool will be beneficial for AV developers to increase their understanding of how to interpret the movements of road users, along with increasing knowledge of when implicit and explicit communication techniques should be used.

Keywords: Automated Vehicles, Communication and Interaction, Road Safety, Pedestrians, External-HMI, Human Machine Interface, Human Factors

1. Introduction

Vehicle manufacturers and OEMs are conducting trials of automated vehicles (AVs) across the US, Asia, and Europe (e.g. HumanDrive Project 2020, L3pilot 2020). If these vehicles are to be implemented successfully into current traffic systems, they will need to interact appropriately with other road users, including pedestrians and drivers of manually driven vehicles (Fuest, Sorokin, et al. 2018, Schieben et al. 2019). Knowledge of any patterns emerging in these interactions will be essential to ensure that automated vehicles can successfully, safely, and predictably move through the traffic environment; and can adapt to variations in infrastructure and environment (Fuest, Sorokin, et al. 2018, Madigan et al. 2019). Thus, the purpose of the current paper is to validate a methodology for evaluating the characteristics of pedestrian-vehicle interactions in different circumstances and locations, allowing the identification of common interaction patterns.

While the initial development of AVs has mainly focused on collision avoidance principles and sensors for obstacle detection (Urmson et al. 2008), there has been a growing body of research into the communication requirements of these vehicles (e.g. Habibovic et al. 2018, Merat et al. 2018, Rothenbucher et al. 2016), with the proposal of various visual- and auditory-based solutions. However, there is still very little understanding of where and when these communication solutions should be implemented to promote effective pedestrian-AV interactions.

In recent years, there has been a proliferation of studies examining the factors which *pedestrians* use to make crossing decisions, around both conventional and automated vehicles. Research has shown that the level and criticality of interactions between vehicles and pedestrians is influenced by three main factors – environment/situational characteristics, road user characteristics, and vehicle characteristics (see Madigan et al. 2019 for a review). Examples of environment and situational characteristics which may influence pedestrian gap acceptance and cautiousness at crossing points include road infrastructure e.g. zebra crossings, traffic lights, intersections (Hamilton-Baillie 2008, Havard and Willis 2012); traffic density (Harrell 1991); time of day (Hagel et al. 2014); visibility, and weather conditions (Li and Fernie 2010). Relevant road user characteristics include pedestrian and driver gender (Clamann 2015, Díaz 2002, Harrell 1991, Rosenbloom et al. 2004), age (Bernhoft and Carstensen 2008, Díaz 2002, Oxley et al. 2005), number of pedestrians present (Hamed 2001, Marisamynathan and Vedagiri 2013), pedestrian eye movements (Guéguen et al. 2015, Schneemann and Gohl 2016), head and hand movements (Schmidt and Färber 2009), and pedestrian attention e.g. mobile phone or headphone use (Hatfield and Murphy 2007, Schwebel et al. 2012). All of these factors have been shown to have an impact on the likelihood and speed of pedestrian crossings.

Other papers have investigated the factors influencing *driver* decision making, including factors such as vehicle speed and distance from the crossing point (Várhelyi 1998), and pedestrian eye contact (Guéguen et al. 2015). Research has also been conducted to understand the mechanisms used by drivers to convey their intent, including explicit communication gestures such as hand movements, flashing lights or indicator signals (Sucha et al. 2017); and implicit signals, such as speed adaptation (Sucha et al. 2017, Várhelyi 1998), and vehicle positioning (Fuest, Michalowski, et al. 2018).

Although there is a wide body of support for the importance of implicit communication conveyed through the movement patters of both pedestrians and vehicles (Fuest, Michalowski, et al. 2018, Schmidt and Färber 2009),there have been mixed reports on the importance of more explicit communication cues, with some studies finding evidence of their use (e.g. Guéguen et al. 2015, Rasouli et al. 2017, Sucha et al. 2017) and others finding little or no evidence of explicit communication in traffic interactions (Dey and Terken 2017, Lee et al. 2020, Risto et al. 2017, Straub and Schaefer 2019, Sucha et al. 2017). Thus, it is important to establish if, and when, this type of explicit communication might be expected to occur. A number of studies have aimed to address this issue through observational analyses of pedestrian-vehicle interactions, based on video data collected at ground level (e.g. Domeyer et al. 2019, Rasouli et al. 2017, Risto et al. 2017). Video observations provide a cost effective method for collecting large amounts of data over a relatively short period of time. However, it is possible that the video data is not capturing all of the information used by pedestrians and drivers in their interaction decisions. For example, it may be possible for on-site pedestrians to see driver actions that are not visible in videos due to camera positioning or resolution, and thus a video-based analysis of pedestrian decision-making criteria might miss important elements of an interaction.

Lee et al. (2020) present initial results from a recently developed methodology designed to facilitate *on-site* observations of when, and where, specific types of implicit and explicit communication techniques are used, by capturing a detailed description of potential interaction locations, along with the actions of both vehicles and pedestrians in these locations. The current paper builds on these findings, by presenting a detailed explanation and validation of the observation methodology used. It assesses its value for both on-site and video observations, by comparing the categories selected for the same interactions across the two forms of data collection. By providing a framework for coding traffic interactions in a consistent manner, and capturing the observable environmental, road user, and vehicle variables which occur during these interactions, common patterns of behaviour can be identified, and these can be used to influence AV design. The comparison of the on-site and video-based applications of this methodology will provide a measure of the reliability of this tool for capturing pedestrian-vehicle interactions, along with allowing an evaluation of the strengths and weaknesses of both observation types.

2. METHOD

2.1 Data Collection

Video data, and on-site observations, were collected at an intersection in Leeds, UK. This intersection was selected as it was an accessible and busy location, where the number of pedestrians crossing the road, and the presence of low-speed vehicles, provided opportunities for frequent interactions between vehicles and pedestrians (see Lee et al. 2020). Ethical approval was gained from the University of Leeds Ethics Committee (AREA 17-010), and data was collected on weekdays between 1st November and 20th December 2017. Data was collected at different times of the day, but rush-hour times were avoided as it became too difficult to accurately capture the details of interactions between two specific individuals. Weather conditions during data collection were mostly sunny or overcast. A total of 200 pedestrian – vehicle observations were recorded by three on-site researchers using a specially designed observation protocol (see Section 2.2, also Dietrich and Ruenz 2018). Video data for all of these observations was collected by an outdoor HD wireless IP Foscam FI9803P camera, mounted on the roof of a nearby building (see Point X on Figure 1). This positioning of the camera enabled a view of the whole intersection, but some of the approach to the intersection was obscured. The camera was composed of a colour sensor CMOS, a wireless antenna, and an Infrared lamp array. It collected data at a resolution of 1920 X 1080 at approximately 30 frames per second. During the observations, observers used a laptop to connect via SSH to a distant laptop in the waterproof box in order to activate the camera. Once the observations were finished, the camera could be left running, or deactivated, as required. Approximately 600 hours of video data was collected in total. However, the video recordings were not switched on for every on-site observational analysis.

In order to extract the relevant data clips from each video, the time stamps of the videos and observation protocols were all matched, and the video content was verified by checking the descriptive features mentioned in the observation protocol e.g. pedestrian wearing hoody / vehicle approaching from the left. Through this process, it was possible to extract 123 matched videos, where data was available from the on-site observation. For the current analysis, 50 video observations were selected at random from the on-site data collection process.

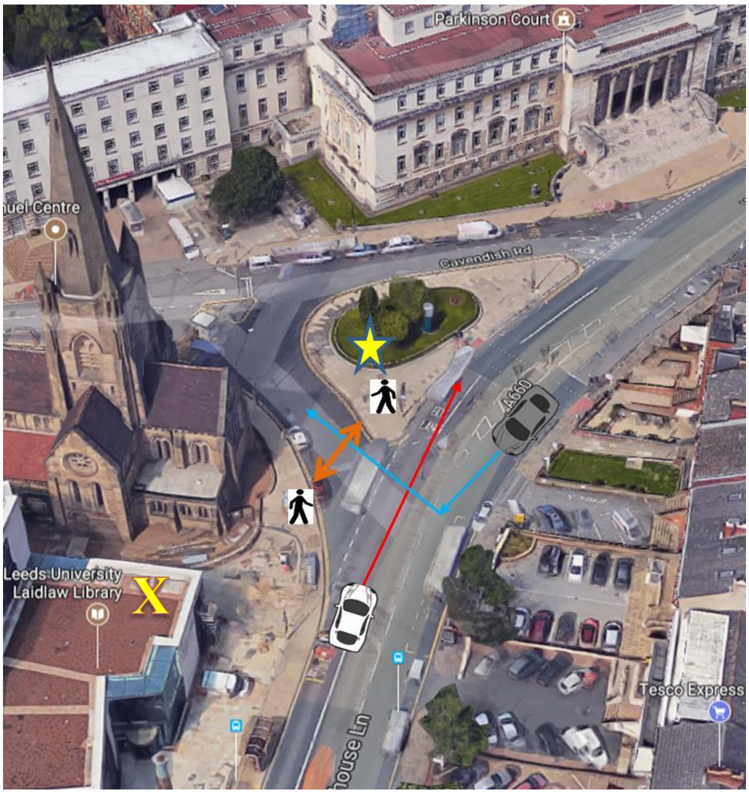

Figure 1: Location of pedestrian-vehicle interactions

2.2 Observation Protocol

Prior to the initial data collection phase, an observation protocol for coding pedestrian–vehicle interactions was developed, based on previous literature in the area. This protocol was designed to capture the presence, or absence, of particular observable elements of pedestrian and vehicle behaviour, which was based on the pedestrian's approach to the intersection, and their road crossing behaviour, such as speed, body movements, and looking behaviours. Through an exploration phase of 70 observations of pedestrian-vehicle interactions, the protocol was refined to consist of 113 descriptors in total - 99 'event types', which captured the observable behaviours of the pedestrians and vehicles as they interacted with each other at each location, and 14 descriptive categories capturing environmental information such as time of day, weather, and possible distractions. The final protocol was incorporated into an html-based application consisting of 4 sections. The first section included four categories of pedestrian movement descriptors (e.g. stepped forward / turned head left), and six categories of driver / vehicle movement descriptors (e.g. decelerated for pedestrian / honked horn), designed to capture the behaviour of pedestrians and vehicles during their approach to the intersection (see Table 1). Here, the **approach phase** was defined as the moment the pedestrian was selected for observation, until they reached the edge of the pavement. The second section of the application included 8 categories, to capture pedestrian and driver / vehicle behaviours during the **crossing phase**, which was defined as the point at which the pedestrian stepped out onto the road, until the point at which one of the actors had passed the other. Pedestrian and driver behaviours included hand and head movements such as waving or turning to look in the direction of another road user. The third section allowed recording of static information such as weather conditions, pedestrian demographics, and group status, and the fourth section provided a sketching tool for observers to show the locations and movement directions of the subjects of interest. This section was included to allow the on-site observations to be linked to the video data.

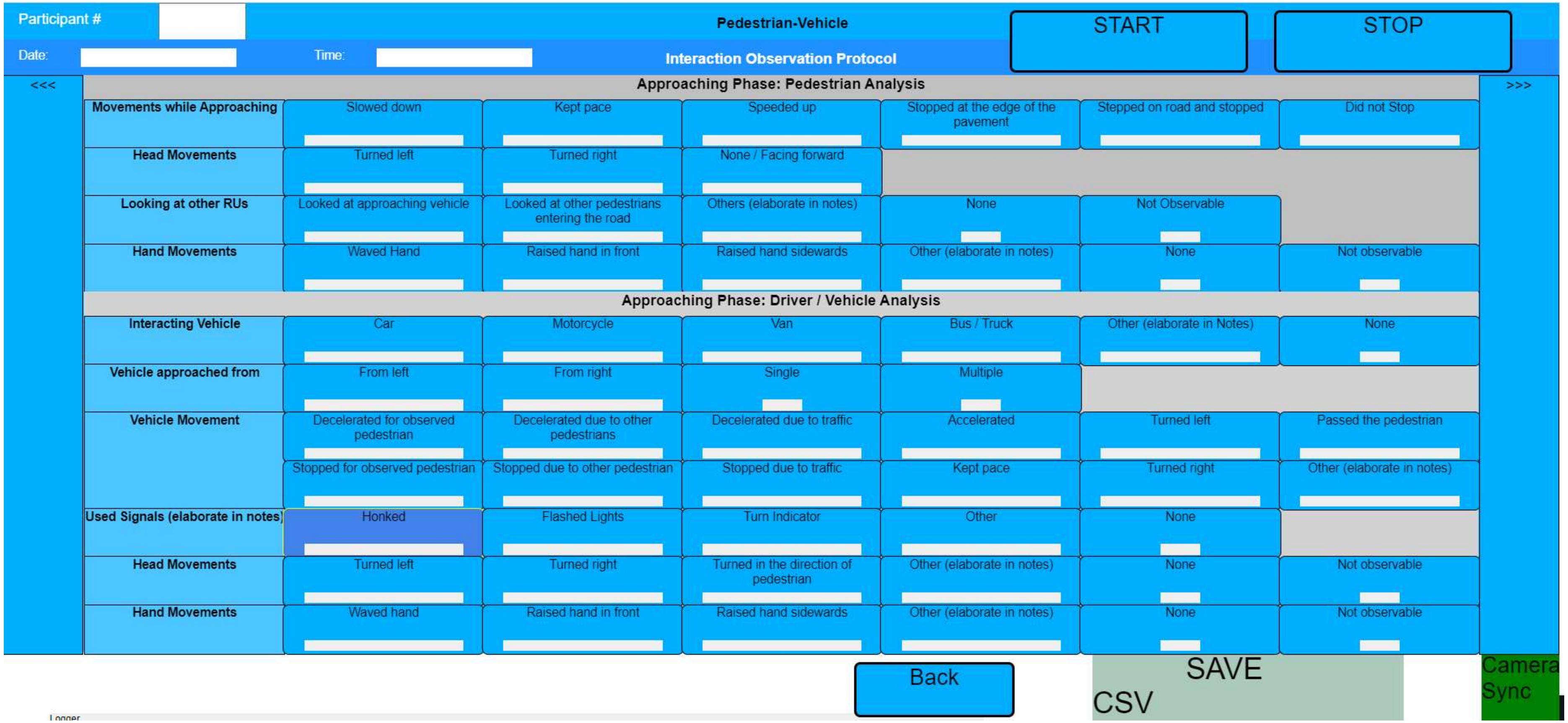

Figure 2: A screen shot of the first page of the observation protocol (see Appendix A for all four pages and Table 1 for a list of all categories included in the protocol)

2.3 Procedure

For the on-site observations, observers were positioned near the star shown in Figure 1, and worked together to identify and agree on when a vehicle-pedestrian pair took part in an ‘interaction’, one observing the vehicle and driver behaviour, and one observing the pedestrian behaviour. At the beginning of each observation, they selected the next pedestrian whose trajectory suggested that they would be crossing the road, making sure there was also a vehicle approaching that the pedestrian may have to interact with. Once an approaching pedestrian and interacting vehicle were selected, each observer described out loud how their subject moved, communicated, and reacted to the other subject, in order to identify the correct sequence in which behaviours occurred. An interaction was considered complete when the pedestrian reached the far-side of the road, or the vehicle passed the pedestrian location. Once the interaction was complete, the observers filled out one observation protocol together, selecting each appropriate descriptor in the order in which it was observed. This enabled the capture of both the actual actions, and the sequence in which they happened. It was possible to select multiple descriptors within each movement phase and category, so the selection of one descriptor did not preclude the selection of another one in the same category.

For the video-based observations, the clips containing the selected on-site observations were extracted from the overall videos of the intersection. These clips started as the pedestrian of interest entered the camera’s field of view, and ended once the interaction was complete. The duration of the video clips ranged between 12 and 39 seconds (M = 21.04 s, SD = 4.61). A screenshot identifying the subjects of interest (based on the information from the on-site observations) was provided for each video clip by an independent researcher (see Figure 3). Two video coders were instructed to focus on the movements of these two parties, and were provided with detailed instructions on how to use the observation protocol to classify the movements of the pedestrians and vehicles.

Figure 3: Example of a screenshot provided to video analysists identifying pedestrian and vehicle of interest (blue circles)

The video coders watched each of the videos and selected the relevant descriptor buttons in the order in which they occurred. They could play the videos at a slower frame rate, and could stop and rewind the videos as often as required, to make sure they had selected the correct codes. After an initial pilot phase of 7 videos, the coders completed the first 10 video-analyses separately, and then met to discuss their coding selections. At this point they determined a need to define each of the coding categories and descriptors more stringently and thus completed the next 20 videos together, developing a definitions table to ensure a shared understanding of each code. Finally, they completed the remaining 20 videos separately, meeting afterwards to discuss any discrepancies in the codes selected (see Figure 4). This process enabled an evaluation of how a shared understanding of the observation protocol application process was developed. Where disagreements in the final codes arose, these were discussed until a consensus was reached. While watching the videos, the coders were also asked to keep a log of any interesting elements or external factors (e.g. movements of other vehicles) which may have influenced the situation, along with identifying any problems with the coding process and suggestions for amendments.

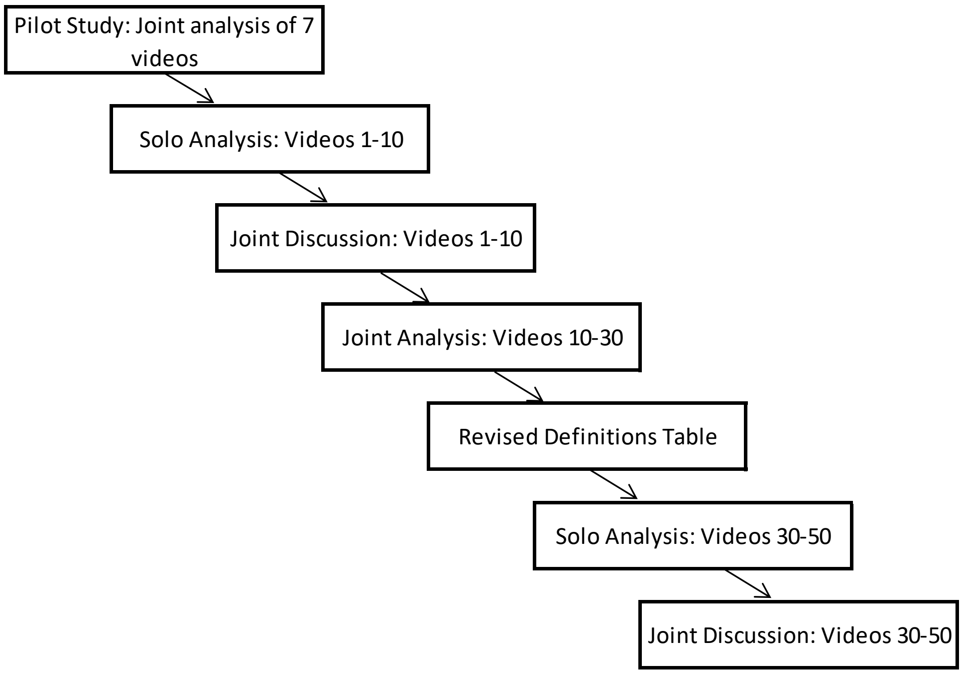

Figure 4: Data analysis process for video-coders

2.4 Data Analysis

Interrater reliability was calculated using the index of concordance (see Wallace and Ross 2006), which provides a percentage agreement. The proportion of codes agreed between two individuals, out of all the possible pairs of codes (selected and unselected), is calculated as follows: (agreements) / (agreements + disagreements). Interrater consensus can then be reported as a figure between 0 and 1, or as a percentage. This method takes into account the cases where coders disagreed, along with providing a method for including situations where there was a difference in the number of codes assigned between coders. A criterion of 70% agreement between coders was adopted as a reasonable minimum, in accordance with Wallace and Ross (2006) and Olsen and Shorrock (2010). The interrater reliability of the two video coders was compared before and after their joint analysis. Their joint final coding selections were then compared to those of the on-site observers. A series of Fisher's exact tests for small sample sizes were conducted to evaluate whether there were any significant differences between the observer groups in the number of selections of a particular coding category. This analysis allows the examination of the significance of association between two categorical variables.

3. Results and Discussion

A. 3.1 Definition of Coding Categories

The first step in the video observation analysis was to refine the definitions of each of the categories defined for the observation protocol, to ensure a shared understanding was achieved between the coders, and to identify any potential shortcomings with the protocol. It was also important to ensure the protocol was suitable for video-based analysis, compared to on-road observations which the app was originally designed for.

For all interactions, data collection started when the pedestrian of interest was visible on the scene. The ***approaching*** phase for both the pedestrian and the vehicle lasted from that point, until the point at which the pedestrian reached the edge of the kerb. The ***crossing*** phase for both the pedestrian and the vehicle started from the point at which the pedestrian stepped onto the road, until they reached the other side, or had been passed by the vehicle. Table 1 provides the definitions for each action phase category and descriptor (see text in bold) from the final video coding analysis, along with any suggested changes. The video coders found most of the original definitions to be straightforward. However, there were a number of suggestions about how the descriptors could be changed to improve the usability in future studies. In particular, there was some question about which phase certain movements belonged in, with the coders feeling that it was important to be very specific about when one phase ended and the next begun. When the term 'other' was selected, there was an opportunity for coders to elaborate on this matter, by adding a text-based response on the third page of the application. The coders also felt that there should be more opportunity to elaborate on the specific demographic features of the observed pedestrian.

Table 1: Observation Protocol category definitions and suggestions for refinement (Descriptors marked in bold in the definition column)

Action Phases	Categories	Descriptors	Suggested Adjustments
Approaching Phase: Pedestrian Movements	Movements while Approaching	**Slowed down**: The pedestrian reduces walking speed	
		Kept pace: The pedestrian reaches kerb without slowing down / speeding up.	
		Speeded up: The pedestrian increases walking speed on approach to the road, or starts running.	
		Stopped at the edge of the pavement: The pedestrian stopped at the edge of the pavement before crossing	
		Stepped on road and stopped: Stepped on road and stopped.	It might be more appropriate to include this in the crossing phase, as the pedestrian has entered the roadway.
		Did not stop: the pedestrian kept moving onto the road without any change in pace.	

	Head movement	If a pedestrian is **turning their head left** or **right** before slowing down, then this should be marked in the sequence before slowing down. If the pedestrian is not turning their head, but just looking forward, then '**None/Facing forward**' should be marked with the same sequence as slowing down. If the pedestrian is slowing down first and then turning head, than this should be marked in the sequence as it happened, and 'None/Facing forward' should not be marked If the pedestrian 'kept pace' and 'turned head' then mark those as being performed at the same time.
	Looking at other Road Users	Mark where the pedestrian was looking from '**looked at approaching vehicle**', '**looked at other pedestrians entering the road**', and '**other**'. If it is not possible to see where the pedestrian is looking, select **not observable**. If it is possible to see a movement and infer a reason, then select others and elaborate your thoughts in the comments section. If you can clearly see that the pedestrian did not look towards other road users select '**none**'.
	Hand Movements	Mark which signals have been used from '**waved hand**', '**raised hand in front'**, '**raised hand sideways'**, and '**other'**. Apply the same rules for looking behaviour to make selections around **'not observable', 'other',** and '**none**'.
Approaching Phase: Vehicle / Driver Movements	Interacting vehicle	Type of the interacting vehicle: **car, motorcycle, van, bus/truck, other, none.**
	Vehicle approached from	**Left:** - When the vehicle approached from the pedestrian's left **Right**: - When vehicle approached from the pedestrian's right **Single / Multiple vehicles**: If more than one vehicle was present which may have had an impact on the interaction, multiple should be selected
	Vehicle movement	Assess the actions of the approaching vehicle for variables; '**decelerated for observed pedestrian'**, '**decelerated due to other pedestrians**', '**decelerated due to traffic**', '**accelerated**', '**stopped for observed pedestrian**', **stopped due to other pedestrians'**, **stopped due to traffic**' and '**kept pace'**. Mark the variables '**turned left**', '**turned right'** and **passed the pedestrian** if the vehicle completes the turn before the pedestrian starts crossing.
	Used signals	Mark which signals have been used from '**honked**', '**flashed lights'**, **'turn indicator**', **'other**', **'none'**.
	Head movements	Mark which head movements have been used by the driver from '**turned left'**, '**turned right'**, **'turned in direction of pedestrian'**, '**other**. If you are unable to see whether the driver moved their head select '**not observable**', and if you can clearly see that they did not move their head, select '**none**'.

	Hand movements	Mark which hand signals have been used by the driver from '**waved hand'**, '**raised hand in front'**, '**raised hand sideways'**, and **other**. Apply the same rules for **none** and **not observable** as used for driver head movements.	
Crossing Phase: Pedestrian Movements	Movements while crossing	Mark what the pedestrian did during the crossing. **Initiated crossing:** Started crossing on own initiative, without any prompt of interacting driver or other road users. Do not mark if the crossing was initiated by the driver (e.g. the driver flashing lights for the pedestrian). If it is not clear, mark other and clarify in notes. Select other categories as appropriate from '**stepped back on pavement**', '**slowed down/stopped while crossing**', '**Speeded up while crossing**', and '**other**'.	No option for kept pace in the crossing movements. Should be incorporated into future iterations. The literature also suggests that an option for "changed crossing trajectory to move around vehicle" should also be included.
	Head movements	Mark the pedestrian's head movement from '**turned left'**, '**turned right**', '**nodded**', and '**none/facing forwards**'	
	Looking at other road users	Mark where the pedestrian was looking from '**looked at vehicle**', '**looked at driver'**, '**looked at other pedestrians entering the road'**. If it is not possible to see where the pedestrian is looking, select **not observable**. If it is possible to see a movement and infer a reason, then select **other** and elaborate your thoughts in the comments section. If you can clearly see that they did not move their head, select '**none**'.	
	Hand movement	Mark the pedestrians hand movement if observable from '**waved hand'**, '**raised hand in front'**, '**raised hand sideways**' and '**other**'. Apply the same rules for **none** and **not observable** as used for driver head movements.	
Crossing Phase: Vehicle / Driver Movements	Vehicle movements	Assess the actions of the approaching vehicle for variables; '**decelerated for observed pedestrian'**, '**decelerated due to other pedestrians'**, '**decelerated due to traffic**', '**accelerated**', '**stopped for observed pedestrian**', **stopped due to other pedestrians'**, **stopped due to traffic**' and '**kept pace'**. Mark the variables '**turned left**', '**turned right**' and **passed the pedestrian** if the vehicle passes the point where the pedestrian would cross and so is no longer a factor in the pedestrian's crossing behaviour. Select **other** if a different movement pattern is observed.	Based on previous literature (e.g. Madigan et al. 2019), it would be good to add a category of 'changed trajectory of movement for observed pedestrian'
	Used signals	Mark if the signals (**honked, flashed lights, turn indicator, other, none)** are initiated while the pedestrian is crossing. If the signal indicator was on the approach stage, do not mark.	
	Head movements	Mark which movements have been observed. (**turned left, turned right, turned in the direction of the pedestrian, other, none, not observable)**	
	Hand movements	Mark which movements have been observed (**waved hand, raised hand in front, raised hand sideways, other, none, not observable).**	
Additional Information	Weather	Mark as accurately as possible if it is **sunny**, **overcast**, **raining**, or **icy**	
	Single pedestrian	Mark if assessment is for the interaction of the single pedestrian only. Mark the gender (**male/female**) as appropriate.	

Group	Mark if the interaction is for more than one pedestrian. Select the **number of males** and the **number of females** observed.	There was an option to select if the observed pedestrian was the leader of the group. However, for the video observation analysis it was not always clear which particular pedestrian had been observed by the on-site observers.
Age	Select the most appropriate age category where possible to estimate (**Child**: under 13 years, **Teenager**: 13-18y, **Young Adult**: 18-30y, **Middle Adult**: 30-60 y, **Older Adult**: 60+ years)	There was no option to select different ages for groups of pedestrians
Potential distraction	Mark if it can be clearly seen that the pedestrian is distracted by **headphones**, **mobile phone use**, or **clothing**, **other**, and **none**.	

B. 3.2 Interrater Reliability

Prior to resolution of any discrepancies in coding between the two video-based coders, the Index of Concordance was used to evaluate the inter-rater reliability of descriptor variables selected in each category, before and after the coders refinement of the shared definitions table. The results show that inter-coder consistency was well above the 70% threshold at all category phases before and after the refinement process (see Table 2). However, the discussion of shared definitions led to an overall increase in the inter-rater reliability of the coding process. In particular, there was a large increase in the reliability of coding for both the vehicle approach and vehicle crossing phases, and a very slight decrease in the reliability of coding for the pedestrian approach and crossing phases. The values for the Index of Concordance were also above the 70% cut off for the level of agreement between the video and on-site coders in all four movement phases. We believe this confirms that the observation protocol provides a reliable tool for categorising the actions of pedestrians and vehicles during both onsite and video based observations. The development of shared definitions and a shared understanding between coders enhances the accuracy of the observations. This type of tool provides a new way of studying human-vehicle interactions that can facilitate our understanding of the exact circumstances in which both implicit and explicit communication tools are used.

Table 2: Measures of interrater reliability and percentage of coding selections for the onsite and video observations

Categorisation Phase	**Index of Concordance (%)**			**Final Codes: Percentage of Descriptors Selected (% of total possible selections)**	
	Between Video-Coders Pre Definition (N = 10)	**Between Video Coders Post Definition (N = 20)**	**Between Video and Field Coders (N = 50)**	**Video Coders**	**On-site Coders**

Approaching Phase: Pedestrian Movements	75.83	75.83	76.67	25.83	25.50
Approaching Phase: Vehicle Movements	87.14	89.29	82.82	8.36	18.36
Crossing Phase: Pedestrian Movements	83.57	82.86	86.63	19.13	21.75
Crossing Phase: Vehicle Movements	74.17	87.92	81.38	14.00	13.88
Overall	80.58	84.13	82.27	15.52	19.39

The last two columns in Table 2 provide a breakdown of the percentage of descriptor selections (e.g. stepped forward / turned head etc.) made under each phase, out of the total number of possible selections (as Table 1 shows the number of descriptors available for selection was different in each category and phase). The results indicate that less than a quarter of all of the possible descriptor selections were made across the 50 observations. Some individual descriptors were selected in almost all of the observations, whereas others were never selected. The graphs presented in Figure 5 provide a further exploration of these results.

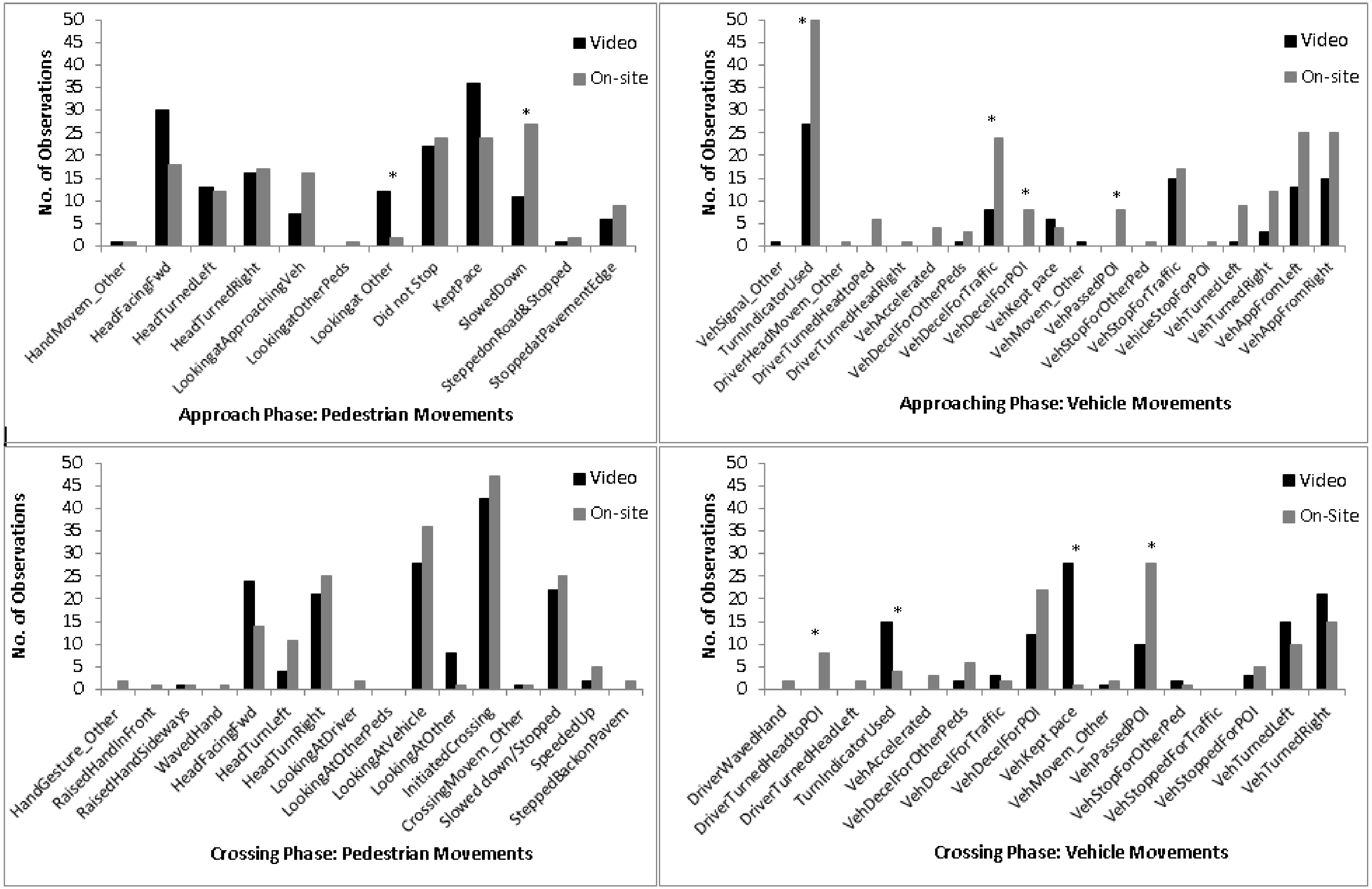

*Figure 5: Comparing the number of descriptors selected within each movement phase for video and on-site observations (*Fishers exact test p<0.01)*

It should be noted that only the descriptors which were selected at least once are included in these graphs. There were a number of descriptors (N = 9) which were never selected by either set of coders. In addition, it was possible to select multiple descriptors within each movement phase and category, so the selection of one descriptor did not preclude the selection of another one.

As Figure 5 shows, there was a great degree of similarity in the codes used by the onsite and video observations, particularly in terms of pedestrian movements, in both the approaching and crossing phases. However, there were some differences in the observations made, particularly in the vehicle movements selected during the approach phase. A series of Fisher's exact tests for small sample sizes were applied, to determine any significant differences at the $p<0.01$ level. These are marked with an asterisk on the graphs. This stringent criteria was selected to offset any concerns surrounding running multiple tests. The results indicate that although the observation protocol could be deemed a reliable measure based on the overall levels of agreement between on-site and video coders, there appeared to be differences in the effectiveness of the observation methods for capturing specific types of actions. These results show that the on-site observers were more likely to evaluate where pedestrians and drivers were looking, suggesting that on-site observations may be more appropriate for identifying and interpreting head and eye-movements. On the other hand, the video observers were more likely to select the category of "kept pace" more often than the on-site observers, while the on-site observers were more likely to select "slowed down". This difference may be due to the capacity to use video timings to make judgements about speed and timings of turns. On the whole both methods show that there were very few instances of explicit communication.

One observation category that was used quite frequently by the video coders during the approach phase was "Looking at other". In all cases, a comment was added that the pedestrians in question had turned their head to look for traffic. During the crossing phase the use of this category was further specified by the coders as relating to traffic other than the interacting vehicle. Also, in one case during the crossing phase, the looking at other referred to a pedestrian looking at their mobile phone. These results suggest one further refinement to the observation protocol of the addition of a 'looking at other traffic' category for pedestrian looking behaviour in both the approaching and crossing phases.

4. Conclusions

The overall purpose of this paper was to present a methodology to facilitate the systematic observation of pedestrian-vehicle interactions, and to validate its use for both onsite and video based observations. The results obtained using the interaction protocol had previously been shown to be useful in identifying the types of explicit and implicit communication mechanisms used by pedestrians and drivers/vehicles (Lee et al. 2020, Uttley et al. 2020), along with enabling the identification of common sequences of behaviours at a given crossing point (Camara et al. 2018). The current paper adds to this research by showing that the tool can be reliably used in both onsite and video observations, although there are benefits and drawbacks of each of these methods.

The results show that the observation protocol provided a consistent method for identifying interaction categories across the two mediums, with high levels of inter-rater reliability emerging between the different observer groups. Across the two studies, it was possible to create an in-depth table of definitions for each of the interaction categories used, with some suggestions for further edits emerging from the video-analysis process. These suggestions mainly focused on the refinement of timings for the approach and crossing phases, along with the addition of a small number of movement descriptor categories. Finally, it was noted that some further refinement of the pedestrian identification process might be required to adequately assess the effects of walking as an individual compared to when in a group setting. It is hoped that the addition of these categories allows this methodology to be used for further understanding of pedestrian – vehicle interactions in a wide variety of contexts and locations e.g. interactions with different types of infrastructure, or interactions with autonomous or other types of vehicles.

In line with the results reported in Lee et al. (2020), the results presented here show that there was limited explicit communication used by pedestrians or drivers / vehicles at this particular intersection. In the majority of cases, pedestrians initiated their crossing movements without any prompt from the interacting driver, or other road users. In addition, there were very few examples of pedestrians using hand or head movements as communication gestures. Instead, both the pedestrians and vehicles seemed to alter their movement patterns to move smoothly around one another, without any requirement for explicit communication.

A detailed comparison of the individual categories selected by the two observer groups showed that there were some differences between the groups regarding the number of times particular categories were selected. Specifically, the onsite coders generally selected codes describing the gestures and looking behaviours of pedestrians and drivers more often than the video coders. This may be a result of greater confidence in interpreting head movements, as the onsite observers had a clearer view of subjects' faces. In relation to pedestrian movement speed, the video observers were more likely to select kept pace while the onsite observers were more likely to indicate that the pedestrian of interest had slowed down. Once again, this may be a result of the quality of the video images, as the frame rate made it difficult for the video observers to detect small changes in speed. However, with higher resolution videos, it should be possible for the observers to make exact evaluations of any change in speed. Similarly, while evaluating vehicle and driver movements during the approach and crossing phases, the onsite observers were more likely to make an interpretation of what the driver was looking at or what a deceleration movement was for. This was most likely linked to their capacity to see the driver in more cases than was possible for the video coders, along with the fact that their experience of any visual looming effects would have been different to the video observers. These results suggest that onsite coding may be beneficial in situations where the aim is to capture any explicit, and perhaps subtle, communication cues.

On the other hand, the results also show some differences between the coders in the phases at which particular actions were selected, with the onsite coders being somewhat more likely to select that the vehicle had turned during the approach phase, or had passed the pedestrian during the crossing phase. This suggests that the delineation of when exactly in the sequence of events the turning movement happened may differ across the coding groups. This could point to an advantage for the use of video coding in situations where knowledge of exact sequences or timings is required, as the video coders have the advantage of being able to rewind and stop the video as needed.

Interestingly, although there was a difference in the number of selections, both the video and onsite category selections suggest that pedestrians took right of way, and reached the other side of the road before the vehicle considerably more often than has been observed in other studies (e.g. Varhelyi, 1998). This may be due to the nature of the intersection, where in almost half of the cases the vehicle was moving slowly after turning right to enter the intersection (this study took place in the UK which has left-hand drive). This finding highlights the importance of understanding context when evaluating pedestrian-vehicle interactions. This issue is likely to become more important in the future, with the development of automated vehicles, which may need to adapt their interaction strategies according to the road structures present.

Overall the results from this study show that the observation protocol provides a useful methodology for capturing patterns of pedestrian-vehicle interactions in both onsite and video based observations. This type of tool is likely to be beneficial for AV developers to understand how to interpret the movements of other road users, along with facilitating an understanding of appropriate implicit and explicit communication techniques. However, it should be noted that this was a relatively small sample, looking at observations in only one location. Therefore, future studies should look at applying this protocol in a wider range of settings and circumstances to understand which behaviour patterns are common across infrastructures and which ones are more likely to change. In addition, the use of higher resolution videos may help determine if subtle communication cues such as head movements can be captured using video observations.

Acknowledgement

This study was conducted as part of the interACT project, which received funding from the European Union's Horizon 2020 research and innovation programme, grant agreement No. 723395.

References

Bernhoft IM, Carstensen G (2008) Preferences and behaviour of pedestrians and cyclists by age and gender. *Transp. Res. Part F Traffic Psychol. Behav.* 11(2):83–95.

Camara F, Giles O, Madigan R, Rothmüller M, Rasmussen PH, Vendelbo-Larsen SA, Markkula G, et al. (2018) Predicting pedestrian road-crossing assertiveness for autonomous vehicle control. *2018 21st Int. Conf. Intell. Transp. Syst.* 2098–2103.

Clamann M (2015) Evaluation of Vehicle-to-Pedestrian Communication Displays for Autonomous Vehicles. *Hum. Factors J. Hum. Factors Ergon. Soc.* 57(3):407–434.

Dey D, Terken J (2017) Pedestrian interaction with vehicles: roles of explicit and implicit communication. *Proc. 9th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* 109–113.

Díaz EM (2002) Theory of planned behavior and pedestrians' intentions to violate traffic regulations. *Transp.*

Res. Part F Traffic Psychol. Behav. 5(3):169–175.

Dietrich A, Ruenz J (2018) Observing traffic--utilizing a ground based LiDAR and observation protocols at a T-junction in Germany. *Congr. Int. Ergon. Assoc.* 537–542.

Domeyer J, Dinparastdjadid A, Lee JD, Douglas G, Alsaid A, Price M (2019) Proxemics and Kinesics in Automated Vehicle--Pedestrian Communication: Representing Ethnographic Observations. *Transp. Res. Rec.* 2673(10):70–81.

Fuest T, Michalowski L, Traris L, Bellem H, Bengler K (2018) Using the Driving Behavior of an Automated Vehicle to Communicate Intentions - A Wizard of Oz Study. *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC* 2018-Novem:3596–3601.

Fuest T, Sorokin L, Bellem H, Bengler K (2018) Taxonomy of traffic situations for the interaction between automated vehicles and human road users. *Adv. Intell. Syst. Comput.* (Springer Verlag), 708–719.

Guéguen N, Meineri S, Eyssartier C (2015) A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing. *Saf. Sci.* 75:87–89.

Habibovic A, Lundgren VM, Andersson J, Klingegård M, Lagström T, Sirkka A, Fagerlönn J, et al. (2018) Communicating Intent of Automated Vehicles to Pedestrians. *Front. Psychol.* 9(August).

Hagel BE, Romanow NTR, Morgunov N, Embree T, Couperthwaite AB, Voaklander D, Rowe BH (2014) The relationship between visibility aid use and motor vehicle related injuries among bicyclists presenting to emergency departments. *Accid. Anal. Prev.* 65:85–96.

Hamed MM (2001) Analysis of pedestrians' behavior at pedestrian crossings. *Saf. Sci.* 38(1):63–82.

Hamilton-Baillie B (2008) Shared space: Reconciling people, places and traffic. *Built Environ.* 34(2):161–181.

Harrell WA (1991) Factors influencing pedestrian cautiousness in crossing streets. *J. Soc. Psychol.* 131(3):367–372.

Hatfield J, Murphy S (2007) The effects of mobile phone use on pedestrian crossing behaviour at signalised and unsignalised intersections. *Accid. Anal. Prev.* 39(1):197–205.

Havard C, Willis A (2012) Effects of installing a marked crosswalk on road crossing behaviour and perceptions of the environment. *Transp. Res. Part F Traffic Psychol. Behav.* 15(3):249–260.

HumanDrive Project (2020) HumanDrive. Retrieved https://humandrive.co.uk/.

L3pilot (2020) L3pilot: Driving Automation. Retrieved https://l3pilot.eu/.

Lee YM, Madigan R, Giles O, Markkula G, Garach-Morcillo L, Fox C, Camara F, et al. (2020) Road users rarely use explicit communication techniques when interacting in today's traffic: Implications for

Automated Vehicles. *Cogn. Technol. Work*.

Li Y, Fernie G (2010) Pedestrian behavior and safety on a two-stage crossing with a center refuge island and the effect of winter weather on pedestrian compliance rate. *Accid. Anal. Prev.* 42(4):1156–1163.

Madigan R, Nordhoff S, Fox C, Ezzati Amini R, Louw T, Wilbrink M, Schieben A, Merat N (2019) Understanding interactions between Automated Road Transport Systems and other road users: A video analysis. *Transp. Res. Part F Traffic Psychol. Behav.* 66:196–213.

Marisamynathan S, Vedagiri P (2013) Modeling pedestrian delay at signalized intersection crosswalks under mixed traffic condition. *Procedia-social Behav. Sci.* 104:708–717.

Markkula G, Madigan R, Nathanael D, Portouli E, Lee YM, Dietrich A, Billington J, Schieben A, Merat N (2020) Defining interactions: A conceptual framework for understanding interactive behaviour in human and automated road traffic. *Theor. Issues Ergon. Sci.* 21(6):728–752.

Merat N, Louw T, Madigan R, Wilbrink M, Schieben A (2018) What externally presented information do VRUs require when interacting with fully Automated Road Transport Systems in shared space? *Accid. Anal. Prev.* 118:244–252.

Olsen NS, Shorrock ST (2010) Evaluation of the HFACS-ADF safety classification system: inter-coder consensus and intra-coder consistency. *Accid. Anal. Prev.* 42(2):437–444.

Oxley JA, Ihsen E, Fildes BN, Charlton JL, Day RH (2005) Crossing roads safely: an experimental study of age differences in gap selection by pedestrians. *Accid. Anal. Prev.* 37(5):962–971.

Rasouli A, Kotseruba I, Tsotsos JK (2017) Agreeing to cross: How drivers and pedestrians communicate. *IEEE Intell. Veh. Symp. Proc.* (Iv):264–269.

Risto M, Emmenegger C, Vinkhuyzen E, Cefkin M, Hollan J (2017) Human-vehicle interfaces: the power of vehicle movement gestures in human road user coordination. *Proc. Ninth Int. Driv. Symp. Hum. Factors Driv. Assessment, Train. Veh. Des.* (Public Policy Center, University of Iowa, Manchester Village, Vermont. Iowa City, IA:), 186–192.

Rosenbloom T, Nemrodov D, Barkan H (2004) For heaven's sake follow the rules: pedestrians' behavior in an ultra-orthodox and a non-orthodox city. *Transp. Res. Part F Traffic Psychol. Behav.* 7(6):395–404.

Rothenbucher D, Li J, Sirkin D, Mok B, Ju W (2016) Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. *25th IEEE Int. Symp. Robot Hum. Interact. Commun. RO-MAN 2016*. 795–802.

SAE (2018) *Definitions for terms related to on-road motor vehicle automated driving systems (J3016B).*

Schieben A, Wilbrink M, Kettwich C, Madigan R, Louw T, Merat N (2019) Designing the interaction of automated vehicles with other traffic participants: design considerations based on human needs and expectations. *Cogn. Technol. Work* 21(1):69–85.

Schmidt S, Färber B (2009) Pedestrians at the kerb - Recognising the action intentions of humans. *Transp. Res. Part F Traffic Psychol. Behav.* 12(4):300–310.

Schneemann F, Gohl I (2016) Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. *IEEE Intell. Veh. Symp. Proc.* 2016-Augus(Iv):38–43.

Schwebel DC, Stavrinos D, Byington KW, Davis T, O'Neal EE, De Jong D (2012) Distraction and pedestrian safety: how talking on the phone, texting, and listening to music impact crossing the street. *Accid. Anal. Prev.* 45:266–271.

Straub ER, Schaefer KE (2019) It takes two to tango: Automated vehicles and human beings do the dance of driving--four social considerations for policy. *Transp. Res. part A policy Pract.* 122:173–183.

Sucha M, Dostal D, Risser R (2017) Pedestrian-driver communication and decision strategies at marked crossings. *Accid. Anal. Prev.* 102:41–50.

Urmson C, Anhalt J, Bagnell D, Baker C, Bittner R, Clark MN, Dolan J, et al. (2008) Autonomous driving in urban environments: Boss and the urban challenge. *J. F. Robot.* 25(8):425–466.

Uttley J, Lee YM, Madigan R, Merat N (2020) Road user interactions in a shared space setting: Priority and communication in a UK car park. *Transp. Res. Part F Traffic Psychol. Behav.* 72:32–46.

Várhelyi A (1998) Drivers' speed behaviour at a zebra crossing: A case study. *Accid. Anal. Prev.* 30(6):731–743.

Wallace B, Ross A (2006) *Beyond human error: taxonomies and safety science* (CRC Press, Boca Raton, FL).

World Health Organization (2013) *Global status report on road safety 2013: supporting a decade of action: summary* (Geneva, Switzerland).

II. Appendix A: Observation Protocol

nt #

Pedestrian-Vehicle

Interaction Observation Protocol

Time:

START

STOP

Approaching Phase: Pedestrian Analysis						
Movements while Approaching	Slowed down	Kept pace	Speeded up	Stopped at the edge of the pavement	Stepped on road and stopped	Did not Stop
Head Movements	Turned left	Turned right	None / Facing forward			
Looking at other RUs	Looked at approaching vehicle	Looked at other pedestrians entering the road	Others (elaborate in notes)	None	Not Observable	
Hand Movements	Waved Hand	Raised hand in front	Raised hand sidewards	Other (elaborate in notes)	None	Not observable

Approaching Phase: Driver / Vehicle Analysis						
Interacting Vehicle	Car	Motorcycle	Van	Bus / Truck	Other (elaborate in Notes)	None
Vehicle approached from	From left	From right	Single	Multiple		
Vehicle Movement	Decelerated for observed pedestrian	Decelerated due to other pedestrians	Decelerated due to traffic	Accelerated	Turned left	Passed the pedestrian
	Stopped for observed pedestrian	Stopped due to other pedestrian	Stopped due to traffic	Kept pace	Turned right	Other (elaborate in notes)
Used Signals (elaborate in notes)	Honked	Flashed Lights	Turn Indicator	Other	None	
Head Movements	Turned left	Turned right	Turned in the direction of pedestrian	Other (elaborate in notes)	None	Not observable
Hand Movements	Waved hand	Raised hand in front	Raised hand sidewards	Other (elaborate in notes)	None	Not observable

nt # | | Pedestrian-Vehicle | START | STOP

Time:

Interaction Observation Protocol

Crossing Phase: Pedestrian Analysis

Movements while crossing	Initiated crossing movement	Stepped back on pavement	Slowed down / stopped while crossing	Speeded up while crossing	Other (elaborate in notes)	
Head Movements	Turned left	Turned right	Nodded	None / Facing forward		
Looking at other RUs	Looked at vehicle	Looked at driver	Looked at other pedestrians entering the road	Others (elaborate in comments)	None	Not observable
Hand Movements	Waved Hand	Raised hand in front	Raised hand sidewards	Other (elaborate in notes)	None	Not observable

Crossing Phase: Driver / Vehicle Analysis

Vehicle Movement	Decelerated for observed pedestrian	Decelerated due to other pedestrians	Decelerated due to traffic	Accelerated	Turned left	Passed the pedestrian
	Stopped for observed pedestrian	Stopped due to other pedestrian	Stopped due to traffic	Kept pace	Turned right	Other (elaborate in notes)
Used Signals (elaborate in notes)	Honked	Flashed Lights	Turn Indicator	Other	None	
Head Movements	Turned left	Turned right	Turned in the direction of pedestrian	Other (elaborate in notes)	None	Not observable
Hand Movements	Waved hand	Raised hand in front	Raised hand sidewards	Other (elaborate in notes)	None	Not observable

nt #

Pedestrian-Vehicle

Time:

Interaction Observation Protocol

START

STOP

General Information

Weather	Sunny	Overcast	Raining	Freezy / Icy		
Single Pedestrian	Individual female	Individual male				
Group	Group	Number of males	number of females	Observed Pedestrian was Leader of the group:	Yes	No
Age	Child (under 13 years)	Teenager (13-18y)	Young Adult (18-30y)	Midage Adult (30-60y)	Older Adult (60+ years)	
Potential Distraction	Listening to headphones	Talking on mobile phone	Looking at mobile phone	Was wearing a hoodie	Other (elaborate in comments)	None

nt #

Pedestrian-Vehicle

Interaction Observation Protocol

Time:

START

STOP

Pedestrian

Vehicle 1

Group

Vehicle 2

Movement

Other vehicle

Braking

Intentional Movement