RESEARCH-ARTICLE

The Role of Emotional Expressions by Autonomous Agents in Road-Sharing: Using Scenario Prompts in Online Video and Virtual Reality Studies

YIYUAN WANG, The University of Sydney, Sydney, NSW, Australia

XINYAN YU, The University of Sydney, Sydney, NSW, Australia

TRAM THI MINH TRAN, The University of Sydney, Sydney, NSW, Australia

MARTIN TOMITSCH, University of Technology Sydney, Sydney, NSW, Australia

Published: 30 November 2024

Citation in BibTeX format

OzCHI '24: 36th Australasian Conference on Human-Computer Interaction
November 30 - December 4, 2024
Meanjin | Brisbane, Australia

OzCHI '24: Proceedings of the 36th Australasian Conference on Human-Computer Interaction (November 2024)
https://doi.org/10.1145/3726986.3727906
ISBN: 9798400715099

The Role of Emotional Expressions by Autonomous Agents in Road-Sharing: Using Scenario Prompts in Online Video and Virtual Reality Studies

Yiyuan Wang
Design Lab, School of Architecture, Design and Planning
The University of Sydney
Sydney, Australia
yiyuan.wang@sydney.edu.au

Xinyan Yu
Design Lab, School of Architecture, Design and Planning
The University of Sydney
Sydney, NSW, Australia
xinyan.yu@sydney.edu.au

Tram Thi Minh Tran
Design Lab, School of Architecture, Design and Planning
The University of Sydney
Sydney, NSW, Australia
tram.tran@sydney.edu.au

Martin Tomitsch
Transdisciplinary School
University of Technology Sydney
Sydney, NSW, Australia
Design Lab, School of Architecture, Design and Planning
The University of Sydney
Sydney, NSW, Australia
martin.tomitsch@uts.edu.au

Abstract

Emotional expressions of robots have been shown to enhance user experience in human-robot interactions. Despite their integration across various robots and contexts, a new research area has emerged with the rise of autonomous agents sharing roads with pedestrians. To date, little work has explored how emotional expressions can support such interactions. This work investigates the effects of an autonomous vehicle (AV) expressing emotions towards pedestrians in four road-sharing scenarios. We designed and simulated these scenarios, gathering feedback from an online video study (N=106) and a virtual reality (VR) study (N=24). Results show that happiness and anger were more easily recognised, while sadness and fear had more varied interpretations. Participants perceived positive emotions as crucial for maintaining social norms and negative emotions as effective for conveying attitudes or situational awareness but not for ensuring behavioural changes. Online video participants were more confident in their emotion recognition choices, while VR participants showed greater mental involvement interpreting the scenarios. We offer design and methodological insights for future research.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; **HCI design and evaluation methods**.

Keywords

Scenarios-Based Prototyping; Emotional Expressions; Autonomous Agents; Autonomous Vehicles; Pedestrians; Road-Sharing; Shared Spaces; Virtual Reality; Online Survey; Video Evaluation; Crowdsourcing; Evaluation Methods

ACM Reference Format:
Yiyuan Wang, Xinyan Yu, Tram Thi Minh Tran, and Martin Tomitsch. 2024. The Role of Emotional Expressions by Autonomous Agents in Road-Sharing: Using Scenario Prompts in Online Video and Virtual Reality Studies. In *36th Australasian Conference on Human-Computer Interaction (OzCHI '24), November 30–December 04, 2024, Brisbane, QLD, Australia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3726986.3727906

1 Introduction

The ability of robots to express emotions has been widely studied for its potential to elicit empathy and improve user experience in human-robot interaction (HRI) [12, 16, 23]. While emotional expressions have been explored in various contexts related to HRI, from domestic companions to public service robots [17], the emergence of new autonomous agents[1] on roads introduces a new area of inquiry. These agents, including delivery robots and autonomous vehicles (AVs), encounter and interact with pedestrians on the street [37]. Recent observational studies have found that they interfered with pedestrians when sharing roads [51, 56, 60]. Studies have found that emotional expressions can enhance the perceived sociability of AVs [54] and evoke empathy in self-moving robots [58]. However, the potential of emotional expressions in handling specific road-sharing situations have not been studied thus far.

In this paper, we investigate the effects of an AV expressing emotions towards pedestrians in four typical road-sharing scenarios involving pedestrians. A scenario in HRI is a detailed narrative that describes a specific situation where humans and robots interact. Using scenarios in the evaluation of emotional expressions enhances viewers' comprehension of why a robot expresses certain emotions and what its intentions may be in the interactional situation [26, 35], thereby revealing the possible impact of these

---

[1] In this paper, we use 'agent' to encompass systems capable of performing tasks autonomously, including robots and autonomous vehicles.

OzCHI '24, Brisbane, QLD, Australia

ACM ISBN 979-8-4007-1509-9/24/11
https://doi.org/10.1145/3726986.3727906

emotional expressions on addressing similar situations in future real-world interactions [48, 57]. We simulate the four scenarios through animations and evaluate them using two methods: videos via an online survey and immersive virtual reality (VR) in a lab study. We aim to answer two research questions (RQs):

- RQ1: What is the perceived impact of emotional expressions by an autonomous agent in road-sharing scenarios?
- RQ2: How are these emotional expressions interpreted differently in scenarios simulated through online videos versus immersive VR?

Our contributions are twofold: first, we provide a foundation for future research to leverage emotional expressions in addressing road-sharing challenges involving autonomous agents; second, we offer insights for evaluating scenario-based emotional expressions of autonomous agents concerning the use of video and VR prototypes.

2 Related Work

2.1 Scenario-Based Emotional Expressions in HRI

The success of designing emotional expressions in robots is often measured by participants' recognition rate, i.e., how well the emotion can be inferred from its expression [31, 44, 45]. Incorporating scenarios can enhance the interpretation of emotional expressions through providing context, helping participants understand the environmental [24], interactional [23], or narrative [38] aspects. A study found that including an appropriate scenario significantly increased the recognition rates of emotions compared to using an inappropriate scenario [35].

Furthermore, scenarios can serve as prompts to evoke the imagination in participants, enriching design implications through elicited user feedback. Applications are evident in fostering social storytelling in young children's development through simple scenarios of robots' emotional expressions [3, 38]. Providing participants with simple behavioural cues or interaction referents can inspire in-depth narratives [7] and the creation of interaction patterns for robots [28].

2.2 Simulation Platforms for Emotional Expressions

Physical prototypes enable in-the-wild evaluations [9, 24] and provide direct access to the look and feel of the robotic device [7, 28, 47]. Many studies opt for video simulations to rapidly animate scenarios or expressions [6, 21, 35, 48] or to distribute evaluations online at scale [21, 48]. In recent years, VR has emerged as a promising tool for evaluating HRI designs. VR platforms allow safe evaluations of safety-critical agents, such as drones [22] and AVs [33], and provide immersive depictions of scenarios [22, 33], demonstrating transferability to real-world testing [30, 32, 52].

2.3 Pedestrian Interaction with Autonomous Agents

Findings from observational studies indicated that current pedestrian interactions and attitudes towards autonomous agents on streets, including AVs and autonomous robots, are influenced by novelty [34, 56, 60], trust and safety perceptions [13, 37, 60], social norms [37, 51, 56], environmental contexts [37, 51, 61], among others. While it is found that longer exposure to AVs tend to instill confidence and trust in pedestrians [13], communication systems should be designed in such autonomous agents to handle road-sharing challenges, such as responding to positive or antagonistic pedestrian behaviours [13, 34], eliciting voluntary assistance from pedestrians [56, 59], and integrating with the socio-material human environment [37, 53].

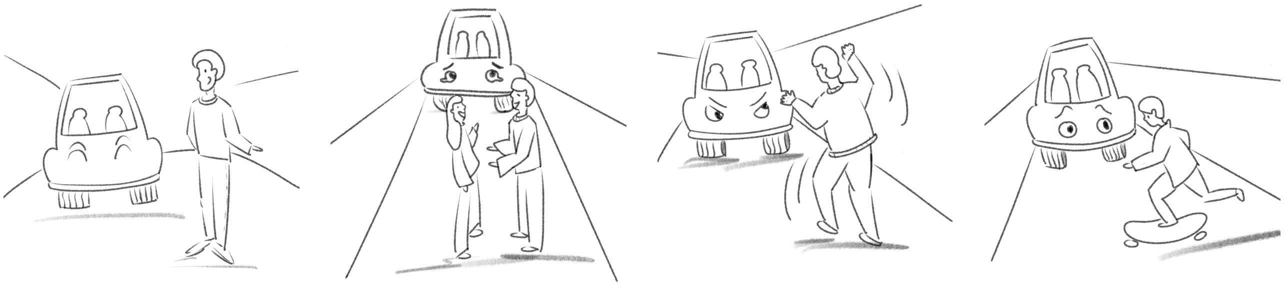
Applications of emotional expressions in agents range from assistive home devices [23, 57] to recreational outdoor robots [9, 24]. However, little is known about how emotional expressions can assist autonomous agents in interacting with pedestrians in traffic [14, 50]. No study has examined their effects on solving specific road-sharing challenges. Our work bridges this gap by eliciting interpretations of AVs' emotional expressions using four road-sharing scenario prompts. Prior research on simulation platforms for autonomous agents suggests that the same scenario prompt can be experienced differently and influence results differently depending on the prototype representation used [25, 32]. Therefore, through online video and VR evaluations, we further uncover insights into the two prototyping methods, supporting future evaluations of scenario-based emotional expressions for agents with real-world testing limitations.

3 Methodology

3.1 Designing Scenario Prompts

We identified four typical road-sharing scenarios from the literature related to autonomous agent-pedestrian interactions (see Figure 1). These scenarios were derived from studies that conducted real-world observations of AV-pedestrian interactions [13, 34, 51] or designed for the behaviour of a self-moving robot [39]. The four scenarios are *granted passage*, *requesting passage*, *antagonised*, and *forced halt*. In *granted passage*, a pedestrian uses a hand gesture to allow the AV to proceed first [40, 51]. In *requesting passage*, two pedestrians conversing in a narrow corridor fail to notice the AV's need to pass [39, 51]. In *antagonised*, a pedestrian intentionally obstructs the AV's path to test or delay it [13, 34]. In *forced halt*, a pedestrian abruptly enters the AV's path, causing it to stop immediately [51].

In human communication, facial expressions play a fundamental role in information conveyance [15], enabling humans to infer emotional states, personality traits, and intentions [18, 36]. In AV-pedestrian research, eye expressions have been found useful for communicating AVs' awareness of pedestrians [11, 42] and manoeuvre intentions [10, 19]. Hence, designing emotional expressions through eyes has the potential to facilitate emotion inferences and the integration with existing AV designs. This work focuses on four basic emotions based on Ekman's framework [15]: happiness, sadness, anger, fear. These four emotions are universally recognised across culture [15] and are commonly used in research on the emotional perception in robots [2, 5]. For the representation style of the eye expressions, we chose a design rated as the most likeable and the most effective in conveying emotions in previous HRI and AV-pedestrian studies [21, 27, 54].



(a) Granted passage **(b) Requesting passage** **(c) Antagonised** **(d) Forced halt**

Figure 1: Illustration of road-sharing scenarios identified in literature.

We mapped the four basic emotions to the four road-sharing scenarios. A happy expression was applied to *granted passage* to convey a positive mood and affirmation towards the pedestrian. Sadness and anger were employed to represent two different response mechanisms in human conflict resolution behaviours [1]: sadness, characterised by pleading and low dominance, signified a polite and deferential approach [46, 49] for *requesting passage*; anger, on the other hand, indicated an assertive response to [1, 43] the *antagonised* scenario. Prior work has found that although robots with commanding behaviours were less socially accepted, they were sometimes more effective than polite requests in achieving desired outcomes [43]. Lastly, a fear expression was applied to *forced halt* to suggest heightened alertness and agitation due to sudden external stimuli [29].

3.2 Creating Video and VR Prototypes

We animated the four scenarios using the Unity game engine. Figure 2 illustrates the animations through frames at representative timestamps. The pedestrian-vehicle shared space and the AV were 3D-modelled in Autodesk 3ds Max and imported into Unity. For creating pedestrian behaviours, we used high-fidelity models of people from a 3D library[2] and customised their movements for the scenarios. The animations were exported as video prototypes in the mp4 format. The same animations were also deployed on the Oculus Quest 2 as immersive VR prototypes.

3.3 User Evaluation

We conducted user evaluations in an online video-supported survey study and a VR-based lab study, using the same scenario prompts in their 2D (videos) and immersive 3D (VR) formats, with the camera positioned at the same location and perspective. The 3D format supported head rotation and physical movement in the VR environment. The video study was hosted on the Qualtrics software and deployed on Amazon Mechanical Turk (MTurk). The VR study was conducted at the University of Sydney. Both studies have been approved by the Human Research Ethics Committee in the University of Sydney.

3.3.1 Data Collection and Procedure. The following data was collected for both the video and VR studies.

- **Comprehension of scenario**: How would you describe the scenario in the [video/VR]?
- **Emotion recognition**: What emotion did this vehicle express? (Choose one from anger, disgust, fear, happiness, sadness, surprise)
- **Justification of choice**: Please justify why you chose the above emotion.
- **Confidence score**: What is your level of confidence in choosing the above emotion? (1: not at all confident – 7: very confident)
- **Second choice**: Would you like to add any additional emotion that could apply to what this vehicle expressed?

Participants in both studies consented to data collection at the beginning of the study session. Participants were presented with the four scenarios one by one in a randomised order from balanced Latin Square. After they watched/experienced each scenario, they were asked to answer all of the above questions before proceeding to the next scenario. Towards the end of the study session, participants completed a demographic questionnaire.

3.3.2 Participants. The video study recruited 106 participants (52% male, 44% female, 4% others; M=39.2 years, SD=12.1 years) through the MTurk platform. Participants had completed more than 1000 human intelligence tasks (HITs) with an approval rate above 95%. They were also required to be at least 18 years old, speak fluent English (professional working proficiency), and pass attention-check questions inserted in the survey. The study took an average of 17 minutes (SD=6.1) to complete, which was estimated at 15 minutes based on a pilot study. Participants were remunerated $2 plus a bonus between $0.25 and $1 for elaborating answers to open-ended questions.

The VR study recruited 24 participants (63% male, 33% female, 4% others; M=27.2 years, SD=4.8 years) through physical flyers and social networks. Participants were required to be at least 18 years old, have normal or corrected-to-normal eyesight, and speak fluent English. The study took an average of 26 minutes (SD=5.5) to complete, which was estimated at 20 minutes based on a pilot study. All study sessions were audio-recorded. Participants did not receive compensation.

3.3.3 Data Analysis. We calculated descriptive statistics for emotion recognition and confidence scores, including the recognition

---

[2] https://renderpeople.com/

The pedestrian is going to cross in front of the approaching AV.

Both the pedestrian and the AV slow down and stop, after which the pedestrian gives way to the AV.

The AV expresses happiness through its eyes.

The AV continues on its way while maintaining the expression.

Two pedestrians are conversing on a narrow path where the AV is going to pass through.

The AV slows down, stops, and waits for the pedestrians to grant passage.

The AV expresses sadness through its eyes.

The AV stays waiting, maintaining the expression.

The pedestrian and the AV approach each other.

The pedestrian walks towards the AV and intentionally blocks the way, causing the AV to slow down and stop.

The AV expresses anger through its eyes.

The AV maintains the expression while the pedestrian keeps blocking the way.

The AV drives on the open space at a constant speed.

The skateboarder suddenly rushes out from the side, causing the AV to slow down and stop.

The AV expresses fear through its eyes at the same time as the interaction happens.

The skateboarder crosses in front of the AV, and the AV maintains the expression.

Figure 2: An illustration of the four animated scenarios used in both online and VR studies. Each scenario is depicted through representative frames at four typical timestamps: scenario start, interaction start, expression start, and scenario end.

rate (the ratio of emotions recognised as intended) and the mean confidence scores, separated by correct and incorrect recognition. Furthermore, we conducted Chi-Squared tests to determine if there were significant differences in recognition rates between the two studies for each emotion. We also performed non-parametric Mann-Whitney U tests to compare differences in mean confidence scores for correct and incorrect recognition between the two studies for each emotion. All quantitative analyses were conducted using the Python SciPy library.

We transcribed the audio recordings from the VR study using a professional AI-supported transcription service[3] and subsequently

---

[3] https://otter.ai/

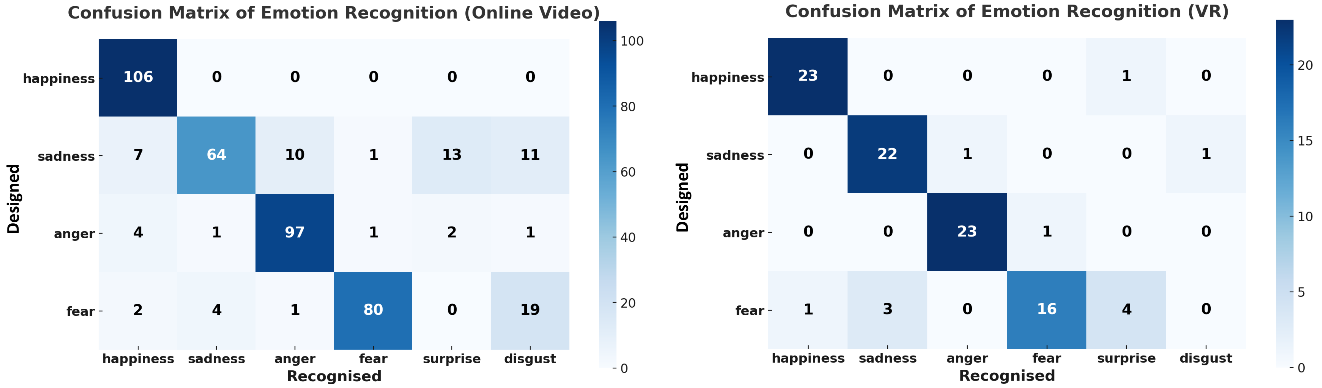


Figure 3: Emotion recognition in the online video study (left) and in the VR study (right).

reviewed the transcripts manually. Data from the video study were directly extracted from the survey platform. The first three authors of this paper collaboratively conducted an inductive thematic analysis [4] on the data collected from both studies. Initially, each coder independently analysed half of the data from both the video and VR studies. The three coders then convened to discuss the emerging themes and sub-themes. Following this, the first author synthesised these discussions into a unified set of themes and developed a codebook agreed upon by all coders. The first author then applied this codebook to code the remaining datasets.

4 Results

4.1 Emotion Recognition

We report on emotion recognition in terms of recognition rate (i.e., the ratio that the emotion was recognised as intended ('correct'), confidence score (i.e., how confident participants were in making the recognition), and reasons for 'incorrect' recognition (i.e., why participants interpreted the emotion expression differently).

4.1.1 Recognition rates. Table 1 presents differences in the emotion recognition rate between online video and VR studies. Figure 3 shows how well the emotions were recognised correctly compared to the counterpart emotion labels. Happiness and anger showed high accuracy in both online video and VR studies. In the VR study, happiness was mistook for surprise by one participant, while there was no mistakes in the online video study. Anger was only identified as fear in the VR study by one participant, but its incorrect recognition was more varied in the online video study, with participants confusing it with happiness, surprise, disgust, and sadness. The other recognition of fear leaned heavily towards surprise in both studies but were more pronounced in the online video study. Finally, the recognition rate of sadness was much lower in the online video study than that in the VR study. For sadness, the incorrect recognition was more varied in the online video study, with surprise, disgust, and anger being common answers, whereas in the VR study, it was equally split between disgust and fear.

The chi-square test result for sadness ($\chi^2 = 7.216$, $p < 0.01$) indicates a statistically significant difference in the ability of participants to recognise sadness in the *requesting passage* scenario between the VR and online studies. Specifically, the recognition rate suggests that participants in the VR study were better at discerning sadness for the *requesting passage* scenario (91.67%) compared to the participants in the online video study (60.38%).

4.1.2 Confidence scores. Table 1 shows how confident participants were in making correct emotion recognition and incorrect emotion recognition for the four scenarios across the two studies. Looking at the mean scores, we found that in general, participants were more confident when their recognition was correct compared to when their recognition was incorrect; participants in the online video study were more confident in making recognition choices, regardless of correctness, compared to the participants in the VR study.

The Mann-Whitney U test results indicates a statistically significant difference between the online video and VR studies in the mean confidence scores for correct recognition for anger (Mann-Whitney U = 1372.0, $p < 0.05$) and for fear (Mann-Whitney U = 819.0, $p < 0.05$). However, no statistical significance was found in correct recognition for happiness or sadness. Furthermore, a statistically significant difference between the online video and VR studies in the mean confidence scores for incorrect recognition for fear was observed (Mann-Whitney U = 154.0, $p < 0.05$). No significant difference was found in incorrect recognition for the other emotions.

4.1.3 Reasons for 'incorrect' recognition. We observed four reasons for participants' incorrect recognition (i.e., why participants interpreted the emotion expression differently), namely 'misread AV intention', 'misread expression design', 'misread pedestrian behaviour', and 'own judgement of suitable emotion'. Table 2 shows the percentages of participants whose qualitative answers indicated these four reasons. The four coding categories are mutually exclusive. Participants misread the intention of the AV only in the online study, with a higher percentage for sadness. A notable percentage of participants in the VR study misread the design of the fear expression, while the same reason was more prevalent for sadness in the online study. Additionally, participants used their own judgement of what emotion was most suitable for the situation. This was prominent for both sadness and fear for both studies.

Table 1: Differences in emotion recognition rates and confidence scores between online video and VR studies ($^{*}p<0.05$, $^{**}p<0.01$).

Emotion (scenario)	Happiness (granted passage)		Sadness (requesting passage)		Anger (antagonised)		Fear (forced halt)	
Study	Online video	VR	Online video	VR	Online video	VR	Online video	VR
Recognition rate	100%	95.83%	60.38%	91.67%	91.51%	95.83%	75.47%	66.67%
p-value	0.41		**0.007****		0.77		0.53	
Mean confidence score (correct recognition)	6.60	6.35	6.22	6.05	6.68	6.22	6.38	5.88
p-value	0.92		0.74		**0.03***		**0.049***	
Mean confidence score (incorrect recognition)	No incorrect recognition	3.00	5.40	4.50	6.33	5.00	6.20	5.50
p-value	N/A		0.21		0.19		**0.03***	

Table 2: Percentages of participants in online video study and VR studies interpreting the emotional expressions differently ('incorrect' recognition), categorised by four reasons observed from qualitative data.

	Happiness (granted passage)		Sadness (requesting passage)		Anger (antagonised)		Fear (forced halt)	
	Online video	VR	Online video	VR	Online video	VR	Online video	VR
Misread AV intention			5.66%		1.89%		1.89%	
Misread expression design		4.17%	16.04%				10.38%	20.83%
Misread pedestrian behaviour			0.94%		3.77%	4.17%		
Own judgement of suitable emotion			16.98%	8.33%	2.83%		12.26%	12.50%

4.2 Perceived Impact of Emotional Expressions

Our qualitative analysis revealed five themes regarding the perceived impact of emotional expressions, as shown in Figure 4. For each theme, we report on its sub-themes, along with a visual comparison (stacked bar charts) between online video and VR studies, highlighting the percentages of participants suggesting each sub-theme per emotion/scenario.

4.2.1 Intent to influence pedestrian behaviour. This theme pertains to the perception of the AV attempting to evoke a response from the pedestrians, such as asking them to *“move out of the way and let it pass” (sadness)* or *“stop being annoying and make way” (anger).* It is worth noting that the sub-theme ‘motivating bystander assistance’ appeared only in the VR data. This denotes participants’ remarks about stepping into the AV-pedestrian interaction to assist the AV. They noted how expressions of sadness and fear can affect *“people surrounding the AV”*, eliciting feelings to *“stop the conversation” (sadness)* or *“’take responsibility to help in a dangerous situation” (fear).*

4.2.2 Indication of AV capability. This theme shows that participants felt the emotional expressions indicated the capabilities of the AV. They noted the AV was able to *“notice the potential danger” (fear)* (risk anticipation) or *“indicate its awareness of the intentional obtrusion” (anger)* (situational awareness). Participants also sensed the AV’s willingness to take responsibility and accountability for its actions: *“I also think that the vehicle is trying to tell the boy that he is worried about the boy might get hurt and it doesn’t want to hurt him” (fear).* Additionally, a higher ratio of VR participants provided feedback around ‘situational awareness’, ‘responsibility and accountability, and ‘safe actions’. The sub-theme ‘understanding and coordination’ appeared only in the online video data, denoting perceptions of the AV’s ability to understand and coordinate with human instructions.

4.2.3 Maintenance of social norms. This theme indicates the emotional expressions were perceived as a way for the AV to practise social norms in its interactions with pedestrians. The happiness expression was effective in reciprocating positive pedestrian behaviours, as suggested by 46% of participants in both studies, mentioning that it conveyed *“gratitude”* and was akin to *“saying ‘thank you’”*. Following this, displaying emotions was considered as symbolising the AV’s intentions to encourage or discourage pedestrian behaviours. Participants noted the expressions conveyed the AV’s *“attitudes”* rather than *“actual emotions”*, or simply to *“achieve its goal”*. Hence, the emotions were thought to *“incentivise the action of giving way” (sadness)* or act *“as a warning to pedestrians” (anger).*

Furthermore, a higher percentage of VR participants felt the AV engaged in social dynamics or enabled human-like interactions. They thought the happiness expression contributed to a *“polite atmosphere”*, while the sadness expression was a “gentle way to ask for help” since it was *“unusual for an AV to go to a pedestrian pathway and ask people to give way”*. Regarding ’enabling human-like interactions,’ participants in the online video study often related the expression design to human facial or bodily reactions, such as *“almost like your knees or hands would shake when a person is scared” (fear).* VR participants more often noted how the emotional expressions personified the AV, mentioning it had a *“character”*, or drew parallels with driver-pedestrian interactions.

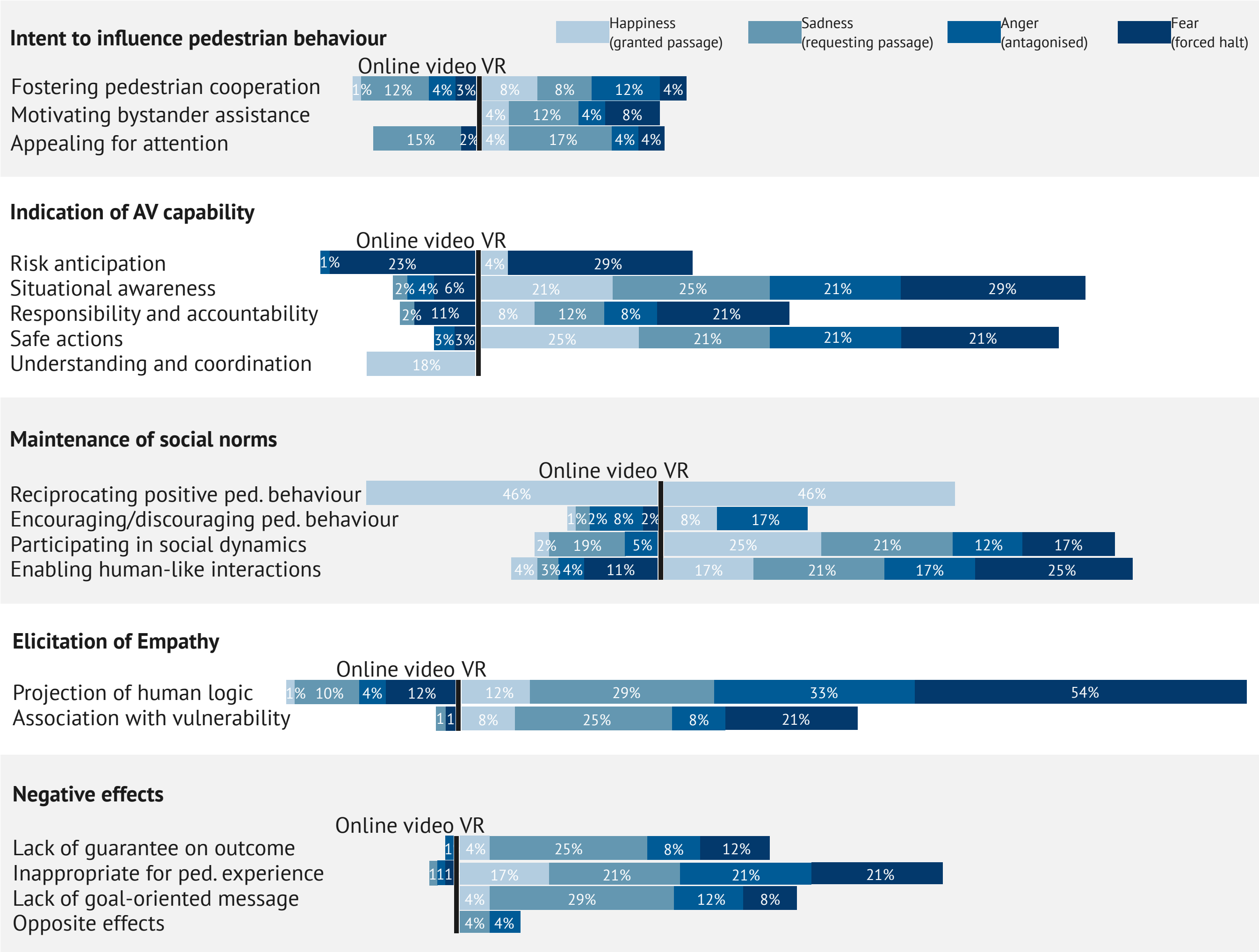


Figure 4: Themes (bold) and sub-themes suggesting the perceived impact of emotional expressions, generated from qualitative analysis.

4.2.4 Elicitation of empathy. This theme suggests that the emotional expressions evoked empathy from participants. Participants projected how they or other humans would respond emotionally in the AV's position, such as *"I'll get frustrated if I were the car" (sadness)* and *"I think part of it is just me bringing my own assumptions that I would be angry in that situation" (anger).* They also viewed the AV as the vulnerable party in the situation, exemplified by remarks like *"Because there was no real driver, I think the car was vulnerable since it couldn't alert the people in front of it" (sadness).* Additionally, the percentage of participants in the VR study who provided feedback related to this theme is noticeably higher than in the online video study.

4.2.5 Negative effects. This theme arises from some negative effects of using emotional expressions as suggested by participants. This concern was mostly raised by participants in the VR study, while rarely mentioned by those in the online study. First, participants were concerned that emotional expressions might not cause behavioural changes in pedestrians since *"people don't care about its emotion"* or *they don't really have time to check the screen"*. Second, displaying emotions might impair the pedestrian experience. Participants suggested pedestrians should still control whether they go first, felt showing emotions was *"too much"* or made the AV emph"less professional", and that negative emotions shouldn't make people feel *"blamed" (sadness)* or show any tendency to *"cause harm" (anger).* Third, participants recommended the AV also display goal-oriented messages, such as instructions for pedestrians, to *"make sure that people get out of the way"*. Lastly, two participants noted that showing a reaction might be gratifying for those taunting the AV, while not reacting might be less entertaining for them.

4.3 Methodological Observations

Our qualitative analysis identified three themes demonstrating patterns in how participants interpreted the scenarios, which have

Table 3: Percentages of participants showing patterns about how they interpreted the scenarios, as observed from qualitative data.

	Happiness (granted passage)		Sadness (requesting passage)		Anger (antagonised)		Fear (forced halt)	
	Online video	VR	Online video	VR	Online video	VR	Online video	VR
Interpreting AV emotion in conjunction with its movement	31.13%	16.67%	29.25%	8.33%	11.32%	8.33%	46.23%	29.17%
Interpreting AV behaviour in temporal sequence	4.72%	29.17%	24.53%	20.83%	16.04%	50.0%	4.72%	20.83%
Judging pedestrian behaviour	16.04%	20.83%	2.83%	8.33%	14.15%	20.83%	15.09%	12.5%

methodological implications in choosing between the two evaluation methods. Table 3 presents the percentages of participants who exhibited these patterns per emotion/scenario.

4.3.1 Interpreting AV emotion in conjunction with its movement. Participants interpreted the AV's emotional expressions in conjunction with its movement during pedestrian interactions. Generally, they agreed that the AV operated at a safe speed and braked effectively. Participants in the online video study frequently mentioned the AV's movement when interpreting its emotional expressions, especially for the *forced halt*, *granted passage*, and *requesting passage* scenarios. Participants in the VR study showed a similar pattern for the *forced halt* and *granted passage* scenarios, though it was less prominent than in the online study.

4.3.2 Interpreting AV behaviour in temporal sequence. Participants interpreted the AV's behaviours in the temporal sequence presented in the animations. This was particularly prominent in the VR study, especially for the *antagonised* scenario, where participants appreciated that the anger expression was delayed for a few seconds after the pedestrian's teasing behaviour. This pattern was also evident in the other three scenarios in the VR study. In contrast, participants in the online video study mostly exhibited this pattern for the *requesting passage* scenario, noting mostly how long the AV waited in the interaction.

4.3.3 Judging pedestrian behaviour. Participants appraised the actions of pedestrians, specifically complimenting or negatively judging their behaviours. In general, this was more evident for the *granted passage*, *antagonised*, and *forced halt* scenarios, but less for the *requesting passage* scenario. In addition, more participants exhibited this pattern in the VR study for *granted passage*, *requesting passage* and *antagonised*, while more in the online video study for *forced halt*.

5 Discussion

Our results revealed the perceived impact of emotional expressions by an autonomous agent in road-sharing scenarios (RQ1) and how the emotional expressions were interpreted differently in scenarios simulated through online videos versus immersive VR (RQ2). Based on these results, we discuss design insights around using emotional expressions of autonomous agents to support road-sharing with pedestrians (subsection 5.1 and 5.2) and methodological insights around evaluating these emotional expressions with online video or VR prototypes (subsection 5.3 and 5.4).

5.1 Ambiguity in Emotion Recognition

Happiness and anger were easily recognised, as shown by the emotion recognition results in Figure 3. However, sadness and fear were often mistaken for other negative emotions. This confusion arose from both the design of the expressions and the participants' subjective judgements (see section 4.1.3). Literature suggests that humans inherently struggle to clearly discern sadness and fear because these emotions utilise similar facial muscles [29]. This makes designing distinct expressions for these emotions challenging. Despite participants selecting other negative emotions in the scenarios, they demonstrated a high level of comprehension of the situations. This suggests that people tended to rely on their judgement to determine the most suitable emotion, reflecting the diversity in individual conflict resolution approaches [1]. The fact that autonomous agents operate in public spaces may make it difficult to personalise strategies for these varied interpretations. From a strategic perspective, happiness being the only positive option likely contributed to its high recognition rate. This suggests that in certain situations, using a representative 'negative expression' could suffice to symbolise the autonomous agent's emotion. Traffic environments are inherently information-rich and cognitively demanding for pedestrians. The introduction of emotional expressions in AVs, while potentially beneficial, may increase cognitive workload if pedestrians are required to discern subtle differences between emotions [55]. Therefore, designing intuitive and easily interpretable emotional expressions [54] is essential for helping pedestrians efficiently infer the AV's intent and awareness without adding to their cognitive burden.

5.2 Functional and Social Messages

Participants identified that the emotional expressions communicated two main types of information: the AV's functional capabilities and social norms (see Figure 4). For functional capabilities, participants inferred causal relationships between the interactional situations and the AV's emotional expressions, such as the AV anticipating risks or being aware of pedestrians' behaviours, aligning with literature suggesting that emotional expressions can effectively convey functional messages, including a person's current states and future actions [18]. Regarding social norms, participants indicated that emotional expressions communicated complex social messages, such as reciprocating pedestrian behaviours or discouraging future actions, which might be difficult to convey through other communication means equipped on autonomous agents [8, 18]. Furthermore, the emotional expressions elicited empathy from viewers and made the AV appear vulnerable in conflict situations. However, negative effects were perceived, as these emotional expressions do not provide clear guidance on behavioural compliance. While emotional expressions can help maintain social norms and implicitly reflect functional capabilities, additional measures are needed to guide behaviours more effectively. Future work should consider exploring how using emotional expressions to convey functional messages or social norms can support public trust and AV adoption in open-road environments.

5.3 Interpretation Confidence in Online Video

Participants in the online video study reported higher confidence scores in recognising emotional expressions within the scenarios, despite having lower recognition rates compared to those in the VR study (see Table 1). Factors that may have influenced the online participants' confidence levels include the anonymity of the online setting, which can make respondents feel more secure in their choices, reducing the fear of judgement and resulting in more confident responses [41]. Additionally, online surveys offer a relatively self-paced environment compared to lab studies, allowing participants to take their time and think through their choices, resulting in a perceived sense of control over the survey process [20]. Furthermore, our methodological observations (see Table 3) revealed that participants often interpreted the emotional expressions of the AV in relation to its movement. This indicates that participants perceived a connection between the AV's emotional expressions and its trajectory, particularly in relation to potential conflict points and overall traffic safety associated with pedestrians. Such interpretations suggest that emotional cues could play a role in signaling the AV's intentions at critical moments, contributing to traffic safety in road-sharing situations.

5.4 Mental Involvement in Immersive VR

Participants in the VR study provided more in-depth answers to qualitative questions, as evidenced by results from the thematic analysis (see Figure 4). Not only were the percentages of participants per theme higher, but some themes also only appeared in the VR study, such as 'motivating bystander assistance'. Reasons behind this outcome include the immersive environments, which could have induced a sense of presence and being part of the situation [22, 33, 52]. Furthermore, based on the interpretation patterns, we found that participants often focused on the temporal sequence of the animations, interpreting the AV's behaviour within that context. Although VR has been shown to mimic real-world traffic scenes and widely used for testing eHMIs [25, 32, 52], the prescribed tasks for participants reduced the naturalness of their interactions. In real-world scenarios, the length of the animations could interfere with traffic efficiency and potentially cause pedestrians to disengage from the eHMIs due to prolonged waiting times.

5.5 Limitations and Future Work

Biases in this work could be induced by uncontrolled variations between the two studies, such as differences in the administration process and the different participant cohorts each method attracted. For example, online participants might have used their own monitors at home, while lab study participants were likely from nearby neighbourhoods. Furthermore, our work used an explicit and straightforward representation of robots' emotional expressions to convey understandable emotions, which allowed us to better evaluate their effects in the given scenarios. It is worth noting that other representations of emotional expressions, such as abstract lights and movements, are also often studied in HRI [24, 44, 45]. Future work should consider replicating the road-sharing scenarios in real life and testing the efficacy of emotional expressions in real-world settings.

6 Conclusion

We presented an evaluation of emotional expressions by an AV in four road-sharing scenarios with pedestrians, using animated prompts of these scenarios in an online video study and an immersive VR study. Our findings revealed the perceived impact of these emotional expressions on AV-pedestrian interactions, indicating their role in communicating functional and social messages and fostering assistance and empathy. Furthermore, the findings highlighted differences in interpretation between the online video and VR studies. Online video participants showed higher confidence in recognising the emotions, while VR participants exhibited greater mental involvement in interpreting the scenarios. This work provides insights into using emotional expressions as a strategy to assist autonomous agents in sharing roads with pedestrians, as well as the effects of online video or VR prototypes in evaluating these scenarios.

Acknowledgments

We thank all the participants in our user studies for their contributions to this research. We thank the reviewers of this paper for their supportive feedback. This study was funded by the Australian Research Council through grant numbers DP200102604 Trust and Safety in Autonomous Mobility Systems: A Human-Centered Approach and DP220102019 Shared-Space Interactions Between People and Autonomous Vehicles.

References

[1] Franziska Babel, Andrea Vogt, Philipp Hock, Johannes Kraus, Florian Angerer, Tina Seufert, and Martin Baumann. 2022. Step aside! VR-based evaluation of adaptive robot conflict resolution strategies for domestic service robots. *International Journal of Social Robotics* 14, 5 (2022), 1239–1260.

[2] Christoph Bartneck, Juliane Reichenbach, and van A Breemen. 2004. In your face, robot! The influence of a character's embodiment on how users perceive its emotional expressions. (2004).

[3] Laura Boccanfuso, Elizabeth S Kim, James C Snider, Quan Wang, Carla A Wall, Lauren DiNicola, Gabriella Greco, Lilli Flink, Sharlene Lansiquot, Pamela Ventola, et al. 2015. Autonomously detecting interaction with an affective robot to explore connection to developmental ability. In *2015 international conference on affective computing and intelligent interaction (ACII)*. IEEE, 1–7.

[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[5] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *International journal of human-computer studies* 59, 1-2 (2003), 119–155.

[6] Mason Bretan, Guy Hoffman, and Gil Weinberg. 2015. Emotionally expressive dynamic physical behaviors in robots. *International Journal of Human-Computer Studies* 78 (2015), 1–16.

[7] Paul Bucci, Lotus Zhang, Xi Laura Cang, and Karon E MacLean. 2018. Is it happy? Behavioural and narrative frame complexity impact perceptions of a simple furry robot's emotions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.

[8] Judee K Burgoon and Jerold L Hale. 1984. The fundamental topoi of relational communication. *Communication Monographs* 51, 3 (1984), 193–214.

[9] Jessica R Cauchard, Kevin Y Zhai, Marco Spadafora, and James A Landay. 2016. Emotion encoding in human-drone interaction. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 263–270.

[10] Chia-Ming Chang, Koki Toda, Xinyue Gui, Stela H Seo, and Takeo Igarashi. 2022. Can Eyes on a Car Reduce Traffic Accidents?. In *14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 349–359.

[11] Chia-Ming Chang, Koki Toda, Daisuke Sakamoto, and Takeo Igarashi. 2017. Eyes on a Car: an Interface Design for Communication between an Autonomous Car and a Pedestrian. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*. 65–73.

[12] Filipa Correia, Samuel Mascarenhas, Rui Prada, Francisco S Melo, and Ana Paiva. 2018. Group-based emotions in teams of humans and robots. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*. 261–269.

[13] Miguel Ángel de Miguel, Daniel Fuchshuber, Ahmed Hussein, and Cristina Olaverri-Monreal. 2019. Perceived pedestrian safety: Public interaction with driverless vehicles. In *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, New York, NY, USA, 90–95.

[14] Jiayuan Dong, Nikhil Gowda, Yiyuan Wang, Mungyeong Choe, Areen Alsaid, Ignacio Alvarez, Sven Krome, and Myounghoon Jeon. 2024. Inside Out: Emotion GaRage Vol. V. In *Adjunct Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 260–263.

[15] Paul Ekman and Wallace V Friesen. 1971. Constants across cultures in the face and emotion. *Journal of personality and social psychology* 17, 2 (1971), 124.

[16] Friederike Eyssel, Frank Hegel, Gernot Horstmann, and Claudia Wagner. 2010. Anthropomorphic inferences from emotional nonverbal cues: A case study. In *19th international symposium in robot and human interactive communication*. IEEE, 646–651.

[17] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and autonomous systems* 42, 3-4 (2003), 143–166.

[18] Chris D Frith and Uta Frith. 2006. How we predict what other people are going to do. *Brain research* 1079, 1 (2006), 36–46.

[19] Xinyue Gui, Koki Toda, Stela Hanbyeol Seo, Chia-Ming Chang, and Takeo Igarashi. 2022. "I am going this way": Gazing Eyes on Self-Driving Car Show Multiple Driving Directions. In *14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 319–329.

[20] Nick Hajli and Xiaolin Lin. 2016. Exploring the security of information sharing on social networking sites: The role of perceived control of information. *Journal of Business Ethics* 133 (2016), 111–123.

[21] Viviane Herdel, Anastasia Kuzminykh, Andrea Hildebrandt, and Jessica R Cauchard. 2021. Drone in love: Emotional perception of facial expressions on flying robots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–20.

[22] Olivia Herzog, Niklas Forchhammer, Penny Kong, Philipp Maruhn, Henriette Cornet, and Fritz Frenkler. 2022. The influence of robot designs on human compliance and emotion: A virtual reality study in the context of future public transport. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 2 (2022), 1–17.

[23] Guy Hoffman, Oren Zuckerman, Gilad Hirschberger, Michal Luria, and Tal Shani Sherman. 2015. Design and evaluation of a peripheral robotic conversation companion. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. 3–10.

[24] Marius Hoggenmueller, Jiahao Chen, and Luke Hespanhol. 2020. Emotional expressions of non-humanoid urban robots: the role of contextual aspects on interpretations. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays*. 87–95.

[25] Marius Hoggenmüller, Martin Tomitsch, Luke Hespanhol, Tram Thi Minh Tran, Stewart Worrall, and Eduardo Nebot. 2021. Context-based interface prototyping: Understanding the effect of prototype representation on user feedback. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[26] Malte F Jung. 2017. Affective grounding in human-robot interaction. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 263–273.

[27] Alisa Kalegina, Grace Schroeder, Aidan Allchin, Keara Berlin, and Maya Cakmak. 2018. Characterizing the design space of rendered robot faces. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 96–104.

[28] Lawrence H Kim and Sean Follmer. 2019. Swarmhaptics: Haptic display with swarm robots. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[29] Daniel H Lee and Adam K Anderson. 2017. Reading what the mind thinks from how the eye sees. *Psychological science* 28, 4 (2017), 494–503.

[30] Franziska Legler, Jonas Trezl, Dorothea Langer, Max Bernhagen, Andre Dettmann, and Angelika C Bullinger. 2023. Emotional Experience in Human–Robot Collaboration: Suitability of Virtual Reality Scenarios to Study Interactions beyond Safety Restrictions. *Robotics* 12, 6 (2023), 168.

[31] Diana Löffler, Nina Schmidt, and Robert Tscharn. 2018. Multimodal expression of artificial emotion in social robots using color, motion and sound. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 334–343.

[32] Martina Mara, Jan-Philipp Stein, Marc Erich Latoschik, Birgit Lugrin, Constanze Schreiner, Rafael Hostettler, and Markus Appel. 2021. User responses to a humanoid robot observed in real life, virtual reality, 3D and 2D. *Frontiers in psychology* 12 (2021), 633178.

[33] Anselmo Martínez, Lidia M Belmonte, Arturo S García, Antonio Fernández-Caballero, and Rafael Morales. 2021. Facial emotion recognition from an unmanned flying social robot for home care of dependent people. *Electronics* 10, 7 (2021), 868.

[34] Dylan Moore, Rebecca Currano, Michael Shanks, and David Sirkin. 2020. Defense against the dark cars: Design principles for griefing of autonomous vehicles. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, New York, NY, USA, 201–209.

[35] Jekaterina Novikova and Leon Watts. 2014. A design model of emotional body expressions in non-humanoid robots. In *Proceedings of the second international conference on Human-agent interaction*. 353–360.

[36] Nikolaas N Oosterhof and Alexander Todorov. 2008. The functional basis of face evaluation. *Proceedings of the National Academy of Sciences* 105, 32 (2008), 11087–11092.

[37] Hannah RM Pelikan, Stuart Reeves, and Marina N Cantarutti. 2024. Encountering autonomous robots on public streets. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 561–571.

[38] Yu Peng, Yuan-Ling Feng, Nan Wang, and Haipeng Mi. 2020. How children interpret robots' contextual behaviors in live theatre: Gaining insights for multi-robot theatre design. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 327–334.

[39] Björn Petrak, Gundula Sopper, Katharina Weitz, and Elisabeth André. 2021. Do you mind if i pass through? Studying the appropriate robot behavior when traversing two conversing people in a hallway setting. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 369–375.

[40] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. 2017. Agreeing to cross: How drivers and pedestrians communicate. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, New York, NY, USA, 264–269.

[41] Lynne D Roberts and Camilla J Rajah-Kanagasabai. 2013. " I'd be so much more comfortable posting anonymously": Identified versus anonymous participation in student discussion boards. *Australasian Journal of Educational Technology* 29, 5 (2013).

[42] Jaguar Land Rover. 2018. The virtual eyes have it. *https://www.fastcompany.com/90231563/people-dont-trust-autonomous-vehicles-so-jaguar-is-adding-googly-eyes* (2018).

[43] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. 141–148.

[44] Sichao Song and Seiji Yamada. 2017. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 2–11.

[45] Sichao Song and Seiji Yamada. 2018. Designing expressive lights and in-situ motions for robots to express emotions. In *Proceedings of the 6th international conference on human-agent interaction*. 222–228.

[46] Vasant Srinivasan and Leila Takayama. 2016. Help me please: Robot politeness strategies for soliciting help from humans. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 4945–4955.

[47] Haodan Tan, John Tiab, Selma Šabanović, and Kasper Hornbæk. 2016. Happy moves, sad grooves: using theories of biological motion and affect to design shape-changing interfaces. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 1282–1293.

[48] Hamish Tennent, Dylan Moore, and Wendy Ju. 2018. Character actor: Design and evaluation of expressive robot car seat motion. *proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–23.

[49] Cristen Torrey, Susan R Fussell, and Sara Kiesler. 2013. How a robot should give advice. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 275–282.

[50] Yiyuan Wang, Luke Hespanhol, and Martin Tomitsch. 2021. How can autonomous vehicles convey emotions to pedestrians? a review of emotionally expressive non-humanoid robots. *Multimodal Technologies and Interaction* 5, 12 (2021), 84.

[51] Yiyuan Wang, Luke Hespanhol, Stewart Worrall, and Martin Tomitsch. 2022. Pedestrian-vehicle interaction in shared space: Insights for autonomous vehicles. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 330–339.

[52] Yiyuan Wang, Marius Hoggenmüller, Guixiang Zhang, Tram Thi Minh Tran, and Martin Tomitsch. 2024. Immersive In-Situ Prototyping: Influence of Real-World Context on Evaluating Future Pedestrian Interfaces in Virtual Reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[53] Yiyuan Wang, Martin Tomitsch, Marius Hoggenmüeller, Senuri Wijenayake, Wai Yan, and Luke Hespanhol. 2025. From passersby to placemaking: designing autonomous vehicle-pedestrian encounters for an urban shared space. *Multimedia Tools and Applications* (2025), 1–25.

[54] Yiyuan Wang, Senuri Wijenayake, Marius Hoggenmüller, Luke Hespanhol, Stewart Worrall, and Martin Tomitsch. 2023. My eyes speak: Improving perceived sociability of autonomous vehicles in shared spaces through emotional robotic eyes. *Proceedings of the ACM on Human-Computer Interaction* 7, MHCI (2023), 1–30.

[55] Yiyuan Wang, Xinyan Yu, and Martin Tomitsch. 2023. Designing Emotional Expressions of Autonomous Vehicles for Communication with Pedestrians in Urban Shared Spaces: Use Cases, Modalities, and Considerations. In *Proceedings of the 35th Australian Computer-Human Interaction Conference*. 454–461.

[56] David Weinberg, Healy Dwyer, Sarah E Fox, and Nikolas Martelaro. 2023. Sharing the sidewalk: Observing delivery robot interactions with pedestrians during a pilot in Pittsburgh, PA. *Multimodal Technologies and Interaction* 7, 5 (2023), 53.

[57] Steve Whittaker, Yvonne Rogers, Elena Petrovskaya, and Hongbin Zhuang. 2021. Designing personas for expressive robots: Personality in the new breed of moving, speaking, and colorful social home robots. *ACM Transactions on Human-Robot Interaction (THRI)* 10, 1 (2021), 1–25.

[58] Yuhui You, Mitchell Fogelson, Kelvin Cheng, and Bjorn Stenger. 2020. EMI: An expressive mobile interactive robot. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[59] Xinyan Yu, Marius Hoggenmüller, and Martin Tomitsch. 2024. Encouraging bystander assistance for urban robots: Introducing playful robot help-seeking as a strategy. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 2514–2529.

[60] Xinyan Yu, Marius Hoggenmüller, Tram Thi Minh Tran, Yiyuan Wang, and Martin Tomitsch. 2024. Understanding the Interaction between Delivery Robots and Other Road and Sidewalk Users: A Study of User-generated Online Videos. *ACM Transactions on Human-Robot Interaction* (2024).

[61] Xinyan Yu, Tram Thi Minh Tran, Yiyuan Wang, Kristina Mah, Yidan Cao, Stine S Johansen, Wafa Johal, Maria Luce Lupetti, Megan Rose, Markus Rittenbruch, et al. 2024. Out of Place Robot in the Wild: Envisioning Urban Robot Contextual Adaptability Challenges Through a Design Probe. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.