UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Psicologia Generale

Brain, Mind, and Computer Science Ph.D. Course

Neuroscience, Technology, and Society curriculum

XXXIV cycle

MORAL DILEMMAS IN VR DRIVING SIMULATIONS: EFFECTS ON BEHAVIOR AND EMOTIONAL EXPERIENCE

Funded by CARIPARO

Supervisor: Luciano Gamberini

Co-supervisor: Gian Antonio Susto

Ph.D. Student: Giulia Benvegnù

Academic Year: 2020/2021

To Luca.

Table of Contents

Abstract

Moral dilemmas, such as the trolley problem and its variants, have long been used as a paradigm for studying moral decision-making. In recent times, two factors have shaken up this field: on the one hand, the diffusion of Virtual Reality (VR), a technology that has proven to be the ideal tool to study dangerous situations that cannot be safely reproduced in the real world (such as moral dilemmas) in a more realistic way. On the other hand, the introduction of the first autonomous vehicles (AVs), which has made moral dilemmas a tool to investigate user preferences in unavoidable collision situations. This Ph.D. thesis combines these lines of research, exploring behavior and emotional reactions in VR versions of moral dilemmas applied in the driving context through three studies.

Study 1 investigated the eventual presence of emotional differences between a human and an autonomous driving modality. Results showed that being the one who decides in life-death situations or experiencing the same decision made by an AV elicited a very different pattern of arousal and valence.

Study 2 explores the effect of different legal frameworks in crashes involving semi-autonomous cars. Main findings revealed that being informed that the legal liability lies with the manufacturing company when the driver has no control of the vehicle did not affect participants' behavior but made them feel less morally responsible for accidents made by semi-autonomous cars.

Study 3 analyzes the effect of the presentation method, text or VR, on decision and emotional reactions during unavoidable accident situations, showing that people are inclined to sacrifice their lives to save pedestrians in textual moral dilemmas, but not in the realistic VR version of the same dilemmas.

The thesis begins with a presentation of the main theories of moral decision-making (chapter 1), followed by a brief description of VR technologies and work that has applied them to the study of moral dilemmas (chapter 2). Relevant literature on moral dilemmas in the driving context is then presented (chapter 3). Finally, the three studies of this thesis are described and discussed.

1. Moral dilemmas: the Trolley and the Footbridge problems

1.1. Greene's dual-process model of moral judgment

A runaway trolley is heading towards five workers who will be killed if it continues its run. The only way to save them is to pull a lever that will cause the trolley to swerve to another track, where it will run over and kill one worker instead of five. Is it morally appropriate to pull that lever?

To the moral dilemma described above, known as the Trolley dilemma (Foot, 1978; Thomson, 1986), many people respond affirmatively (Greene et al., 2001, 2004; Petrinovich et al., 1993; Petrinovich & O'Neill, 1996). However, if instead of pulling a lever, they were asked to push a man off an overpass to slow down the trolley with his body and thus avoid the death of the five workers, would this choice be still morally acceptable? In this version, known as the Footbridge dilemma (Thomson, 1986), only a small minority judge this action appropriate (Greene et al., 2001, 2004; Petrinovich et al., 1993; Petrinovich & O'Neill, 1996), despite the cost-benefit in terms of lives at stake being identical in the two cases (1 vs. 5) (Figure 1.1).

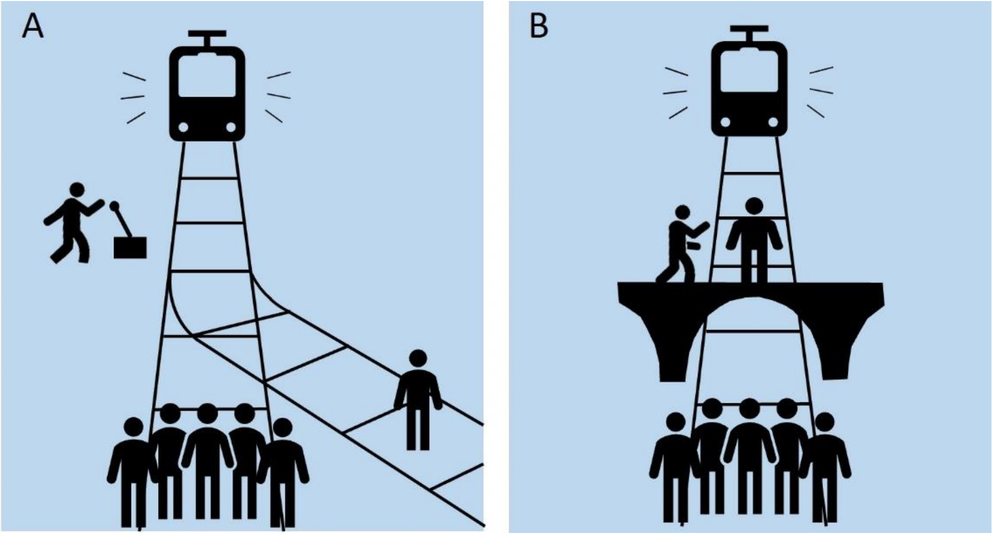


Figure 1.1. a) Schematic representations of the Trolley dilemma and b) of the Footbridge version.

In the Trolley dilemma, there is a tendency to respond in a "utilitarian" way, maximizing the overall good. Instead, in the Footbridge dilemma, people tend to choose a "deontological" resolution, according to which moral rules (in this case, "don't kill") should not be violated regardless of the consequences (Greene et al., 2008). But why does this happen? What makes it morally appropriate to pull the lever but not push the man off the bridge? Philosophers have tried to answer this question by providing normative explanations, seeking principles that justify the judgments commonly given to the two dilemmas. Kant (Kant, 1785/1959) and Aquinas (Aquinas, 1265-1272 /1947) thought that the intentional killing of an individual for a greater good was morally unacceptable. In this light, sacrificing the man in the Footbridge dilemma is not allowed because it is literally the means to save others. The death of an individual, on the other hand, is acceptable if it is an expected but unwanted consequence: in the Trolley dilemma, the worker who is killed if the lever is pulled represents a side effect of saving other lives.

Greene et al. (2001, 2004) sought instead to provide a descriptive rather than normative explanation, thus focusing on how a particular decision is reached rather than how it would be right to respond. Their point of view is articulated on the “personal/ impersonal” dichotomy. A moral dilemma is defined as personal if the violation described complies with the following criteria: 1) it must involve physical harm, 2) it must concern a specific individual or group of people, 3) the damage must not be caused simply by the diversion of an existing threat. These criteria can be reformulated in the terms “I wound you”: the criterion “wound” identifies the most primitive types of violations, such as physical aggression; the "you" criterion ensures that the victim is vividly represented as a specific individual; the "I" condition captures the concept of "agency," which provides that the action derives directly from the will of the agent, seen as an active subject (Greene et al., 2001, 2004; Greene & Haidt, 2002). If a moral dilemma does not meet these criteria, it is classified as impersonal.

The Footbridge dilemma, according to this classification, contains a "personal and direct" moral violation. This violation, following this model, elicits an automatic and immediate negative emotion that prevails over rational reasoning, guiding moral judgment towards non-utilitarian responses (i.e., let the five workers die) (Greene et al., 2001, 2004). On the other hand, the Trolley dilemma required an "impersonal" action that evokes a weaker or absent emotional response. In this case, cognitive control prevails and guides the decision towards utilitarian responses that maximize benefits and minimize costs (i.e., kill one man to save the five workers) (Greene et al., 2001, 2004). The foundation of this interpretation is largely evolutionary. Evidence from the observation of great apes suggests that the human species evolved using emotional responses as a form of behavior regulation (de Waal, 1996), in the apparent absence of moral reasoning (intended as a slow and deliberative process that implies introspection and abstraction). In this light, the strong negative emotions elicited by direct and "personal" moral violations can be interpreted as modalities of behavioral regulation developed during evolution to guarantee the conservation of the human species: incurring such violations would have meant harming the processes of cooperation and social cohesion, essential for survival (Sober & Wilson, 1998; Trivers, 1971).

Greene et al. (2001, 2004) investigated the processes underlying moral dilemmas resolution and the associated brain activity designing and using a set of dilemmas inspired by the Trolley and Footbridge problems. In particular, their moral dilemmas were classified as “personal” (Footbridge-type) and “impersonal” (Trolley-type) based on the three criteria described above.

As previously mentioned, their model argues that both a slow and controlled cognitive system and a rapid emotional one come into play when solving personal/impersonal dilemmas, which makes the model fit into the corpus of double process theories. (Greene et al., 2001, 2004, 2008). These theories are based on the existence of two opposing processes: one unconscious and automatic and another slow, explicit, and deliberative (for example, see Kahneman, 2003; Lieberman et al., 2002).

Although Greene's dual-process model of moral judgment has generally found supporting evidence in behavioral, neuroimaging, psychophysiological, and neuropsychological studies, some conflicting findings, and criticisms have also emerged. The relevant literature is reported in the following subsections.

1.1.1. Behavioral and self-reported data

The first work that highlighted the competition between emotional and cognitive processes during moral dilemmas resolution consisted of a task where participants had to judge whether the proposed moral action was appropriate or inappropriate. When personal moral dilemmas were presented, only a minority of participants considered the action “appropriate,” and, interestingly, this required longer response times (RT) than classifying it as "inappropriate.” The authors suggested that this finding was due to the effort to answer incongruently against the salient and automatic emotional reaction elicited by Footbridge-type dilemmas only (Greene et al., 2001).

In a subsequent study, the same group of authors tested the presence of a causal relationship between controlled cognitive processes and utilitarian resolution. They found that when participants had to make the classic "appropriate/inappropriate" judgment task under cognitive load (a concurrent digit-search task), this selectively interfered with the choice of the utilitarian option, increasing the RT (Greene et al., 2008).

The authors attribute the lack of a significant change in RT in non-utilitarian responses to the fact that these are modulated by emotional processes of a more automatic nature, giving further support to the inclusion of the model in the double process theories (Greene et al., 2008).

Subsequent studies with similar paradigms found mixed results regarding RT data. The work of Manfrinati et al. (2013) did not show a significant difference in decisions times between utilitarian and non-utilitarian resolutions in Footbridge-type dilemmas. In addition, they found slower RT in Trolley-type dilemmas when participants chose not to kill a man to save more lives, compared to the utilitarian choice. Following Greene's model, in such cases, there should be no emotional response to overcome, and consequently, no difference in response times should be expected between the two possible resolutions (Manfrinati et al., 2013). Again, when the two types of dilemma were directly compared (Lotto et al., 2014; Sarlo et al., 2012), conflicting results emerged. While Lotto et al. (2014) highlighted that killing a man in Footbridge-type dilemmas required longer RT than doing the same in Trolley-type dilemmas, in line with the dual-process model (Greene et al., 2001, 2004), in the work of Sarlo et al. (2012) no significant differences were found.

It is possible that cognitive and emotional processes participate in both deontological and utilitarian moral judgments (Manfrinati et al., 2013). The two mechanisms could play a role in Trolley-type and Footbridge-type dilemmas, although to varying degrees.

As stated by Cushman et al. (2010), debating over whether a moral judgment is reached exclusively by reason as opposed to emotion might be overly simplistic. Rather, moral decision-making could be the product of complex interactions between emotional and cognitive processes (Cushman et al., 2010; Manfrinati et al., 2013).

Studies investigating the emotional impact elicited by the dilemmas using subjective evaluations align with this last consideration. Following Greene's dual-process model, an increase in the self-reported unpleasantness and arousal in Footbridge-type dilemmas should be expected compared to the Trolley-type ones. Despite some studies effectively finding a higher negative valence in Footbridge-type dilemmas (Manfrinati et al., 2013; Sarlo et al., 2012), others failed to replicate this result (Lotto et al., 2014; Pletti et al., 2015). A similar pattern was also found for arousal, with a consistent number of works that did not highlight any significant difference between the two types of dilemmas in terms of self-reported emotional activation (Lotto et al., 2014; Manfrinati et al., 2013; Pletti et al., 2015; Sarlo et al., 2012).

Other authors have investigated which specific emotions come into play when solving moral dilemmas. The work of Choe & Min (2011) dealt with this last aspect, analyzing which emotions were felt most intensely during the choice of utilitarian resolution in personal (Footbridge-type) dilemmas. Guilt was the most frequently reported emotion, followed by sadness, disgust, and anger. According to the authors, the fact that guilt was the predominant emotion may be due to two possible reasons: it could be the result of foreshadowing the killing of someone before carrying out the act, or it could be a post-decision emotion. Instead, sadness, disgust, and anger were chosen as the most intense emotions in dilemmas with particular characteristics and contents. Sadness was predominately elicited in the scenarios where the protagonist had to kill his own offspring to save more lives. On the other hand, disgust was mainly reported in dilemmas in which one was forced to take drastic actions for selfish reasons. Finally, anger was associated with scenarios in which a particularly unfair situation was presented (Choe & Min, 2011). Overall, the study shed some light on the nature of the “strong and aversive” emotional response elicited in Footbridge-type dilemmas and showed that emotions are present (and perhaps could play a role) not only in the deontological resolution but also in the utilitarian one.

More recently, a more in-depth analysis on the role of emotion in both Trolley-type and Footbridge-type dilemmas was carried out by Pletti et al. (2016). Several decision-making models, such as the regret theory (Bell, 1982; Loomes, G., & Sugden, 1982) and the decision affect theory (Mellers et al., 1999), suggested that people anticipate how they would feel after having chosen each of the different options, and then select the alternative that minimizes the emotional cost. On this basis, Pletti et al. (2016) considered in their research the emotions reported by participants not only after their decision but also after imagining to have chosen the alternative option (the “counterfactual emotions”). Their findings showed that the difference between emotional intensities related to the two alternatives predicted the decision in both moral dilemma types, highlighting how, even in Trolley-type dilemmas, participants chose the option with the lowest emotional cost (Pletti et al., 2016). More in detail, the idea of not intervening and letting die people in Trolley-type dilemmas elicited higher levels of anger and "inaction regret" (the Italian "*rimpianto,"* the regret experienced for not having done a certain action), driving the decision towards utilitarian resolution (Pletti et al., 2016).

In conclusion, conflicting evidence emerged on the dual-process model of moral judgments (Greene et al., 2001, 2004) from behavioral data and subjective emotional evaluations. Specifically, two main points are contrary to the model: first, the fact that the emotional response may play a role in judging in a utilitarian way (Choe & Min, 2011; Manfrinati et al., 2013; Pletti et al., 2016), and second, that emotions seem to modulate the choice even in Trolley-type dilemmas (Pletti et al., 2016). Regarding this last point, it could be speculated that the emotional system plays different roles in the two moral dilemmas. In Footbridge-type dilemmas, an immediate adverse reaction is elicited by the idea of directly killing a man and contrasts with a rational cost-benefit analysis, as theorized by the dual-process model (Greene et al., 2001, 2004). In the Trolley-type ones, the strongest emotional response seems to be related to the idea of letting die a higher number of people rather than indirectly causing the death of only one person (Pletti et al., 2016). In this case, the emotional and the cognitive system appear to drive the decision in the same direction toward a utilitarian resolution.

1.1.2. Neuroimaging, Electrophysiological and neuropsychological data

The first neuroimaging study that compared impersonal (Trolley-type) and personal (Footbridge-type) moral dilemmas highlighted in the latter a greater activation of brain areas associated with emotional processes, such as the medial frontal gyrus (BA 9 and 10), the posterior cingulate gyrus (BA 31), and the superior angular gyrus/temporal sulcus (BA 39) (Greene et al., 2001). On the other hand, the Trolley-type dilemmas showed higher activation in the right dorsolateral prefrontal cortex (DLPFC, BA 46) and the parietal lobe (BA 7/40) (Greene et al., 2001), both known to be areas associated with working memory and therefore with cognitive processes (Cohen et al., 1997). These findings were replicated by the same authors in a successive study. They also found a previously unobserved bilateral increase in amygdala activity for Footbridge versus Trolley-type dilemmas (Greene et al., 2004). Furthermore, in “difficult” Footbridge-type dilemmas (i.e., dilemmas in which an extreme immoral action is favored by strong utilitarian considerations), the choice of the utilitarian option was associated with the engagement of areas related to the deployment of cognitive control and the detection of conflict (the DLPFC and the anterior cingulate cortex, respectively) (Greene et al., 2004).

However, Borg et al. (2006), despite confirming the main results of previous neuroimaging study (Greene et al., 2001, 2004), also found that some non-utilitarian responses (i.e., refrain from killing) could be mediated not only by emotion but also by reason (Borg et al., 2006).

Electrophysiological studies mainly aligned with the dual-process model (J. D. Greene et al., 2001, 2004) and helped understand the temporal dynamics of emotional and cognitive processing underlying moral decision-making. A larger P260 component was found in the frontal and frontopolar regions in Footbridge-type dilemmas compared to Trolley-type dilemmas, possibly reflecting an immediate emotional reaction during the early stage of preference formation between the two alternatives (Sarlo et al., 2012). The authors suggested this neural event could represent an “alarm-bell” emotion that signals aversion to intentional harm and pushes the decision toward non-utilitarian resolutions (Sarlo et al., 2012, 2014). In addition, this component showed a positive correlation with “personal distress” (i.e., an affective dimension of empathy), suggesting that the possible motivation behind this "alarm bell" emotion could be a selfish motivation to alleviate the distress related to the idea of killing a man, rather than an altruistic feeling of concern for the victim (Sarlo et al., 2014).

The neuropsychological studies in this field focused on understanding whether emotional processes play a causal role in moral decision-making, filling the gap left by purely correlational fMRI and electrophysiological works (Ciaramelli et al., 2007; Koenigs et al., 2007). To this aim, they recruited patients with bilateral damage in the ventromedial prefrontal cortex (vmPFC), a brain area that contributes to the generation of emotions and, specifically, social emotions (Beer et al., 2003; Damasio et al., 1990). VmPFC patients, compared to healthy controls, showed a selective increase in utilitarian responses in Footbridge-type moral dilemmas while no significant difference was found in Trolley-type or non-moral dilemmas (Ciaramelli et al., 2007; Koenigs et al., 2007).

The same pattern was obtained when their performance was compared to that of patients with brain damage in structures unrelated to emotions (Koenigs et al., 2007).

On this basis, the ventromedial prefrontal cortex appears to be essential to oppose a "personal and direct" moral violation (Ciaramelli et al., 2007). Previous works already suggested that this region plays a crucial role in predicting the long-term emotional consequences of actions during decision-making and that it contributed to preventing their implementation (Bechara, 2005). This mechanism may work although the presence of immediate costs: in Footbridge-type dilemmas, judging a moral violation inappropriate involves as immediate cost the death of a certain number of people but, in the long term, it avoids the prospect of feeling responsible for killing a man (Ciaramelli et al., 2007). In patients with vmPFC injury, the absence of an emotional reaction to the idea of sacrificing someone probably leads them to rely on explicit norms that support the promotion of the greater overall good (Koenigs et al., 2007).

These findings highlighted how emotions are necessary to avoid killing in Footbridge-type dilemmas but, to test the double-process model properly, a double-dissociation between "utilitarian" patients with vmPFC lesions and "anti-utilitarian" patients with DLPFC lesions should be proven (Moll & de Oliveira-Souza, 2007). If cognitive control is necessary to override the emotional tendency to avoid personal moral violations, the lesser amount of cognitive control derived from a DLPFC injury should result in a decrease in utilitarian moral judgments. Although it seems to be no studies with DLPFC patients in this field, direct modulation of brain areas by non-invasive electrical or magnetic stimulation represents a valid alternative to evaluate a causal relationship (Kuehne et al., 2015).

Results regarding the causal role of DLPFC (and, consequently, cognitive processes) are mixed. Jeurissen et al. (2014) used transcranial magnetic stimulation (TMS) to alter the activity of the right DLPFC or the right temporoparietal junction (TPJ), an area associated with emotional processing and theory of mind (Saxe & Kanwisher, 2003; Young et al., 2010). Their results supported the dual-process theory (Greene et al., 2001, 2004): in the case of DLPFC disruption, non-utilitarian responses in Footbridge-type dilemmas increased, while by stimulating the TPJ, a greater number of non-utilitarian responses in the Trolley-type moral condition were recorded (Jeurissen et al., 2014). However, Tassy et al. (2012) found a very different pattern: increased utilitarian responses when suppressing the right DLPFC with rTMS, especially in high-conflict moral dilemmas. Similarly, increasing the activity of left DLPFC by anodal transcranial direct current stimulation (tDCS) resulted in a decrease in utilitarian moral behavior (Kuehne et al., 2015).

To summarize, neuroimaging, electrophysiological and neuropsychological studies outlined a complex picture. On the one hand, they showed how emotional processes are mainly involved in Footbridge-type dilemmas (Ciaramelli et al., 2007; Greene et al., 2001, 2004; Koenigs et al., 2007), acting in an early stage of decision-making as an "alarm-bell" that helped individuals to abstain from personal moral violation (Sarlo et al., 2012, 2014). Possibly, the reason behind this negative reaction is to prevent choice with a long-term emotional cost (Ciaramelli et al., 2007; Sarlo et al., 2012), avoiding the personal distress related to killing a man as a direct mean to save others (Sarlo et al., 2014). On the other hand, the findings did not exclude that non-utilitarian responses could also be mediated by reason in some cases (Borg et al., 2006), and, more generally, they did not unambiguously demonstrate the role of cognitive processes in driving the decision toward utilitarian resolution (Kuehne et al., 2015; Tassy et al., 2012). Together with studies using behavioral data and subjective emotional evaluations, these findings suggest that emotion and cognition do not always interact in the rigid competitive way outlined by Greene's dual-process model. Their relationship could be more complex, assuming different roles in relation to deontological and utilitarian responses (Cushman et al., 2010).

1.2. Cushman's version of the dual-process model

As described in the previous paragraphs, Greene's dual-process model (Greene et al., 2001, 2004) showed some limitations. More recent approaches criticize the simple contraposition between cognitive and affective systems (Cushman et al., 2010; Huebner et al., 2009; Moll et al., 2008) without necessarily contesting the existence of distinct processes that contribute to moral judgment. Cushman's version of the dual-process model recognizes that two systems compete (and, sometimes, complement each other) in reaching a moral decision, but it states that both have to involve some affective and cognitive elements (Cushman, 2013). In simpler words, the system that pushes the decision toward utilitarian resolution does not simply limit its role to the representation of the factual content, “five lives are more than one life,” but possibly it conveys the affective message “deciding to save five lives is better than the choice to preserve one.” Similarly, some cognitive aspects, in terms of information processing, have to also be involved in the mechanism responsible for the deontological response (Cushman, 2013).

More in detail, Cushman’s distinction can be framed in terms of action- vs. outcome-based evaluation. One process assigns value directly to the action (e.g., negative value to the representation of pushing the man off the bridge), while the other selects an action on the basis of its expected outcome (e.g., negative value to the representation of the damage inflicted to the person pushed) (Cushman, 2013).

A series of works provides some evidence for such division in the moral field (Miller et al., 2014; Miller & Cushman, 2013). In those studies, participants were first asked to evaluate their negative emotional state concerning different situations created in a way that dissociated outcomes from actions. For example, “see a person stepping on broken glass shards” put the focus on the painful outcome, while “fire a bullet at a friend, with his consent, while he is behind bulletproof glass” has no harmful consequences but represents an adverse action (Miller et al., 2014). Then, participants had to judge the acceptability of third-party harmful behavior in moral dilemmas. Deontological judgments, particularly in Footbridge-type dilemmas, were strongly predicted by the aversion to committing harmful actions but not by aversion to harmful outcomes. The result was maintained even with a delay of two years between the two tasks (Miller et al., 2014). Other works found evidence that some types of violent actions possess an intrinsic aversive quality independent from their consequences. Cushman et al. (2012) highlighted how participants exhibited an increased physiological aversion when they had to perform simulated harmful actions (e.g., hit a person’s fake leg with a hammer) compared to witnessing the experimenter executing the same actions, even knowing that no harm could occur in either case. Similarly, significantly greater aversion to performing simulated harmful actions was also shown in comparison with the execution of “metabolically matched” actions. (e.g., using a hammer to hit a wooden block) (Cushman et al., 2012).

Based on this, the model hypothesized that choice in Footbridge-type dilemmas, in which a harmful action is a sub-task necessary to reach the output “saving more lives,” is mainly guided by action-based value representations (Cushman, 2013), with the result that the decision is pushed away from non-utilitarian resolution. Differently, maximizing lives in Trolley-type dilemmas does not require an action to which it is possible to assign an intrinsic negative value (e.g., flipping a switch). For this reason, these kinds of moral dilemmas are probably more influenced by outcome-based value representations, facilitating the choice of the utilitarian option. As previously reported, both action- and outcome-based processes have some affective contents responsible for motivating the respectively endorsed decision.

On this basis, if we follow this version of the dual-process model (Cushman, 2013), enhancing the role of emotional processes should increase people's motivation to choose the utilitarian resolution, especially in Trolley-type dilemmas. On the other hand, if Greene's dual-process model (2001, 2004) is correct, higher emotional processing should be related to a reduced proportion of utilitarian choices in both types of moral dilemmas (or, at least, no change should occur in Trolley-type dilemmas). In the next chapter, studies using emotionally activating Virtual Reality (VR) simulations of moral dilemmas will be presented, shedding some light on the role of emotion and the applicability of those models in "real" decision-making.

2. Virtual Reality: from moral judgment to moral actions

2.1. Definition and key concepts of VR

It is not an easy task to define VR. Several authors have made various and heterogeneous proposals, with the result that a precise and unified definition is still missing (Kardong-Edgren et al., 2019). Some of them focused on the technological aspects; for example, Riva (2002) defined VR as "A collection of technologies that allow people to interact efficiently with 3D computerized databases in real-time using their natural senses and skills."

However, a substantial part of the existing definitions is more oriented to delineate VR in terms of the psychological effects elicited during the interaction with it. For example, Schroeder (1996) described VR as "A computer-generated display that allows or compels the user (or users) to have a sense of being present in an environment other than the one they are actually in and to interact with that environment." Similarly, Sherman & Craig (2003) define it as "A medium composed of interactive computer simulations that sense the participant's position and actions and replace or augment feedback to one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world)." The concepts of "immersion" and "presence", often incorporated into the definitions of VR, play a fundamental role in the understanding of this medium.

2.1.1. Immersion and Presence

Immersion and presence are intimately connected constructs. Immersion is related to the technological capability of a system (Slater & Sanchez-Vives, 2016). It refers to "the degree of physical stimulation impinging on the sensory systems and the sensitivity of the system to motor inputs" (Bohil et al., 2011). The range and number of motor and sensory channels engaged by a virtual scenario directly influence the degree of immersion. The more accurately a system responds to motor inputs (e.g., precise tracking of head movements, unambiguous commands through hand gestures) and provides extensive sensory stimulation (e.g., higher fidelity of the visual display, use of three-dimensional spatialized sound), the higher the level of immersion elicited in the user (Bohil et al., 2011).

Presence is instead considered the psychological product of immersion (Bohil et al., 2011; Slater & Sanchez-Vives, 2016). Different definitions can be found, however, a commonly used one defines it as “the sensation of ‘being there’, in the virtual space, instead of the physical environment” (Biocca, 1997; Lombard & Ditton, 1997; Slater & Sanchez-Vives, 2016). Importantly, this phenomenon occurs despite the user being cognitively aware that the virtual simulation s/he is experiencing is not real (Slater & Sanchez-Vives, 2016). Other approaches emphasize instead how the sense of presence depends on interacting with the virtual environment (Gamberini & Spagnolli, 2015) and receiving timely feedback from it (Slater, 2009).

However, beyond its different definitions and conceptualizations, the construct of immersion has the practical implication of classifying the VR technologies in “immersive”, “semi-immersive” and “not-immersive” systems (Bamodu & Ye, 2013).

2.1.2. Classification of VR systems

It is possible to classify the VR systems into three major categories, on the basis of the level of immersion and type of interfaces utilized (Bamodu & Ye, 2013):

Non-immersive. These systems, also known as "desktop-VR" (Feng et al., 2022) or “low-immersive VR” (Martirosov et al., 2021), allow users to visualize 3D virtual environments through a 2D monitor (generally a high-resolution display, but also the screen of smartphones and tablets) (Figure 2.1). Mouse and keyboard are the most commonly used devices to interact with the simulation. Compared to other VR systems, they are relatively economic but less immersive (Bamodu & Ye, 2013).

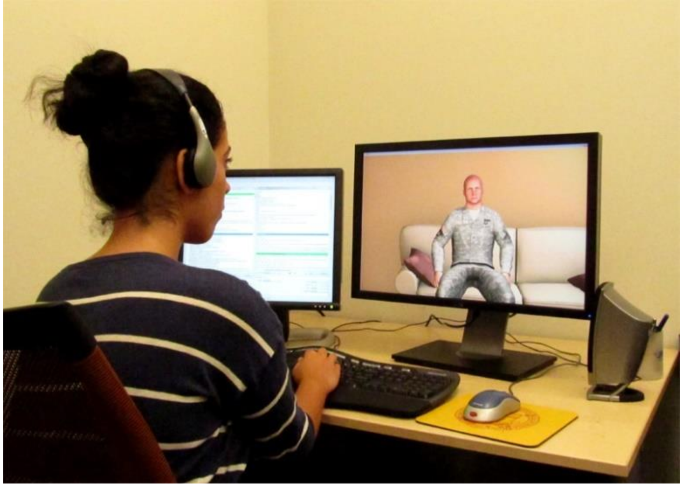


Figure 2.1. Example of desktop-VR simulation system.

Semi-immersive. A semi-immersive VR system typically includes a high-performance graphics processing system coupled to a large monitor or projector. Compared to a desktop-VR system, it has the advantage of being more immersive thanks to the involvement of a larger portion of the user's field of view, although in some cases the use of projection devices can limit the quality of the image. Some authors also include in this category the CAVE (Cave automatic virtual environment, Cruz-Neira et al., 1993), a system composed by three or more walls with images projected onto each, in which the user experiences 3D stereovision by wearing stereoscopic LCD shutter glasses synced with the projector (Figure 2.2).

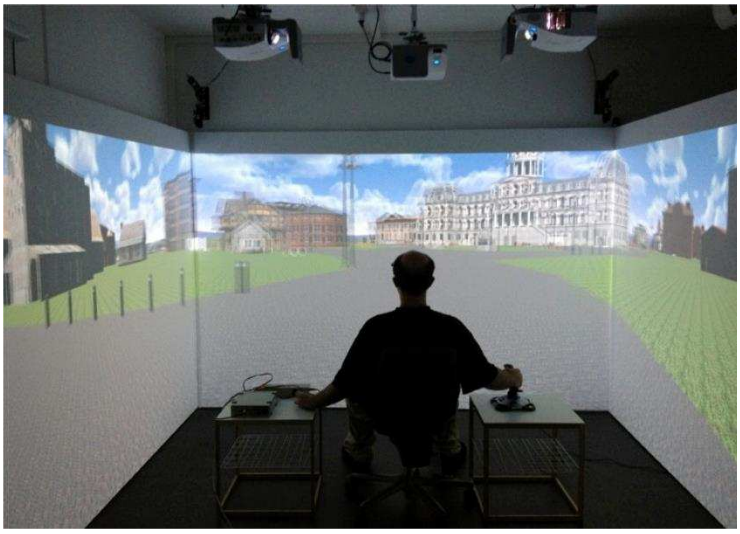


Figure 2.2. Example of a CAVE system.

Immersive. These systems, also called "fully immersive" VR (Martirosov et al., 2021), are probably the best-known VR tools. They include Head Mounted Display (HMD) (Figure 2.3), and, for some authors, CAVE (despite others, such as Slater & Sanchez-Vives (2016) considering this last one a less immersive system). The HMD uses small monitors positioned in correspondence of each eye in such a way as to provide stereoscopic images. Interaction with the virtual environment generally occurs through the use of controllers or, more recently, through hand gestures.

Figure 2.3. Example of an immersive VR system (HMD).

2.2. Moral dilemmas in Virtual Reality

In recent years, VR has revealed to be the ideal tool to study dangerous situations that cannot be safely reproduced in the real world or difficult to recreate in a laboratory setting for practical or ethical reasons (Blascovich et al., 2002; Slater & Sanchez-Vives, 2016) such as training for risky jobs (Pulijala et al., 2018), safety procedures (Chittaro & Buttussi, 2015) and emergencies (Spagnolli et al., 2021). This technology guarantees both ecological and internal validity, allowing the researchers to maintain methodological rigor and generalizability at the same time (Rovira, 2009). It was only a matter of time before VR was also applied to the field of moral decision-making.

The first virtual transposition of a moral dilemma was made by Pan & Slater (2011). In their pilot study, they used an alternative version of the trolley dilemma, in which the participant had to operate the lever of a platform to allow visitors to move between the two floors of an art gallery. At one point, when there were five people on the upper floor and one on the lower floor, a seventh visitor still on the platform began firing on the people on the upper floor. The participant then had to choose between letting the 5 visitors die or bringing the elevator down, sacrificing the visitor on the ground floor. The authors compared a semi-immersive VR version of this dilemma (presented in a CAVE-like system) with a non-immersive one (desktop version). Interestingly, they chose to adopt a minimalist graphic, in which the visitors were portrayed as stick figures.

Despite the presence of a trend, the percentage of utilitarian responses between the CAVE (89%) and desktop (67%) conditions was not significantly different, however, it should be considered that, being a pilot experiment, the sample size was small. Moreover, participants in the CAVE exhibited greater panic than those in the desktop condition, shown by the fact that a higher percentage of them pressed the button controlling the elevator multiple times (Pan & Slater, 2011). A more complex version of this pilot work was later implemented using immersive VR (Friedman et al., 2014).

In the subsequent study on the subject (Navarrete et al., 2012), an immersive VR version of the classic trolley problem was used. The focus was on comparing an "action" condition (participants could act to save the five avatars) and an "omission" condition (if participants did not act five would be saved). In both cases, about 90% of the subjects opted for the utilitarian resolution. However, those who had to actively perform an action to save the five people exhibited greater physiological arousal (as measured by skin conductance response and level) than those who could obtain the same result by doing nothing (Navarrete et al., 2012). Furthermore, higher levels of arousal were negative related with a propensity to choose the utilitarian outcome, in line with Greene's dual-process theory (Greene et al., 2001, 2004). Despite the more adequate sample size than the previous pilot study, even in this work there is no direct comparison between dilemmas in virtual and textual form (the only one was with data of previous surveys conducted by other authors).

The comparison between action and omission conditions was expanded upon by a subsequent study, which also included a VR version of the Footbridge dilemma (McDonald et al., 2017). Again, a higher level of arousal emerged in the action condition. Furthermore, greater emotional activation and a tendency to prefer the deontological resolution were found in the Footbridge dilemma compared to Trolley one, suggesting that even in its virtual transposition the Footbridge dilemma elicited a stronger emotional response than the Trolley problem (McDonald et al., 2017).

Instead, the work of Skulmowski et al. (2014) was the first to harness the potential of immersive VR to create versions of the Trolley dilemma in which, in addition to the number, other characteristics of the potential victims (the gender, ethnicity, and body orientation) were manipulated. Besides the classic preference for utilitarian responses, a greater tendency to sacrifice male avatars over female ones was found, while ethnicity and body orientation did not seem to have an influence on the choice. The study also measured the emotional arousal on a physiological level as measured by pupil diameter, finding a specific temporal signature displaying a peak in arousal around the moment of decision (Skulmowski et al., 2014).

Only with the studies of Patil et al. (2014), Niforatos et al. (2020) and Francis et al. (2016, 2017, 2018, 2019) a clear comparison between text and VR dilemmas was done. In the first one, the Trolley dilemma was used again, adapted in this case for desktop-VR administration. Interestingly, a within-subject design was adopted, whereby participants addressed the dilemmas in both virtual and textual form in two separate sessions approximately 100 days apart to avoid spillover effects (Patil et al., 2014). A clear discrepancy emerged between the two methods of administration, with participants choosing the utilitarian option significantly more often and exhibiting a higher skin conductance in the VR condition than in the text one. Importantly, the authors also compared the moral dilemmas with control situations, finding a significant difference in arousal only in VR condition. This implies that the increase in emotional activation found in virtual dilemmas versus textual versions was not a simple effect of the presentation mode but a combination of modality and moral content (Patil et al., 2014). An increase in utilitarian decisions in Trolley-like virtual dilemmas was also found by Niforatos et al. (2020), using, in this case, an immersive-VR setup. In addition, they also tested different type of victims, finding a preference to spare children (Niforatos et al., 2020).

To the best of my knowledge, the only works that compared virtual and textual versions of the Footbridge dilemma were the ones of Francis et al. (2016, 2017, 2018, 2019). The findings were similar to studies using the Trolley problem: participants more often chose to kill the man to save the five workers in immersive-VR than in the text condition. Physiological results went in the same direction, with an increase in arousal (measured, this time, by heart rate) for the virtual dilemmas compared to the textual versions (Francis et al., 2016, 2018, 2019). Similar to Patil et al. (2014), control situations were used to assess the role of the content in addition to the presentation method. The difference in heart rate between control tasks in VR and textual administration was not significant, suggesting that the VR modality alone was not responsible for this increased arousal, but the presence of a moral content also played a role (Francis et al., 2016).

Instead, the attempt to make the virtual scenario even more realistic, adding haptic feedback (e.g., with a robotic manipulandum or interactive life-like sculpture) did not show an effect on the choice compared to the "simple" VR scenario (Francis et al., 2017).

To summarize, the adoption of VR technologies (both immersive and non-immersive) has made it possible to study in a more ecological way how people act in a situation of moral dilemma and their emotional reactions. Part of the studies on the subject analyzed the differences between textual and virtual versions of the dilemmas, unexpectedly finding an increase in utilitarian responses (Francis et al., 2016, 2017, 2018, 2019; Niforatos et al., 2020; Patil et al., 2014) and physiological arousal (Francis et al., 2016, 2018, 2019; Patil et al., 2014) in the virtual presentation modality. These latest results seem to be in line with Cushman's version of the dual-process model (Cushman, 2013, see chapter 1). These authors suggested that the contextual saliency (e.g., see the potential victims) of the virtual versions of Trolley and Footbridge dilemmas increased the negative value assigned to the outcome-based representations for not acting and allowing the people on the tracks to be killed, increasing utilitarian resolutions in both types of dilemmas (Francis et al., 2016; Patil et al., 2014). On the contrary, in the textual versions, the reliance on imagination and the absence of salient features has perhaps led to attributing a more negative value to the action of harming and less to the outcome of not acting (Francis et al., 2016; Patil et al., 2014).

Other works on the topic have instead used only virtual versions of the dilemmas, exploiting the advantages that this medium offers to go beyond the classic Trolley problem and investigate the effect of manipulating different characteristics of potential victims (Niforatos et al., 2020; Skulmowski et al., 2014). These studies, and the concomitant introduction of the first autonomous and semi-autonomous vehicles, opened the door to a new line of works in which the Trolley dilemma (suitably modified) was used to investigate driver preferences in unavoidable crash situations.

These works were discussed in the next chapter.

3. Autonomous Vehicles: moral dilemmas in real contexts

3.1. Autonomous and semi-autonomous vehicles

Fully autonomous cars can be defined as vehicles that “can operate without human control and do not require any human intervention” (Bimbraw, 2015) and that are “capable of sensing their local environment, classifying the types of objects that they detect, reasoning about the evolution of the environment and planning complex motions that obey the relevant rules of the road” (Campbell et al., 2010).

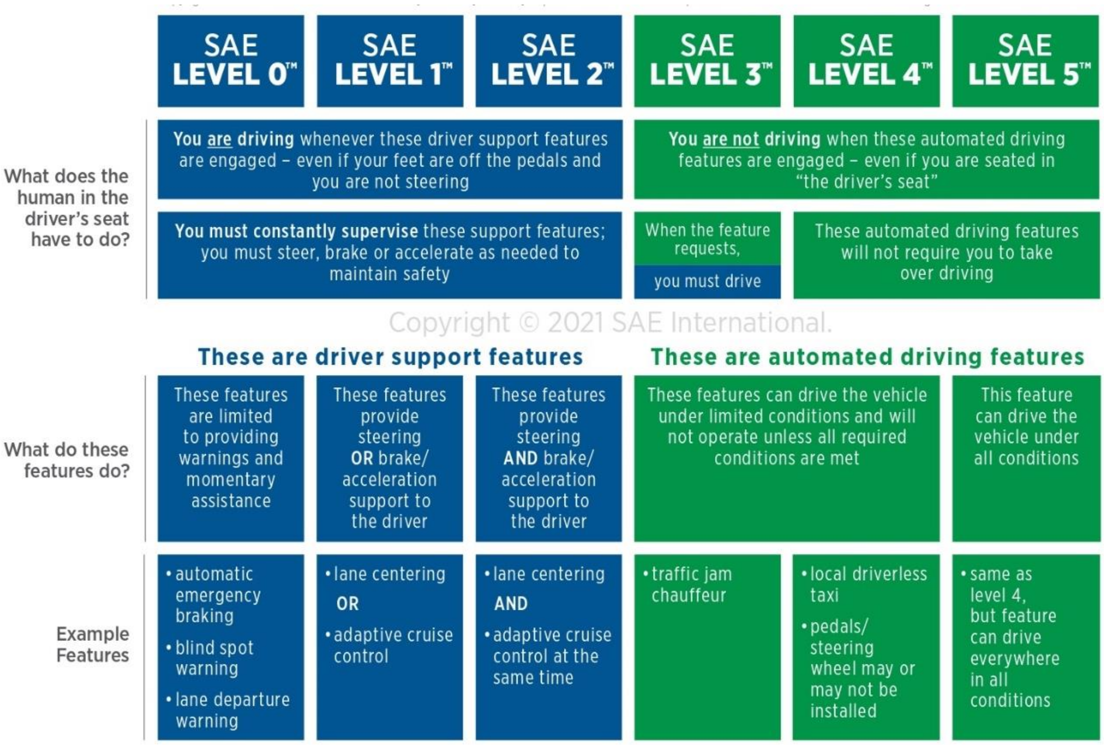
However, the concept of autonomous driving cannot be summarized in the simple dichotomy “presence/absence” of an autonomous driving system but includes several self-driving levels. Among the best-known classifications of automation levels is the one made by the Society of Automotive Engineers (SAE), which includes six different levels (Figure 3.1.). Their taxonomy ranges from no driving automation (Level 0) to full driving automation (Level 5), according to the following scheme (SAE, 2021):

1. Level 0: No Driving Automation
2. Level 1: Driver Assistance
3. Level 2: Partial Driving Automation
4. Level 3: Conditional Driving Automation
5. Level 4: High Driving Automation
6. Level 5: Full Driving Automation

The most relevant transition is between SAE Level 2 and Level 3, where there is a shift from human driving (even if facilitated by support features) to a situation where the car drives autonomously (at least as long as the automated driving features are engaged). However, in SAE Level 3, the driver still has the responsibility to intervene if the automated system requires it (SAE, 2021).

Although there are no products yet available on the market with the highest levels of automation (e.g., SAE Level 4 and 5), substantial progress has been made in the development of AVs in recent years. The introduction of these technologies is expected to result in a significant reduction in the number of accidents, as they are largely caused by human error (Fleetwood, 2017; Gao et al., 2014). Other expected benefits include a reduction in traffic congestion (Van Arem et al., 2006), a decrease in pollution levels, and an increase in mobility of the physically impaired (Gao et al., 2014).

Beyond these benefits, the introduction of vehicles in which control is delegated in whole or in part to the driving system has raised several ethical questions, including how the vehicle should behave in the event of an unavoidable collision with one or more human targets (Bonnefon et al., 2016; Goodall, 2014). And it is on this basis that the Trolley dilemma (suitably modified and adapted to the driving context) has been rediscovered and used to investigate what is acceptable for human drivers dealing with extreme accident situations in order to create a bottom-up framework for the "decisional system" of AVs.



	SAE LEVEL 0™	SAE LEVEL 1™	SAE LEVEL 2™	SAE LEVEL 3™	SAE LEVEL 4™	SAE LEVEL 5™
What does the human in the driver's seat have to do?	**You are driving** whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			**You are not driving** when these automated driving features are engaged – even if you are seated in "the driver's seat"		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
	These are driver support features			**These are automated driving features**		
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering **OR** brake/ acceleration support to the driver	These features provide steering **AND** brake/ acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met		This feature can drive the vehicle under all conditions
Example Features	• automatic emergency braking • blind spot warning • lane departure warning	• lane centering **OR** • adaptive cruise control	• lane centering **AND** • adaptive cruise control at the same time	• traffic jam chauffeur	• local driverless taxi • pedals/ steering wheel may or may not be installed	• same as level 4, but feature can drive everywhere in all conditions

Copyright © 2021 SAE International.

Figure 3.1. SAE levels of driving automation (Copyright © 2021 SAE international).

3.2. Moral dilemmas in the driving context

The renewed interest in "Trolley dilemma"-like situations triggered by the introduction of AI-driving systems has led to two parallel lines of research. The first one aims at understanding the "cold" preferences of people. To this end, moral dilemmas were used in text form (Bonnefon et al., 2016; McManus & Rutchick, 2019; Zackova & Romportl, 2018) or with simple graphical representations (vignettes) (Awad et al., 2018; Awad, Levine, et al., 2020; Rhim et al., 2020; Yokoi & Nakayachi, 2020), often in a third-person perspective. Participants were not given a time limit to answer the dilemmas and generally were able to do the task from the comfort of their own home via online surveys (Awad et al., 2018; Bonnefon et al., 2016; McManus & Rutchick, 2019; Zackova & Romportl, 2018).

However, the use of these presentation methods to investigate moral judgments can lead to responses that are flawed by social desirability bias and do not reflect the actual preferences of users (Tan et al., 2021). Although some attempts have been made in more recent studies to control for this bias (e.g., Mayer et al., 2021), this remains an important problem. On the other hand, text/graphic versions of dilemmas have the advantage that they can be easily administered online, thus obtaining enormous sample sizes. For example, in the famous "Moral Machine experiment", Awad et al. (2018) collected 40 million decisions in ten languages from millions of people in 233 countries, finding some universal trends, such as saving humans instead of animals, and important cultural differences (e.g., the preference to spare the young instead of the old is much less pronounced for Confucian and Islamic countries) (Awad et al., 2018).

The second line of research aims to investigate not conceptual judgement but the “actual” behavior during unavoidable accidents, using VR to investigate users’ decisions in a more ecological and realistic way (Grasso et al., 2020). As reported in the previous chapter, the adoption of VR allowed researchers to finely manipulate different factors of the simulations (e.g., the level of graphical realism or the duration of the decision window) and characteristics of the potential victims. Among the first factors investigated are the number and age of victims. Several studies have clearly shown a preference for maximizing the number of lives, and saving younger people (especially children) over older ones (Bergmann et al., 2018; Faulhaber et al., 2019; Frison et al., 2016; S. Li et al., 2019; Wintersberger et al., 2017), consistent with the results of online surveys (Awad et al., 2018).

Some finer manipulations involve the likelihood of injury of the potential victims and their compliance with traffic laws. Regarding the first one, Bergmann et al. (2018) and Faulhaber et al. (2019) modified the position of pedestrians to make them more or less susceptible to injury in case of collision, revealing how participants actually tended to run over people standing as opposed to those kneeling (and thus at greater risk due to low head position). Another way to influence the probability of injury is by changing the type of target with which the collision may occur. For example, a pedestrian is more likely to be fatally injured, even at low speed, while a collision with another car might cause less damage. Li et al. (2019) have in fact found a greater tendency to run over a motorcycle or other car rather than a pedestrian. Interestingly, if the pedestrian was not obeying the rules of the road (i.e., was crossing at a red light) this trend reversed, highlighting a preference to spare the innocent motorcycle/car (S. Li et al., 2019). The observance of road rules therefore seems to be a relevant factor in the decision process, or at least hierarchically more important than the likelihood of injury, but less than the number of potential victims. Indeed, users preferred to run over an innocent pedestrian/s on the sidewalk to save a greater number of people in the middle of the road (Bergmann et al., 2018; Faulhaber et al., 2019; Kallioinen et al., 2019).

The level of realism of the simulation is another factor that has been manipulated. This has been done in several ways, including changing the time allowed to decide, making it more like that of a real accident situation. Most studies have allowed a time window of 4 seconds to act (Bergmann et al., 2018; Faulhaber et al., 2019; Sütfeld et al., 2017), which is enough time to recognize the different characteristics of potential victims but not as long to make complex reasoning. Shorter time frames (under 2 seconds) resulted in a decreased preference for saving the young over the elderly. However, this decrease in victim age bias appears to be due to the difficulty in recognizing the physical characteristics of targets in such a short time window (Sütfeld, Ehinger, et al., 2019). In the recent work of Lucifora et al. (2021), a "hot" (more ecological), and a "cold" (more conceptual) VR scenario were compared. In the hot one, participants had to react in real time to an accident situation and make a quick choice, while in the cold scenario the simulation stopped before the accident, allowing them to decide without time limits between the possible alternatives. In the strictly ecological situation, 92% of the participants chose to kill the child in the middle of the road (the cause of the accident was precisely the sudden crossing of the young pedestrian). In the cold scenario only 35% opted for this choice, while most chose to turn left and kill two construction workers, highlighting a strong dissociation between the two conditions, with a more utilitarian behavior in the ecological one (Grasso et al., 2020; Lucifora et al., 2021).

Another way to make the situation more realistic is to move away from the typical dichotomy of moral dilemmas, leaving more options for the user. For example, in the studies of Ju et al. (2019, 2016), the participants had to react to the sudden appearance of three pedestrians in the middle of the street. In this case, they could try to brake, attempt dangerous evasion maneuvers, or throw themselves off a cliff to avoid running them over. Although the other choices resulted in the death of the three pedestrians, only one participant chose to sacrifice himself in order not to kill them (Ju et al., 2019), suggesting that in more ecological situations people tend to behave less altruistically.

Among the other factors investigated, but which were not found to have a significant impact on the choice, are the familiarity with the potential victims (i.e., one of the targets is the driver's best friend) (Frison et al., 2016; Wintersberger et al., 2017) and the presentation mode (immersive VR vs. desktop VR) (Sütfeld, Ehinger, et al., 2019).

As seen so far, Trolley-type dilemmas have been extensively employed in the driving context, however, a number of criticisms emerged regarding their application in this sector. The main ones have been considered and discussed below.

Trolley-type dilemma situations are too improbable. AVs are supposed to drastically decrease the number of accidents (Fleetwood, 2017). Some might argue that the probability of an inevitable collision situation between two possible targets with no other option than killing one is so low that it does not deserve attention (Awad, Dsouza, et al., 2020). However, while it is difficult for such accident scenarios to occur, their consequences could have a huge impact on public opinion. Even now, the few fatal incidents involving AVs have received far more media coverage than the positive coverage of progress in their performance (Awad, Dsouza, et al., 2020). It is not difficult to imagine how even a single Trolley-type accident can impact the trust in AVs and consequently their adoption (Awad, Dsouza, et al., 2020; Othman, 2021)**,** thus making the study of these dilemmas definitely worthy of attention.

Trolley-type dilemma situations are too simple. A common criticism concerns the excessive simplicity of Trolley-type dilemmas (Awad, Dsouza, et al., 2020). In a real crash there are generally more options than killing one target than the other. However, these dilemmas need to represent a simplified situation in order to cleanly capture basic preferences and draw general conclusions beyond the highly specific set of circumstances of real (and more complex) accidents (Awad, Dsouza, et al., 2020). In any case, as discussed in this chapter, more realistic forms of dilemmas (in VR, with more options, limited time to decide, etc.) have already begun to be used (e.g., Ju et al., 2019; Lucifora et al., 2021).

Trolley-type dilemma situations are not ethical. Some argue that basing the decisional system of AVs on people's preferences in moral dilemma situations is unethical and leads to discrimination based on personal physical features such as age and gender (Kochupillai et al., 2020). How an AV resolves ethical trade-offs should be decided by ethics experts and policymakers, using a top-down rather than a bottom-up approach (Awad, Dsouza, et al., 2020). However, informing policy experts about people's preferences can help them best serve the public interest, regardless of whether they ultimately decide to accommodate those preferences. Knowing in which cases public opinion and regulations are not aligned is a first step in understanding potential problem areas and how to address them (Awad, Dsouza, et al., 2020).

To summarize, despite some criticisms, Trolley-type dilemmas represent a useful paradigm for studying moral judgment and moral action in the context of driving. However, research in this area is far from exhaustive, and there are still fundamental aspects that have not been adequately investigated. The present thesis project, described in the next chapter, aims to shed some light on these points.

4. The present project

As seen in the previous chapters, since VR technology was adopted in the moral dilemmas field, two different theoretical pathways emerged. The first tried to replicate classic Trolley and Footbridge dilemmas through virtual simulations (Francis et al., 2016, 2017, 2018; Friedman et al., 2014; McDonald et al., 2017; Navarrete et al., 2012; Pan & Slater, 2011; Patil et al., 2014; Skulmowski et al., 2014). Despite the augmented realism added by VR, situations represented remain not particularly ecological (Ramirez, 2018).

An alternative approach is to adapt the Trolley problem to a more ecological situation: driving simulations. As in the classic version, the subject has to choose who to kill but instead of diverging a trolley, the participant drives a car (Bergmann et al., 2018; Faulhaber et al., 2019; Frison et al., 2016; Ju et al., 2019, 2016; Sütfeld et al., 2017; Sütfeld, Ehinger, et al., 2019). In this type of studies, the experimental focus shifted from the study of the competition between emotional and cognitive processes to the "behavioral" output. Indeed, the promising introduction of the first autonomous cars created a renovated interest in "Trolley dilemma"-like situations, as a means to collect data about how people decide during extreme road accident situations. The general idea behind these works is to understand which are the typical choices of the drivers in order to create a bottom-up framework for the "moral system" of autonomous vehicles.

To reach this goal, different situations have been tested, changing number, sex, age, ethnicity and position of the potential victims (Bergmann et al., 2018; Faulhaber et al., 2019; Frison et al., 2016; Ju et al., 2019, 2016; Sütfeld et al., 2017; Sütfeld, Ehinger, et al., 2019). However, very little attention was paid to emotional processes in these new versions of the Trolley problem. Considering not just the behavioral output but also the entire experience of the "driver" can help to highlight the motivations behind those choices, allowing researchers to test classic moral judgement theories in more "concrete" and realistic situations and, last but not least, can help obtain useful insight on possible effects of the introduction of autonomous-driving technology.

On this basis, the main goal of the present Ph.D. thesis was to investigate the behavior and emotional reactions in these "driving versions" of the Trolley dilemma, using VR technology to recreate these accident situations in a more ecological way.

To this end, three studies have been conducted to explore the following three aspects:

1. **The eventual presence of differences in valence and arousal in human driving vs. autonomous-driving simulations (study 1).** In fact, not only we are not sure about what the participants' emotional reactions in these "driving versions" of the dilemma are, but also we do not know whether they realistically reflect the emotions experienced in an autonomous vehicle confronting the same situations. To choose to kill a person or to passively assist to the same choice made by an autonomous system could elicit different effects and it is important to explore them.

2. **The influence of different legal frameworks on emotions and choices (study 2).** At the moment there is no clear standard for who is responsible during accidents involving autonomous cars. The liability could be on the driver, the motorist, the manufacturer or some proportion of these parts (Luetge, 2017). It is important to assess if and how different positions about legal responsibility affect emotional processes and moral decision-making.

3. **The effect of the presentation method adopted (text or VR) on decision and emotional reactions during unavoidable accident situations (study 3).** Previous studies with classic moral dilemmas found an increase in utilitarian choice and in the arousal dimension (using only physiological measures) in VR compared to the traditional text version (Francis et al., 2016; Patil et al., 2014). However, whether the method also plays a role in the new versions of the dilemmas applied to the driving context is unclear and requires further investigation.

The role of emotional processes in realistic moral dilemmas is still an open topic. Previously mentioned VR works did not analyze all the dimensions of emotion, particularly none of them assessed valence; arousal was measured in some of these studies but only at a physiological level (McDonald et al., 2017; Navarrete et al., 2012; Patil et al., 2014; Skulmowski et al., 2014). In the present project the introduction of self-report measures could help the in-depth examination of the role and nature of emotional processes.

To the best of the author's knowledge, only one study (Pletti et al., 2015) investigated the role of evaluating the legal consequences, but as limited to textual versions of the

Footbridge and Trolley problems. No one has assessed it in more realistic situations or in VR simulations. Similarly, only one study directly compared textual and virtual dilemmas applying them to the more ecological driving context, but its findings did not replicate the effect found in the virtual transposition of Trolley and Footbridge dilemmas. Furthermore, it did not even assess emotional reactions (Sütfeld, Ehinger, et al., 2019), making additional investigations on this topic highly needed.

To summarize, each of the three studies proposed permits to shed new light on moral decision-making in the driving context, exploring the effect on emotions and behavior of different factors: presence/absence of an autonomous driving system (Study 1), being or not being legally responsible (study 2), the use of VR/textual presentation method (Study 3).

5. Study 1 - Virtual Morality: Using Virtual Reality to Study Moral Behavior in Extreme Accident Situations[1]

5.1. Introduction

In the last years, Virtual Reality (VR) has become a widely employed technology to recreate reality-like scenarios in realistic but safe, interactive, and immersive 3D Virtual Environments (VE). Several works have shown that people respond to situations presented in VR as if they were real (Rovira, 2009; Slater et al., 2006). One of the greatest advantages of VR technologies is then the possibility of investigating in an ecological way human behavior in complex and dangerous situations that cannot otherwise be reproduced in the real world without incurring in unacceptable risks. This made VR the ideal tool for studying reactions to various types of emergencies such as fires (Gamberini et al., 2003; Kinateder et al., 2014; Ronchi et al., 2015), floods (Fujimi & Fujimura, 2020), earthquakes (Tarnanas & Manos, 2001), terrorist attacks (Shendarkar et al., 2008), correct understanding of safety procedures and trainings (Chittaro & Buttussi, 2015; Leder et al., 2019; Sacks et al., 2013), and, more generally, ways of acting in risky contexts. A particular case of emergency is represented by moral dilemmas, i.e., situations in which the user is forced to choose between two highly undesirable alternatives. In recent years, moral dilemmas have been adapted to more ecological and relevant scenarios: indeed, the promising introduction of AI-driving systems created a renovated interest in "Trolley dilemma"-like situations, as a means to collect data about how people decide during morally salient situations (Bergmann et al., 2018; Kallioinen et al., 2019; Sütfeld et al., 2017; Sütfeld, König, et al., 2019), in order to effectively inform the design and development of advanced technological systems in which AI must take the place of the human decision-maker. The main application is in the automotive sector: despite some guidelines for ethical decision-making for AVs (Luetge, 2017) already existing, it is still important to understand what is acceptable for human drivers dealing with extreme accident situations and what their reactions are in order to create a bottom-up framework for the "decisional system" of AVs.

[1] The material presented in this chapter has been partially published in Benvegnù, G., Pluchino, P., & Garnberini, L. (2021, March). Virtual Morality: Using Virtual Reality to Study Moral Behavior in Extreme Accident Situations. In 2021 IEEE Virtual Reality and 3D User Interfaces (VR) (pp. 316-325). IEEE, doi: 10.1109/VR50410.2021.00054.

Although the most famous study in this sector, "The moral machine" (Awad et al., 2018), has used pictorial representations of these “driving versions” of the dilemma to investigate moral judgment, a growing number of works are instead using VR systems to shed light on drivers’ moral actions rather than their conscious beliefs (Faulhaber et al., 2019). Additionally, the adoption of VR allows researchers to finely manipulate these simulations, by testing, for example, the effect of different types and numerosity of possible victims (Bergmann et al., 2018; Faulhaber et al., 2019; Skulmowski et al., 2014), their position (Bergmann et al., 2018), vehicle type (S. Li et al., 2019) and observance of road rules (Bergmann et al., 2018; Faulhaber et al., 2019; S. Li et al., 2019), various levels of time pressure (Sütfeld, Ehinger, et al., 2019), and different degrees of the lethality of the accident (Frison et al., 2016; Wintersberger et al., 2017)

However, despite the considerable efforts spent in understanding people’s preferences and behaviors in such situations, consumers remain skeptical about AVs (Edmonds, 2019) and are generally averse to the concept of machines making moral decisions (Bigman & Gray, 2018). In this light, two important clarifications need to be made:

First, the experience of being the actual driver making a choice in a situation of unavoidable collision could be different compared to the experience of being in an AV facing the same situation. A great part of the previously mentioned studies investigated how a human driver chooses, but this does not necessarily imply that the same choice made by an AV would be found equally acceptable and would have the same impact on the driver. More attention needs to be paid to possible differences between human and autonomous driving modalities. This could lead to a better understanding of the aversion toward self-driving vehicles.

Second, the majority of studies on moral choices in traffic dilemmas focused only on behavioral results (or on the self-reported acceptability of the decision). Considering the emotional reaction to these situations could help obtain a more exhaustive view of the driver’s experience. Despite classic moral judgment models recognizing the importance of emotional processes in decision-making (Cushman et al., 2006; Greene et al., 2001, 2004; Haidt & Hersh, 2001), the role of emotions in traffic moral dilemmas is still an open topic.

In the present chapter, a study in the automotive sector was described. To fill the previously reported gaps in the literature, I choose to investigate the behavioral and emotional reactions in two driving modes, an autonomous modality and a human one, through the manipulation of the level of interaction with the VR simulation. In addition to scenarios of moral dilemmas, control situations (non-moral dilemmas) were also included, with the dual purpose of making the simulation more ecological and of investigating the driver's experience in more common driving conditions. I chose to focus on two dilemmas: a revisited Trolley problem (a three versus one situation, from now on "Numerosity dilemma") and an "elder versus child" case (from now on "Age dilemma"). Finally, since VR simulations can suggest different possibilities for action because of their contextual richness, I have chosen to overcome the typical dichotomy of the dilemma by leaving the participant a third possibility: to not pick any of the two roads, avoiding the death of someone else and instead crashing their vehicle by going straight at an obstacle present at the end of the street he/she is traveling on. More in detail, I state the following hypothesis:

H.1 The interaction with the moral dilemma scenarios in the human driving mode increases the perceived arousal and unpleasantness more than experiencing the same scenarios in a more passive way in the autonomous driving condition.

H.2 The more interactive VR scenario, where participants actually drive the car, elicits a greater sense of responsibility than the autonomous driving system, always experienced in a VR environment.

H.3 The moral acceptability assigned to the actions is not influenced by the level of interaction with the scenario, and therefore by the driving mode, but only by the presence/absence of the moral content in the proposed choice situation.

H.4 Although sacrificing themselves in order to avoid run over anyone can be considered a socially desirable choice, the participants in the virtual scenario prefer to take actions that lead to saving the highest number of lives or the youngest potential victim, but without sacrificing themselves to save all parties.

In the following section, the relevant literature on the topic will be presented, framing each research hypothesis.

5.1.1. The role of emotion in virtual moral dilemmas

The first studies that applied VR technologies in this sector mainly dealt with comparing the effect of different methods of administering dilemmas. The basic idea of these works was not only to verify if people act consistently with the judgments expressed in the textual forms of the dilemma, but to use VR to test the most well-known theories in the field of moral judgment. Indeed, according to the dual-process theory (Greene et al., 2001, 2004), during the resolution of moral dilemmas, both cognitive and emotional processes come into play and compete. The former would lead to favoring utilitarian choices, maximizing the number of lives saved, even at the cost of sacrificing someone else. Emotional processes, on the other hand, would guide the choice in the opposite direction: the idea of killing a person, even if to save more lives, would elicit a strong negative emotional response that leads to the deontological decision not to kill (Greene et al., 2001, 2004).
On this basis, the use of interactive and activating virtual simulations of the dilemmas should have intensified the impact of emotional processes, leading to an increase in non-utilitarian resolutions. However, with the exception of the work of Navarrete et al. (2012), the other studies on the topic (Francis et al., 2016, 2017, 2018; Patil et al., 2014) did not confirm the dual-process theory, finding instead a concomitant increase in utilitarian choice and in the arousal emotional dimension (measured with electrodermal activity and heart rate) in VR compared to the traditional text version of the dilemmas. These results suggested that moral judgment and moral action could be distinct constructs (Francis et al., 2016, 2017; Patil et al., 2014). In the "arid" textual dilemmas, the dual-process theory (Greene et al., 2001, 2004) remains valid, but in the virtual versions of the dilemmas, the greater “contextual richness” offered by VR simulations makes the possibility of letting people die, rather than sacrificing just one to save them, more salient and emotionally aversive, probably motivating the participant to act in a way to minimize this distress (Francis et al., 2016; Patil et al., 2014).

From these works, it clearly emerges that emotional aspects play a key role in virtual moral dilemmas resolution. However, so far, the study of emotional processes has remained focused on comparing dilemmas in virtual and textual form, while the present study aims to offer a contribution on emotional aspects in moral dilemmas exclusively in virtual form, comparing two different driving conditions. At present, to the best of my knowledge, no work has analyzed the differences in participant's emotional reaction when dealing with moral dilemmas as a driver or as an occupant of a self-driving car, measuring self-reported arousal and emotional valence.

However, previous VR studies comparing classic Trolley dilemma with its "omission condition" variant (i.e., the runaway trolley is about to run over a single worker. If the participant pulls a lever the trolley will be diverted onto another track, where it will kill five workers) found a decreased physiological arousal when reaching the utilitarian resolution required to avoid to act (i.e., in the omission condition) (McDonald et al., 2017; Navarrete et al., 2012). Despite the difference between these studies and the current one, it is possible to hypothesize that the "omission" version of the Trolley had some similarities with the autonomous driving condition used in the present work because in both cases abstaining from action is required to maximize the number of lives saved.

On this basis, the necessity of actively sacrificing one person to save a higher number of potential victims (or the younger one), as in the human driving mode, can be expected to elicit a stronger and more aversive emotional response than being in the more passive situation of letting the car doing it (i.e., autonomous driving mode).

5.1.2. Responsibility and acceptability of choices made by human drivers and self-driving cars

According to the German Ethics Commission for Automated and Connected Driving, "in the case of automated and connected driving systems, the accountability that was previously the sole preserve of the individual shifts from the motorist to the manufacturers and operators of the technological systems and to the bodies responsible for taking infrastructure, policy, and legal decisions" (Luetge, 2017). People perception of responsibility and blame attribution seems to be coherent with this "responsibility diffusion" phenomenon. In J. Li et al. (2016) and Pöllänen et al. (2020), blame attribution to different road transport system actors following crashes was assessed through a survey. In accidents involving fully autonomous cars, vehicle users received low blame, while vehicle manufacturers and the government were highly blamed. In addition, McManus &

Rutchick (2019) found that users of manufacturer-programmed fully-autonomous vehicles were considered less responsible compared to those that manually drove the car or to those that chose the “behavioral style” (selflessly or selfishly) of their AV’s algorithms. To the best of my knowledge, only one VR study (Wilson & Theodorou, 2019) investigated the perception of “moral culpability” between a condition with a human driver and an autonomous driving condition facing moral dilemmas, finding that the AV was perceived as less culpable than the human driver. On this basis, I expect to find also in the present VR simulation an increased perceived responsibility in the human driving mode because of control on the vehicle seemingly paying a key role in responsibility attribution.

Concerning the perceived acceptability, despite a general aversion to letting cars make moral decisions (Bigman & Gray, 2018), when choices made by AVs or human drivers have been directly compared, no difference in acceptability emerged. These results were found in studies using textual dilemmas (e.g., Bonnefon et al., 2016) and, more recently, in a work using VR simulations of moral dilemmas (Kallioinen et al., 2019).

However, it is interesting to note, from a methodological point of view, that both VR works mentioned in this section used a "passive" human driving mode. In the first one, participants were told that they were to be a passenger sat in the driver seat of a car controlled by the experimenter (Wilson & Theodorou, 2019). In the second one, participants had no control over the car in both human and autonomous driving mode; the only difference between the two modalities was the absence of the steering wheel in the autonomous driving mode (Kallioinen et al., 2019). In the present study, I chose to let participants act while in the human driving mode in order to be closer to a driver’s experience and assess the perceived acceptability of the moral decision in a more realistic context.

5.1.3. Numerosity and Age dilemmas and current AVs guidelines

VR simulation of moral dilemmas has permitted to highlight similarities and discrepancies between how people act in these situations and the current guidelines for programming autonomous guide systems. In the present study, two of the most investigated dilemma situations were examined, involving the number and age of possible victims. Regarding the “numerosity” factor, as previously reported, a consistent number of studies have shown that people tend to maximize the number of lives saved (e.g., Bergmann et al., 2018; Faulhaber et al., 2019; Skulmowski et al., 2014) and I expect that participants will behave in a similar way in the present study. This tendency is also in line with guidelines promoted by the German Ethics Commission for Automated and Connected Driving: "In the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical, or mental constitution) is strictly prohibited" while "General programming to reduce the number of personal injuries may be justifiable." (Luetge, 2017).

Regarding the "age" factor, despite the prohibition to make distinctions between possible victims (Luetge, 2017), people showed a clear preference to sacrifice the older avatar in the event of unavoidable accident situations (Awad et al., 2018; Bergmann et al., 2018; Faulhaber et al., 2019; Frison et al., 2016; Sütfeld, Ehinger, et al., 2019). This tendency is particularly strong when the other avatar involved is a child, suggesting that the remaining lifespan of the potential victims influences moral decision-making (Bergmann et al., 2018). On this basis, the majority of the participants in the present study are expected to save the younger avatar when in human driving condition.

5.1.4. Immersion, realism, time constraint, and self-sacrifice in moral dilemmas

Differences between textual and virtual versions of the dilemmas were found in both immersive (Francis et al., 2016, 2017, 2018) and non-immersive (Patil et al., 2014) VR simulations. However, studies directly comparing moral dilemmas using immersive and desktop VR systems found in the former a more intense experience of panic, a greater number of errors (Pan & Slater, 2011), and more recently, a slight preference for saving female potential victims (Sütfeld, Ehinger, et al., 2019). From Pan & Slater (2011) and other works outside the field of moral dilemmas (e.g., Pallavicini et al., 2019), it therefore appears that immersive VR elicits greater emotional activation: since one of the objectives of the present study is the analysis of the emotional experience of participants, immersive VR was adopted.

The effect of other characteristics of a virtual environment, such as graphic realism, are less clear. Indeed, in the study of (Pan & Slater, 2011), despite the use of a simplified environment with stick figures, the participants reported being emotionally activated.

However, with the partial exception of (Sütfeld, Ehinger, et al., 2019), there is still no direct comparison in this field between a hyper-realistic virtual scenario and a simpler one. In the simulation used in the present work, the virtual environment was created in a graphically realistic way, but without showing potentially disturbing images (i.e., the screen goes black shortly before impact with an obstacle/human avatar), as the experience of violent scenes in VR, compared to other media, can be particularly stressful for the participant (Madary & Metzinger, 2016; Slater et al., 2020).

More efficient ways to increase the realism of virtual moral simulations include the incorporation of haptic feedback (e.g., using a robotic manipulandum or interactive life-like sculpture; Francis et al., 2017) or increasing the ecological validity of the situation represented (Ju et al., 2019). With respect to this last point, some works have tried to create situations as similar as possible to those of a real accident, without explaining in advance to the participants that they would have to face a moral dilemma and leaving short windows of time to react (Ju et al., 2019, 2016).

The presence of time constraints is an important factor because, if the time window is short enough, it allows researchers to investigate actions less mediated by cognitive processes, minimizing the effect of bias often present in classic surveys such as social desirability. Participants in the studies of Ju et al. (2019, 2016), for example, had just under two seconds to react to the sudden appearance of three pedestrians in the middle of the street. Some participants tried to brake, others potentially endangered pedestrians by not braking or carrying out dangerous evasion maneuvers, and only one participant sacrificed himself to avoid running over them. However, having too short of a time window may not allow the correct recognition of potential victims (Sütfeld, Ehinger, et al., 2019). In the present study, participants had 4 seconds to react (as in Bergmann et al., 2018; Faulhaber et al., 2019; Sütfeld et al., 2017), a time window capable of making the pressure felt but still long enough to correctly identify potential victims.

Differently from the studies of Ju et al. (2019, 2016), other works found instead that participants acted more selflessly and that the probability of self-sacrifice increased with the number of avatars saved by that choice (Bergmann et al., 2018; Faulhaber et al., 2019; Frison et al., 2016; Wintersberger et al., 2017). However, it is possible that this result is partially influenced by other factors, for example, in Frison et al. (2016) and Wintersberger et al. (2017), the simulation was blocked to allow the participant to choose without time limits, an aspect that could break the sense of presence and make the decision more subject to the social desirability bias. To minimize this risk, in the present study, the self-sacrifice option has been made less obvious by letting it happen as a "default" resolution in case the participant did not steer either left or right. I expect that with the use of the time constraint and the presence of more options, the simulation gained realism and the possibility of sacrifice became less salient, making people less likely to act in this altruistic way to be seen in a favorable light.

5.2. Method

5.2.1. Experimental Design

A 2 (driving mode) × 4 (dilemma type) within-participants design was employed. The two levels of the first variable were human driving and autonomous driving. The dilemma type variable consisted of 4 levels: two moral dilemmas and two non-moral control situations, respectively: three vs. one (Numerosity dilemma), elder vs. children (Age dilemma), one vs. empty road, and elder vs. empty road (non-moral dilemmas). In the autonomous driving mode, the car always chooses to kill the single person in the Numerosity dilemma, the elder in the Age dilemma, and to pick the empty road in the non-moral dilemmas (in line with the most frequent behavioral choices made by participants in previous studies, thus making the results of the two driving conditions comparable). Additionally, two training trials (one in human and one in autonomous driving mode) were employed at the start of the simulation to help participants to familiarize themselves with the technology and the task. The training task used a "closed road" sign vs. empty road situation instead of showing a human avatar, and the results of these tasks were not included in the analysis. The participants were exposed to all eight resulting trials, presented in a random order (except for the two training trials, always presented at the beginning of the simulation).

5.2.2. Participants

15 participants took part to the study (8 males; age M = 23.83, SD = 2.32). Participants were recruited from the University of Padua, and they voluntarily chose to participate in the experiment (no monetary incentive was provided). For one of the variables under consideration (Responsibility), the analysis was performed on only 11 participants due to

a possible bias linked to social desirability (for more information, see subsection 5.3.2). To participate, a native-level understanding of Italian, having an age between 20 and 35 years, and having possessed a driving license for at least one year were required. Exclusion criteria included having a history of migraines or motion sickness, having experienced previous car-related trauma, having had a seizure or history of epilepsy, and presenting vision problems that cannot be corrected with the use of glasses.

5.2.3. Material

The VR simulation was administered using an HTC Vive Pro HMD (resolution: 2880×1600, 1440×1600 per eye; refresh rate: 90 Hz; Field of view: 110 degrees; Hi-Res audio headphones). The HMD was connected wirelessly to an Intel computer with an i7 CPU with 32 GB of RAM and an NVIDIA GeForce GTX 1080 GPU. The virtual scenario (Figure. 5.1.) was built using Blender 2.79 and Unity 2019.2.9. Participants were sitting in an adjustable seat, placed in the same position and orientation of the virtual one. The level of interaction with the scenario was manipulated to create the two driving modalities: only in the human driving mode, the participant had the possibility to influence the direction of the car, pressing the right or the left button (or neither to go straight) positioned in the trackpad of the HTC Vive controller. During the post-trial questions, after selecting an answer with the trackpad, they had to confirm the choice by pressing the trigger button.

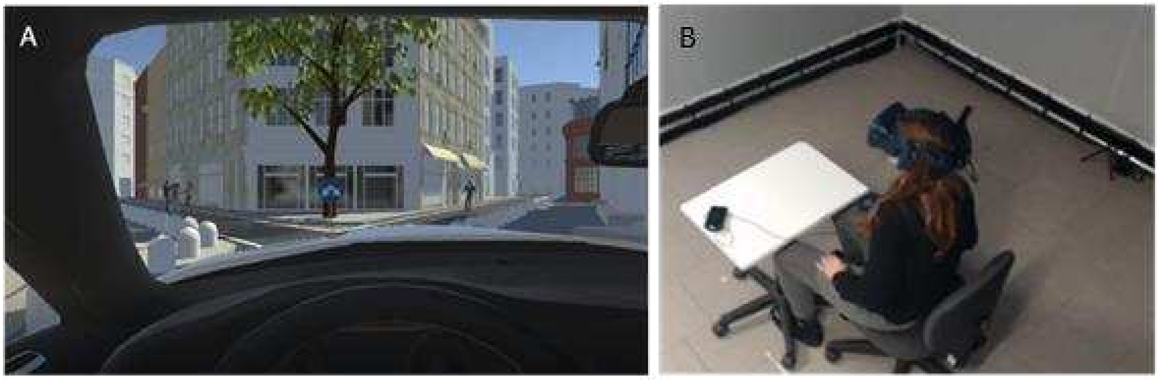


Figure 5.1. a) The virtual scenario from the point of view of the participant and b) a participant during the virtual session.

5.2.4. Measures

Behavioral Measures

For each participant in every trial in human driving mode the following data were collected:

1. Action (type of victim ran over).
2. Response times (RT).
3. Frequency of buttons pressed.
4. Duration of each press.

Self-Reported Measures

After each trial in both driving modalities, the perceived responsibility, acceptability, arousal, and affective valence related to the action made by the driver or by the car were assessed. For the first two variables, a 9-point Likert scale was employed. For the latter two, the self-assessment manikin (SAM) (Lang et al., 2008) was adopted. At the end of the virtual session, a series of questionnaires were administered in random order on a computer screen:

1. the Presence Questionnaire (PQ) (Witmer et al., 2005; Witmer & Singer, 1998).
2. the Marlowe-Crowne Social Desirability Scale -brief form- (MC-SDS-b) (Manganelli-Rattazzi et al., 2000)
3. A demographic questionnaire and post-experiment control questions.

5.2.5. Procedure

After the compilation of the informed consent and the assessment of inclusion and exclusion criteria, participants read the instructions for the task: they will have to drive a vehicle that could shift from a fully autonomous driving modality to a human one. The vehicle has a brake system failure, so it cannot stop but only turn. In the human driving modality (signaled by a flashing red light), they will have to act and choose the direction –left, right, or none- that the car will go towards. After every single action (made by them or by the vehicle itself in the autonomous driving mode) has been completed, a series of questions will be presented. Participants were further informed that they could stop the experiment without reason at any moment.

The virtual session started with the two training trials to familiarize the participant with the VR and usage of the controller. Participants were then exposed to each of the eight trials in random order. Similarly to Bergmann et al. (2018), Faulhaber et al. (2019) and Sütfeld et al. (2017), every trial started with the participant sitting in the driver seat of a car traveling at the constant speed of 36 kph (10 m/s) in a city (Figure. 5.2.). The human driving mode was signaled by a red light on the car head unit flashing for 5 s at the beginning of the respective trials. After 160-120 m (inconsistent distance was used to avoid habituation), a Y junction with two roads (one on the left and one on the right of a tree) appeared. In each road, avatars of the possible victims became visible 55 m before the impact point. In the human driving mode, participants had 4 s (40 m) to react and act to pick the direction of the car. Once the car had taken the junction (15 m from the impact), no other actions could be done. In both modalities, 5 m before the impact, the participant's sight was blocked by 3 s of black screen. The collision and death of the victim/s was signaled by the sound of an object hitting the hood of the car. When the black screen disappeared, participants answered the four questions regarding responsibility, acceptability, arousal, and valence directly in VR to maintain immersion. When the last question was completed, a new trial began after a brief delay. At the end of the VR session, a series of questionnaires were administered using a survey made with Google Forms on a computer screen. At the end of the experiment, participants were asked if they felt uncomfortable about the content of the VR simulations, but none reported any negative effects, and a second informed consent was compiled by the participants.

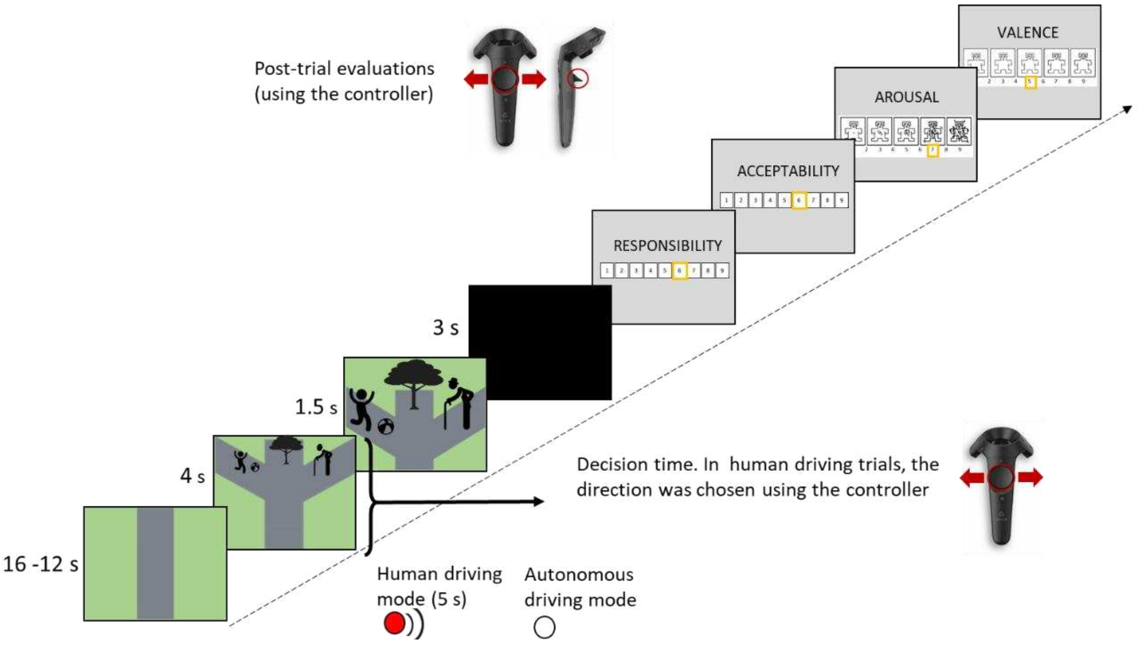


Figure 5.2. Graphical representation of the events of a trial.

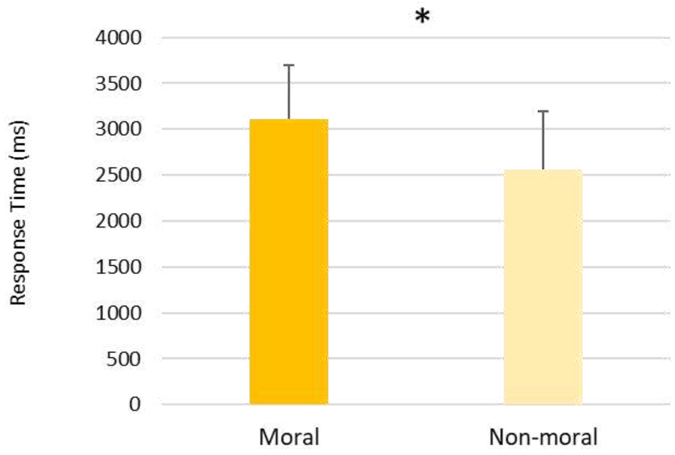
5.3. Results

5.3.1. Behavioral Results

In the human driving mode trials, the totality of the participants acted in a utilitarian way, preferring to sacrifice one man to save three lives. Concerning the "age" factor, the majority of the participants killed the older avatar. Of the 120 trials (8 for each participant), 6 concluded with an unintentional crash due to technical problems or participants failing to press the correct button during the decision time window, which they clearly reported in the post-experiment control questions. No one chose to sacrifice him/herself to save the pedestrians. There were two cases in which the child was killed instead of the elder, but they did not influence the frequency of the "typical" decision (Fisher's Exact Test, P =.22).

Two Wilcoxon signed-rank tests for paired samples were conducted to compare RT in the human driving mode in moral and control situations (Numerosity + Age vs. non-moral dilemmas) and in moral dilemmas only (Numerosity vs. Age dilemmas). Participants exhibited significantly longer RT when they had to act in moral dilemma situations compared to the control ones (V = 252, P < .001, mean ranks = 34.44, 20.36) (Figure 5.3.). No differences emerged between Numerosity and Age dilemmas. Only trials with typical results were included in these analyses, because trials with atypical results (i.e., killing the child instead of the old man) were insufficient (only two cases) to make a statistical comparison.

The same analyses were performed on the frequency of buttons pressed in the decision time window, and on the mean duration of those press, in moral and non-moral situations and then in moral dilemmas only. No significant results were found.



	Moral	Non-moral
M	3114.32	2562.07
SD	585.44	627.65

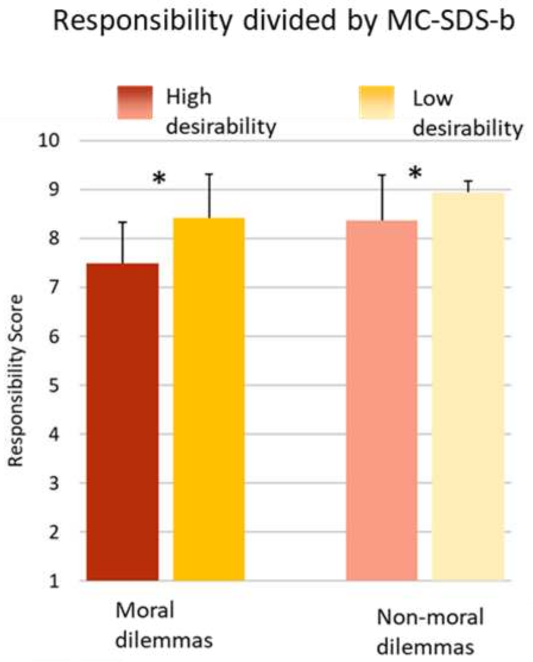
Figure 5.3. Mean and standard deviation of RTs for the typical choice in moral and non-moral situations.

5.3.2. Self-Reports Results

Social Desirability Control Check

Since the self-reported assessment of the perceived responsibility and acceptability in life/death decisions could be biased by social desirability, an initial control check on these two variables was performed. Particularly, it may be possible that people with a high tendency to respond to questionnaires in a way to present themselves in a favorable light (high social desirability) could minimize their responsibility in socially condemned acts (e.g., killing someone) to offer a better image of themselves. For this reason, participants were divided on the basis of their scores on the MC-SDS-b, an Italian validated brief form of the original MC-SDS, with 9-item and a dichotomic -true/false- response scale. In particular, the score of each participant was calculated by summing the number of items with "true" answers and used to assign the participant to one of the following two subsets: "high desirability" group (score> 5, with 5 being the average score) and "low desirability" group (score ≤ 5).

A series of Mann-Whitney tests were used to assess possible differences between the two groups. When in human driving modality, participants in the high desirability subset reported a significantly lower responsibility compared to participants in the low desirability one, in both moral (W = 26, P = .03, mean ranks = 5.5, 8.42) and non-moral dilemma situations (W = 53, P = .02, mean ranks = 6.37, 8.95) (Figure 5.4.). No differences were found in Acceptability scores or considering the distinction between Numerosity and Age dilemmas. As hypothesized, high desirability participants showed significantly lower responsibility scores in dilemma situations in which their actions involved the death of an individual than did low desirability participants. In the next section, to avoid biased results, the analyses of the responsibility scores were implemented in the low desirability subset only (11 participants) and not in the complete sample. In this way the results should be more realistic and less biased by the attempt to put oneself in a better light.



M	7.50	8.42	8.38	8.95
SD	0.84	0.90	0.92	0.22

Figure 5.4. Mean and standard deviation of Responsibility in human driving modality, in high and low desirability groups.

Post-Trial Evaluations Results

A series of Friedman rank-sum tests were conducted to analyze: 1) the general impact of the driving mode (human vs. autonomous) and the presence/absence of a moral dilemma (Numerosity + Age vs. non-moral dilemmas) on post-trial evaluations (Figure 5.5.); 2) the eventual presence of differences in post-trial evaluations between the two types of moral dilemma (Numerosity vs. Age). Comparisons with Wilcoxon signed-rank tests for paired samples with Benjamini-Hochberg correction were performed only when a Friedman test was significant. Only trials with typical results were included in these analyses, because, as previously reported, trials with atypical results were not sufficient to make a meaningful statistical comparison.

Regarding the impact of driving modalities and presence/absence of a moral dilemma situations, all the four DVs were significant (Responsibility: $\chi^2(3) = 47.92$, $P < .001$; Acceptability: $\chi^2(3) = 62.53$, $P < .001$; Arousal: $\chi^2(3) = 56.12$, $P < .001$; Valence: $\chi^2(3) = 60.82$, $P < .001$).

From Wilcoxon comparisons emerged an increased sense of responsibility in human driving mode compared to the autonomous mode in both moral (V = 0, P < .001, mean ranks = 8.42, 2.09) and non-moral decisions (V = 0, P < .001, mean ranks = 8.95, 1.18). In addition, participants reported an increased sense of responsibility when their self-driving car killed the pedestrian, compared to when the car chose an empty road (V = 48, P = .04, mean ranks = 8.95, 2.09).

Concerning Acceptability rates, the only significant effects were between the moral and non-moral decisions, with higher acceptability for actions that did not involve killing a person (Human driving: V = 0, P < .001, mean ranks = 8.96, 4.41; autonomous driving: V = 0, P < .001, mean ranks = 8.53, 4).

Participants experienced more arousal when they had to deal with moral decisions in human driving mode compared to the autonomous one (V = 4, P = .002, mean ranks = 6.59, 5.27) and were more activated when facing a moral situation compared to non-moral one in both driving modalities (Human driving: V = 276, P < .001, mean ranks = 6.59, 1.82; autonomous driving: V = 399, P < .001, mean ranks = 5.27, 1.73). No differences emerged between non-moral actions in the autonomous and human driving conditions.

Concerning Valence scores, an effect of the moral content was found: participants experienced lower pleasantness when a person was killed in both human (V = 0, P < .001, mean ranks = 2.07, 7.79) and autonomous driving modality (V = 0, P < .001, mean ranks = 2.67, 7.33) compared to non-moral situations. Finally, the driving modality showed an opposite pattern of influence on valence scores: actively killing a person in human driving mode was perceived as more unpleasant than letting a car perform the same action (V = 87, P = .02, mean ranks = 2.07, 2.67); in contrast, when the situation did not involve killing someone, as in the non-moral dilemmas, letting the car decide was rated less pleasant than actually acting in the same way as a driver (V = 2.5, P = .02, mean ranks = 7.33, 7.79).

Regarding the differences between Numerosity and Age dilemmas, only the Friedman tests for Responsibility and Arousal were significant (Responsibility: $\chi^{2(}3)$ = 17.52, p < .001; Arousal: $\chi^2(3)$ = 14.88, p = .001). However, no differences between the two types of moral dilemma reach the significance threshold in the Wilcoxon comparisons analysis.

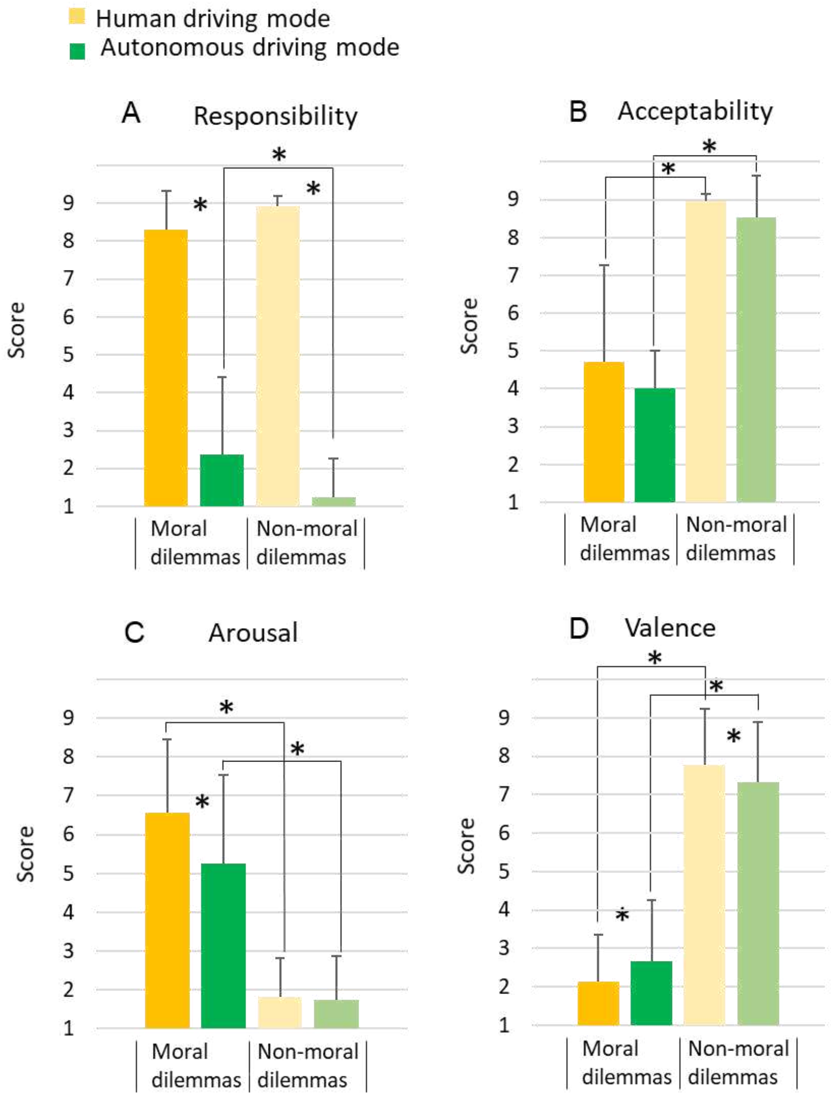


Figure 5.5. Bar graph by dilemma type and driving mode of a) responsibility, b) acceptability, c) arousal, and d) valence scores. No distinction was made between Numerosity and Age dilemmas.

Questionnaires Results

A Wilcoxon signed-rank test was used to compare the real median with the theoretical median of the PQ. Results showed that participants perceived a more than an adequate sense of presence in the virtual environment used in the present study (V = 120, P < .001, Median = 5.23, 4).

The MC-SDS-b was used to verify a possible effect of social desirability on self-reported Responsibility and Acceptability of the action made in the dilemmas. Results were reported in the Social Desirability Control Check section.

The post-experiment control questions revealed that all participants found the instructions clear, felt they had sufficient time to decide and had no trouble seeing the avatars. In some cases, technical problems or errors in pressing the button related to the selected choice were reported: all these trials (6 out of 120) have been identified and excluded from the analysis. All participants reported no cybersickness associated symptoms (nausea and headache).

5.4. Discussion

The present work described a study in the automotive sector, in which the level of interaction of the VR environment was manipulated to give the participants the possibility of experiencing a human driving modality and an autonomous one. Differently from other studies that focused almost exclusively on users' behavioral reactions (Bergmann et al., 2018; Kang et al., 2019; S. Li et al., 2019; Sütfeld et al., 2017), self-reported measures were also used in this work to obtain a more comprehensive frame of the participants' experience. To obtain timely measurements and to minimize the possibility for the participants of experiencing cybersickness and breaking the sense of presence, self-reported levels of arousal, valence, perceived responsibility, and acceptability were collected directly in the VR simulation after every single trial. Indeed, participants reported high levels of presence, measured with the PQ (Witmer et al., 2005; Witmer & Singer, 1998), and none experienced cybersickness symptoms (such as headache and nausea), as reported in the post-experiment control questions, implying that the virtual simulation used in the present study effectively immersed and involved users.

From a theoretical point of view, the present contribution breaks away from the debate on the applicability of dual-process theory (Greene et al., 2001, 2004) to dilemmas in virtual rather than textual form (e.g., Francis et al., 2016; Patil et al., 2014), focusing instead on the emotional processes of the participant in virtual moral dilemmas only. What emerged was that although the "contextual richness" of VR likely played a role in evoking an emotional response by allowing participants to "experience" the situation (instead of only mentally simulating it as in classic textual dilemmas), it is the need to actively react to the moral dilemmas presented which most affected the participant's emotional response. Indeed, the first hypothesis was confirmed: when participants were immersed in the more interactive VR scenario, in which they could decide how to behave in the face of the moral dilemma situation, arousal and negative valence increased significantly.

As expected, in accident situations involving the death of someone, being the actual driver was perceived as an intense, stressful experience, more than being in the same situation but without the possibility of influencing the events as in the autonomous driving modality.

This result could be related to the perceived sense of responsibility: the inability to interfere with the actions of a fully autonomous vehicle may have made the driver feel less responsible for the consequences of that actions, with a consequent decrease in emotional response. According to this interpretation, the results clearly highlighted a significant reduction of the perceived responsibility when in autonomous driving modality trials (confirming the second hypothesis). This result integrates and confirms previous works that found a decreased attribution of blame/responsibility for users of fully-autonomous vehicles compared to those that manually drove the car (J. Li et al., 2016; Wilson & Theodorou, 2019).

Finally, it is important to make a clarification: the participants showed a significant increase in arousal only in moral dilemma situations, while the comparison between human and autonomous non-moral dilemmas, which represented "neutral" and common driving situations, is not significant. This implies that the results are due to the moral content of the act and not to a general activation related to performing any action in a virtual environment.

Interestingly, an increased negative valence was found in these non-moral dilemmas in the less interactive autonomous driving modality compared to the more active human one. This suggests that only in situations with very negative outcomes, such as moral dilemmas, people find it less unpleasant to let the car act in their place, but in simple, neutral decisions (i.e., picking an empty road when driving), they prefer to be the ones to decide. The possibility to behave freely and exert control over the environment is fundamental for the individual (Leotti et al., 2010), and it is possible that the "loss of agency" that characterizes the autonomous driving condition has led to this result.

Overall, the results of research questions 1 and 2 outline an interesting point for AVs design: currently, semi-autonomous driving systems (SAE level 3) are designed to act autonomously under normal driving conditions, but the driver has to regain control in the event of a critical situation, while the participants of the present work seem to prefer an exactly opposite scheme.

As expected, the moral acceptability assigned to the actions was not influenced by the level of interaction with the scenario, and therefore by the driving modality, supporting previous studies (Bonnefon et al., 2016; Kallioinen et al., 2019) and confirming hypothesis 3. Differently from other VR experiments (Kallioinen et al., 2019; Wilson & Theodorou, 2019) in which participants were placed in a passive position even in the human driving condition (e.g., in the role of passengers sitting in a vehicle fully controlled by the experimenter), the driver in the present work was let to actively control the car when in the human driving mode, in order to recreate more realistically the real driver's experience. In the simulation, the participants were never totally passive, not even in autonomous driving mode: even though they could not influence the actions of the car, they could move around in the seat and look around, as every person in a real car does. The degree of acceptability seems to be related more to the content of the action itself and less to the agent who made it, be it the autonomous vehicle, the driver (or an external human entity such as the experimenter, like in Wilson & Theodorou, 2019).

Regarding the behavioral results, when in human driving condition, participants showed longer RT when reacting to moral scenarios compared to control situations (non-moral dilemmas) where they can choose to pick an empty road. This is consistent with the emotional and cognitive conflict elicited by moral dilemmas and with the consequent necessity to analyze the physical features of the avatars in order to reach a decision and act accordingly, increasing resolution times.

Concerning the moral actions, despite it being possible to sacrifice themselves instead of killing others, no one did that, confirming hypothesis 4. The behavioral findings fit into the broader debate on allowing AVs to sacrifice their passengers to save more pedestrians. Previous survey studies found that people are in favor of passenger sacrifice but would not buy an AV designed in this way. The present results showed that totality of the participants behaved in a utilitarian way in the Numerosity dilemma, and most of them sacrificed the older avatar in the Age dilemma consistently with previous studies (Bergmann et al., 2018; Faulhaber et al., 2019; Kallioinen et al., 2019; S. Li et al., 2019; Skulmowski et al., 2014; Wintersberger et al., 2017), but never put their own lives at risk. This suggest that the strong avoidance of "self-sacrifice" actions is due to the design choice to only allow a limited time to act and to give more possibilities for actions to the participant in order to limit the impact of the social desirability bias. It is possible that the participant's tendency to respond in line with social norms instead of their own true beliefs can be decreased by designing more realistic VR scenarios, which elicit a greater sense of presence. An increased tendency to self-sacrifice was, in fact, found in works that used classic dichotomic situations ("kill oneself vs. kill a variable number of potential victims") (Bergmann et al., 2018; Faulhaber et al., 2019; Frison et al., 2016; Wintersberger et al., 2017), or that let unlimited time to act (Frison et al., 2016; Wintersberger et al., 2017). Instead, works like that of Ju et al. (2016), in which participants had to deal with the sudden and unexpected appearance of avatars in the middle of a road and could try different resolutions (e.g., hitting the brake, trying to pass the people, taking the foot off the accelerator, turning right and falling down off a cliff) showed that only one participant sacrificed him/herself to avoid endangering someone else. In support of this hypothesis, both the simulation of Ju et al. (2016) and the one in the present study have been shown to elicit a high level of presence in the participants. Designing highly ecological and engaging VR simulations, in which the user truly feels "present", is a fundamental point to study human behavior in a realistic way, and unfortunately, there are still too few works in the field of moral dilemmas that measure presence. Another way to control the social desirability bias is through its measurement. In the present study, participants with high scores in the MC-SCD-b (Manganelli-Rattazzi et al., 2000) reported a diminished sense of responsibility after moral actions made in human driving mode compared to participants with lower scores. It is possible that people with such a strong tendency to be influenced by social norms may thus prefer not to report feeling responsible for the death of someone in order to avoid being blamed for it.

In conclusion, the main results showed that the principal differences between human and autonomous driving modes, at least in the present study, are at the emotional level and in the perceived sense of responsibility experienced by the driver, while no differences in terms of moral acceptability of the choices emerged. In other words, when people abstractly think of self-driving cars making moral choices instead of humans, a clear aversion toward this possibility emerged (Bigman & Gray, 2018), but when people were immersed in a VR environment and faced moral dilemmas in a self-driving car, as in the present study, they experienced less unpleasantness and arousal than when they were the actual drivers, probably because without any possibility to interfere with the car's actions, they perceived a diminished responsibility for the tragic consequences. On the other hand, when they have to deal with more normal and common driving situations, participants found to be in the human driving modality more pleasant. This pattern showed an opposite direction compared to the current state of semi-autonomous systems design (SAE levels 3), implemented so as to allow the car to drive autonomously in normal conditions and to require the driver to regain control in a critical situation. These results, if confirmed by future studies, could have potential real-world implications and could be used to inform the design of AVs. Particularly, they suggest two future lines of research. The first one concerns the investigation of the factors that contribute to making it more emotionally aversive to face critical situations as a driver rather than as an AV user. In the present work, the personally perceived sense of responsibility has been investigated, but other factors, such as the possible foreshadowing of legal consequences, should be addressed. If I know that I am equally legally responsible even when the car chooses autonomously, does the emotional difference between the two driving modes remain? The second line of research concerns the study of the emotional experience of autonomous and human driving modes in quiet driving conditions. This work investigated very simple situations (i.e., turning into a free road at an intersection), while it would be interesting to test more complex situations and mixed driving modes (e.g., including semi-autonomous conditions). Taken together, the results of ours and future researches on this topic could help to better understand people's reactions to AVs and, possibly, to gain relevant insights on how to increase trust in self-driving vehicles (Yokoi & Nakayachi, 2020).

Some limitations of the present study should be acknowledged. The first and most important is the small sample size that greatly limits the generalizability of the results. Therefore, further studies should extend these findings with larger and more heterogenous samples. Second, emotion was considered only from a conscious and subjective point of view, while the additional adoption of physiological measures could have helped to achieve a more exhaustive understanding of the participants' emotional experience. Third, the Vive controller's touchpad was used for route selection rather than a steering wheel. Although similar systems have been used in previous works (Faulhaber et al., 2019; Sütfeld, Ehinger, et al., 2019) and that this has sped up and simplified the procedure (the same button was used to answer the post-trial questions), it may have diminished the realism of the simulation. Lastly, although the participants reported high scores in the PQ, showing that the scenario elicited a great involvement and sense of presence, and the technical precautions adopted to avoid habituation, the repetition of similar situations through multiple trials may have influenced participants' reactions. Future studies should use more realistic moral dilemmas, recreating a single, sudden, and unexpected accident situation.

6. Study 2 - Who is accountable while driving a semi-autonomous car? A Virtual Reality Study on the Influence of Legal Liability in Moral Dilemmas

6.1. Introduction

Significant progress has been made in the development of autonomous vehicles (AVs) in the last years, with the result that AVs in the next future will become a concrete reality. There are several expected advantages of a widespread adoption of AVs: a reduction in congestion caused by traffic, an increase in mobility for physically impaired people, lower pollution levels, and, above all, an important reduction in traffic accidents (Fleetwood, 2017).

However, before the era of fully autonomous cars becomes a reality, we have to face a delicate transition period where control will be shared between machines and humans (Awad, Levine, et al., 2020). This scenario raises unprecedented questions for the automotive world: we have always been accustomed to the driver being solely responsible for the behavior of the vehicle (*Vienna Convention on Road Traffic*, 1968), but if the control is shared, who has the legal liability in case of an accident?

For automation levels 0-2, as defined by the Society of Automotive Engineers (SAE), it is clear that the driver is always the one in control of the car even if assisted by "support features" (SAE, 2021). Therefore, accountability still lies with the car's driver, as established by the Vienna Convention on Road Traffic (*Vienna Convention on Road Traffic*, 1968). In contrast to this, for automation levels 3-5, it is the car that is driving while the "automated features" are engaged (SAE, 2021), and recent guidelines proposed that the accountability could shift from the driver to the companies that built the vehicle (Luetge, 2017; Lütge et al., 2021). However, for now, the legal framework remains unclear, despite a growing number of studies showing that people are highly concerned about liability issues in crashes involving AVs (Cunningham et al., 2018; Dogan et al., 2021; Kyriakidis et al., 2015; Othman, 2021; Richardson et al., 2017) and that the uncertainty about the extent of corporate accountability for this type of accidents could slow down the adoption of automated vehicles (Geistfeld, 2017). On this basis, it appears fundamental to understand people's reactions to the adoption of different legal frameworks in accidents involving semi-autonomous vehicles. In the present experiment, I investigated whether knowing who is legally responsible in the event of an unavoidable collision (the driver vs. the company that built the car vs. no information about liability) affects people's behavior and emotional experience. To this end, I decided to use modified versions of the "Trolley problem" (Foot, 1967; Thomson, 1985), in which the driver has to choose between two morally salient and mutually exclusive options: let the semi-autonomous car run over one or more potential victims on the main road or save them by assuming control of the vehicle and turning onto a secondary road where another person stands.

More in detail, two types of moral situations were addressed: "unbalanced" dilemmas, in which one of the two options is prominently preferable to the other (i.e., a "three people vs. one" and a "child vs. elder" situation), and "balanced" dilemmas, where there is no strong incentive for one of the options (i.e., a "woman vs. man" and a "man on the crosswalk vs. man not on the crosswalk" situation). The dilemma situations were administered using desktop Virtual Reality (VR), which allowed us to recreate a realistic scenario and implement time pressure into the decision-making process.

VR has already been effectively applied in several studies on moral decision making in the driving context (Bergmann et al., 2018; Frison et al., 2016; Ju et al., 2016; Sütfeld et al., 2017; Wintersberger et al., 2017), and its adoption is highly recommended when the goal is, as in this case, to gain a better understanding of the emotional reactions and behaviors of drivers (Kochupillai et al., 2020).

To the best of my knowledge, only one study assessed the effect of considering legal consequences on moral dilemmas resolution and the related emotional experience (Pletti et al., 2015), but not in a driving context or using realistic VR versions of the dilemmas. For this reason, I advance no specific hypotheses, but I want to address the following three main research questions:

1. Is the choice in an unavoidable collision situation influenced by knowing who will be held legally responsible for the consequences?
2. Does information about legal liability affect the driver's emotional experience?
3. Does the level of balance between the dilemma's options affect the driver's choice and emotional reactions?

In addition, participants' judgments on AVs were collected through a series of questions (after the virtual simulation and with no time pressure) to assess whether the manipulation could also affect "cold" preferences that are more mediated by cognitive processes.

In the following section, the present study was framed within the relevant literature on the topic.

6.2. Related works

While the world is waiting for a top-down deliberation about the delicate theme of legal liability for AVs, some works have begun to investigate how the general public attributes blame to different actors in crashes involving fully and semi-autonomous vehicles (Awad, Levine, et al., 2020; J. Li et al., 2016; McManus & Rutchick, 2019; Pöllänen et al., 2020). Understanding how people judge this type of accidents could provide useful insights for policymakers. Regarding fully autonomous cars, in the survey work of Pöllänen et al. (2020), manufacturers and the government received more blame than vehicle users. Similarly, when a self-driving car was compared to a human driver, more responsibility was allocated to manufacturers and the government in case of accidents (J. Li et al., 2016).

For semi-autonomous cars, the picture is quite different, with a tendency to assign more responsibility to the driver and less to the vehicle (and consequently, to the companies that produced it). Indeed, while McManus & Rutchick (2019) found a decrease in the responsibility assigned to users of fully AVs, even when they pre-programmed the AV to act selfishly in Trolley-like accident situations, Pöllänen et al. (2020) found that drivers/users of semi-autonomous vehicles received the same amount of blame then drivers of a regular car. In addition, the results of Awad, Levine, et al. (2020) showed that in human-machine shared control vehicles, if both the car and the driver make errors, the blame attributed to the machine is reduced. These works highlight how the control on the vehicle seemingly plays a crucial role in responsibility attribution and that even if in semi-autonomous driving this control is reduced (but not absent), this is sufficient to attribute greater responsibility to the driver. It is interesting to note that even in real accidents involving semi-autonomous vehicles, such as the one involving the Tesla Autopilot car in 2016, the responsibility has been allocated to the driver, even if both the driver and the driving system made mistakes (National Highway Traffic Safety Administration, 2017).

All the studies cited above focused on how responsibility was allocated between different road transport system actors, while in the present study I took a different perspective: the participant was not an external judge who decided which parties to blame, but a driver directly involved in the extreme accident who already knew who would be legally responsible for the different possibilities of action. The goal is to understand if and how different legal responsibility frameworks influence participants' behavior and emotional experience. The only study that assessed the effect of legal liability on moral decision-making was the work of Pletti et al. (2015), with classic Trolley and Footbridge-type dilemmas that were not specific to the driving context. Interestingly, it highlighted how participants who considered the legal consequences of their actions showed a dampened emotional experience at both the subjective and neural levels, exhibiting a decrease in emotional valence and arousal ratings and an increased *bereitschaftspotential* (suggesting a lower moral conflict at the level of action preparation). However, they still made the same behavioral choices of participants who were previously informed that none of the options presented was legally prosecutable (Pletti et al., 2015). It is possible that, at least with classic moral dilemmas in textual format, knowing the legal consequences of actions provide a sort of reference point on which people can rely to support their decision. At the same time, decision-making might become more complex without this information, thus increasing the emotional burden of choice to make (Pletti et al., 2015).

6.3. Method

6.3.1. Experimental Design

A 3 (*Legal Liability*) × 2 (*Dilemma Type*) mixed-participants design was employed. The levels of the first variable were the three conditions: "Driver Responsibility" (DR), "Company Responsibility" (CR), and "No information" (NI). The within *Dilemma Type* variable consisted of 2 levels: "unbalanced" situations, namely moral dilemmas in which we know from the literature that one of the two options is clearly preferred over the other (i.e., "three people vs. one" and "child vs. elder" dilemmas); and "balanced" situations, where there is not a strong preference for any of the options (i.e., "woman vs. man" and "man on the crosswalk vs. man not on the crosswalk" dilemmas).

6.3.2. Participants

A total of 66 participants (39 males; age M = 26.85, SD = 5.9) were included in the present study. The sample size of 66 people was calculated through an a priori power analysis conducted with Gpower (Erdfelder et al., 1996), using medium effect size (f =0.25) and a 95% power as input parameters. Recruitment continued until the required sample was reached, removing from the count the participants who: 1) chose to kill a person in filler trials (situations in which, in addition to a road with people, there was an empty road), and 2) did not answer the post-experiment control questions correctly (i.e., one question to check if they had paid attention to the initial message on legal responsibility, and the other to verify the correct recognition of potential victims encountered in the trials). Of the 92 participants who took part in the experiment, 12 were excluded for the first reason and 14 for the second one.

Participants voluntarily chose to participate in the online experiment (no monetary incentive was provided), and they were required to have a native level understanding of Italian and be at least 18 years old.

Exclusion criteria included having experienced a severe car accident or other car-related trauma (e.g., losing a family member or a friend in an accident). The protocol of the present experiment was approved by the local ethics committee, and the study was conducted following the principles of the Declaration of Helsinki.

6.3.3. Material

The VR simulation was implemented using the development platform Unity (version 2019.4.28f1) and Blender (version 2.93). The software artifacts were built using WebGL (a JavaScript API for rendering interactive 2D and 3D graphics within any compatible web browser) and hosted on a Node.js server (version 14.17 LTS) deployed on the cloud application platform Heroku. The experiment was administered directly on the participant's PC screen through an HTTPS connection, ensuring confidentiality. Participants could visually explore at 360° the scenario by moving the mouse and decide whether to turn the car into the secondary road by pressing the left or right button (depending on the side where the secondary road was in the specific trial) (Figure 6.1.).

The mouse was also used to select and confirm the answer to the post-trial questions and the post-experiment questionnaire.

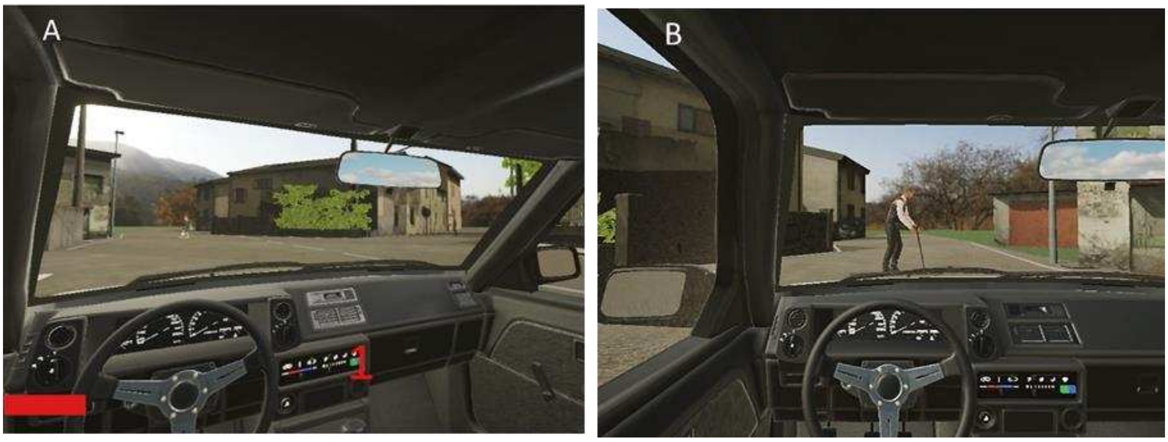


Figure 6.1. a) One trial of the simulation (the Age dilemma) during the choice window and b) after the choice, shortly before the impact.

6.3.4. Measures

Behavioral Measures

For each participant and in each trial, the choice made (i.e., to turn or let the car go straight), the response times (RTs), the frequency of button pressed, and the percentage of time spent with the mouse cursor in the left, central and right portions of the screen during the choice window were collected.

Subjective Measures

After each trial, the perceived responsibility, acceptability, arousal, and affective valence related to the decision were assessed. A 9-point Likert scale was employed for the first two variables, while the self-assessment manikin (SAM) (Lang et al., 2008) was adopted for measuring arousal and valence.

A short demographic questionnaire and an ad-hoc post-experiment questionnaire were administered, the first to gather information about the sample (age, gender, prior knowledge about AVs, and videogame experience) and the second to collect the opinions of the participants on autonomous and semi-autonomous vehicles and on the virtual simulation just experienced (8 items with 9-point Likert scale). Finally, a question about their interest in purchasing an AV, always using a 9-point Likert scale, was presented two times (before the experiment -in the demographic questionnaire-, and after the experiment

-in the ad hoc questionnaire-) to assess possible changes in this variable due to the content of the desktop VR simulation.

6.3.5. Procedure

Participants took part in the experiment via a link. After reading the informed consent and the inclusion and exclusion criteria, they filled the demographic questionnaire. Then, general instructions of the virtual task were provided: they are called to drive a semi-autonomous vehicle that has a breakdown in the braking system, so it can change direction but not stop. The car moves autonomously, following the main road, but they will have the possibility of taking control of the vehicle and changing the default direction in case of crossroads. After every single crossroad, they will have to answer a series of questions about their decision. Participants were further informed that they could stop the experiment without reason at any moment and that closing the browser window before the end of the experiment automatically meant that no data was sent. An automatic assignation to one of three possible conditions followed the instructions. Participants in the DR condition read the following additional instruction before starting the task: "According to Vienna Convention on Road Traffic (1968), the driver is always fully responsible for the behavior of the vehicle". In the CR condition, the additional instruction was: "According to the Ethics Commission for Automated and Connected Driving (2017), if the driver cannot control the car fully in every single situation and is not required to do so, he or she cannot be accountable anymore for the car's behavior, but only the companies who built it" (Luetge, 2017). In the NI condition, no additional instruction about legal liability was showed.

The virtual session started with two training trials to familiarize with the "desktop VR" simulation and the task. No human avatar of potential victims was showed during these trials. Participants were then exposed to the four experimental trials and four "filler trials" (one human target vs. an empty road), all presented in random order. Similar to Bergmann et al. (2018), Faulhaber et al. (2019) and Sütfeld et al. (2017), every trial started with the participant sitting in the driver seat of a car traveling at a constant speed in a city (Figure 6.2.). After 10 s, a Y junction appeared, with the main road on the center and the secondary one randomly on the right or the left (to avoid habituation). From the appearance of the crossroads, the participants had 4 s to decide whether to intervene and turn into the secondary road or let the car continue in autonomous mode for the main road. The 4-second countdown was shown in red characters at the bottom of the screen to make it clear to the participants when the choice window started and ended. Shortly before the impact (2 s) with a human avatar, the screen faded black for 3 s. The collision was signaled by the sound of an object hitting the hood of the car. When the black screen disappeared, participants' arousal, affective valence, perceived responsibility, and acceptability of the decision were assessed. When the last question was completed, a new trial began after a brief delay.

At the end of the last post-trial evaluation, participants answered two control questions to assess their attention during task and instruction and then compiled a brief questionnaire. The whole session lasted about 15 minutes.

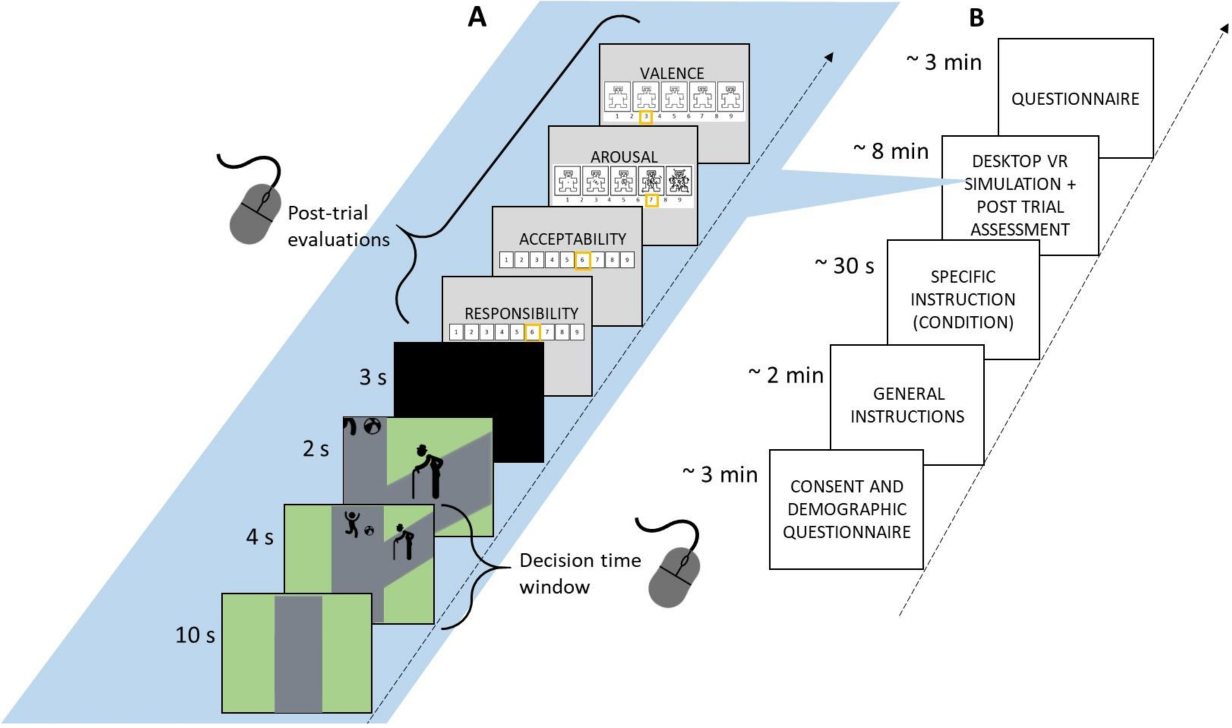


Figure 6.2. Graphical representation of: a) the events of a trial and b) the entire procedure.

6.4. Analysis

Behavioral and subjective data were analyzed with Generalized Mixed Effect Models (GLMMs), including all trials for each participant except the filler and the training trials. GLMMs were chosen because they estimate both fixed and random effects and are especially useful when the dependent variable is binary, or quantitative but not normally distributed (Bono et al., 2021), as in the present case. In case of significance, post hoc contrasts were performed using the Tukey method to adjust p -values for multiple comparisons.

First of all, I investigated whether the condition (*Legal Liability,* three levels: DR, CR, NI) and the type of situation represented (*Dilemma Type,* two levels: unbalanced, balanced) influenced the probability of choosing to turn the car. For this purpose, a mixed effect logistic regression model was built, with Choice (0 = letting the car continue on the main road, 1 = take control and turning) as dependent variable, *Legal Liability*, *Dilemma Type,* and *Legal Liability* × *Dilemma Type* interaction as fixed effects and Participant as a random effect. Similarly, two models were built using RTs and Frequency of Buttons Pressed as dependent variables and the same fixed and random effects of the previous one. For assessing effects on the mouse position, a model was built with the Percentage of Time spent with the cursor in each screen location as dependent variable, while *Legal Liability, Position* (Centre, left, or lateral), *Choice, Dilemma Type,* and their four-way interaction were set as fixed effects and participant as a random effect.

Regarding the subjective data, I first tested whether the condition, the choice, and the type of situation represented affected the perceived sense of responsibility, acceptability, arousal, and emotional valence reported after each trial. To this aim, independent GLMMs were built for each of these four variables, using their score as dependent variable, *Legal Liability*, *Choice, Dilemma Type,* and *Legal Liability* × *Choice* × *Dilemma Type* interaction as fixed effects, and Participant as a random effect.

I then verified whether the condition and background characteristics influenced participants' interest in purchasing an AV, their opinions about autonomous and semi-autonomous cars, and their experience with the virtual simulation. To this end, independent models were built for each question of the post-experiment questionnaire, using their score as the dependent variable. For questions related to AVs (7 out of 8), I used *Gender* (male, female)*, Legal Liability*, *Prior Knowledge of AVs* (yes, no), and their interactions as fixed effects, and participant as a random effect. For the question about the sense of presence experienced during the simulation, the same model was built, but with the factor *Video Gaming Frequency* (Never, A few times a year, A few times a month, Weekly, Daily) instead of *Prior Knowledge of AVs*. For the model regarding the

"Interest in Buying" an AV, in addition to *Gender*, *Legal Liability,* and *Prior Knowledge of AVs,* I also set *Time* (pre-experiment, post-experiment) as a fixed effect.

To select the best model for each analysis, I started from the model including only the random effect, and then introduced the fixed effects one by one, in the order described above. To compare models, the F-test was used.

Finally, additional exploratory analyses were conducted on behavioral data and post-trial evaluations to assess possible gender differences or effects due to prior knowledge about AVs. To this aim, I started from the previously identified best models, and I used them to compute two new GLMMs for each dependent variable: in the first *Gender* was added as fixed effect, and in the other *Prior Knowledge of AVs*.

All statistical analyses were performed using RStudio (R CoreTeam, 2015), with the libraries stats (R Core Team, 2015), lme4 (Bates et al., 2014), glmmTMB (Brooks et al., 2017), emmeans (Length, 2020), and effects (Fox, 1987).

6.4.1. Behavioral Results

Regarding the choice, the best model included the factors *Legal Liability* and *Dilemma Type* but not their interaction ($\chi2(1) = 45.59$, $p < .001$). Only the effect of *Dilemma Type* on the choice was significant ($\chi2(1) = 36.19$, $p < .001$): the probability of deciding to take control of the car and turn was higher in the unbalanced situations compared to the balanced ones (Figure 6.3.A).

Similarly, the best model for RT was the one with only the main effects (*Legal Liability* and *Dilemma Type*; $\chi2(1) = 4.12$, $p = .042$). Also in this case, *Dilemma Type* was significant ($\chi2(1) = 4.16$, $p =.041$), showing that longer RTs were required to take a decision in the balanced situations compared to the unbalanced ones (Figure 6.3.B). General descriptive statistics for choice and RT are reported in Table 6.1. Regarding the frequency of buttons pressed, none of the tested models showed a significant increase in the goodness of fit during the model selection procedure (all $p > .05$). Finally, for the percentage of time spent in the different screen portions, the best model included only two main effects, *Legal Liability* and *Position* ($\chi2(1) = 4.85$, $p = .027$). The *Position* effect was significant ($\chi2(1) = 4.86$, $p = .027$): participants always preferred to keep the mouse cursor in the central part of the screen rather than in the left or right lateral portions.

Finally, about the gender and prior knowledge about AVs, the exploratory analysis still showed significance for the previously described factors. Still, neither gender nor prior knowledge about AVs resulted in being significant.

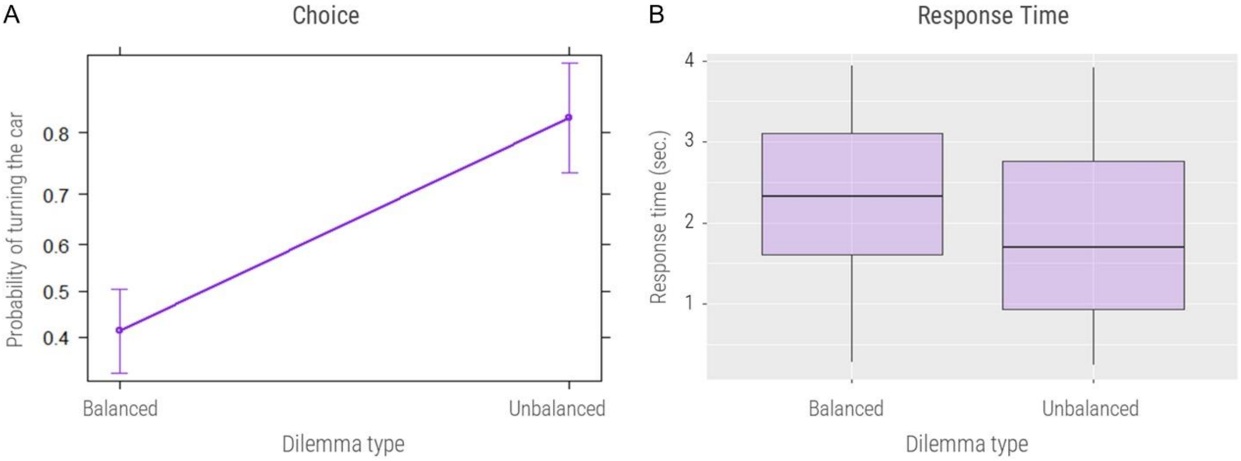


Figure 6.3. a) Effect of the Dilemma Type on the Choice of taking control and turning the car. b) Response time in balanced and unbalanced situations.

Table 6.1. Percentage of the choice of turning and means and standard deviations of RTs (in seconds) as a function of Gender and Dilemma Type. In case of significant main effect, the values are indicated in bold.

	Legal Liability condition			Dilemma Type	
DV	DR	CR	NI	Balanced	Unbalanced
Choice (% to regain control and turn)	68.18	55.68	60.22	**41.67**	**81.06**
RT (M± SD)	1.963 ± 2.287	2.072 ± 1.791	2.021 ± 2.352	**2.268 ± 2.522**	**1.885 ± 1.948**

6.4.2. Self-Report Results

Post-Trial Evaluations Results

Responsibility. For responsibility, the best model included the interaction between *Legal Liability* and *Choice* (χ2(2) = 6.47, p = .039). Both the main effect of *Choice* (χ2(1) = 22.48, p < .001) and the *Legal Liability* × *Choice* interaction (χ2(2) = 6.49, p = .038), were significant (Figure 6.4.A). From post-hoc comparisons emerged that in the CR condition participants reported a lower perceived responsibility for the choice to have the car continue in autonomous mode than that of taking control of the car and turning (p <

.001). This comparison was not significant in the other two conditions: participants felt highly responsible for both choices.

Acceptability. Regarding acceptability and arousal, the best model for both variables was the one with the interactions *Legal Liability* × *Choice* and *Dilemma Type* × *Choice* (acceptability: χ2(1) = 5, p = .025; arousal: χ2(1) = 8.43, p = .003). For acceptability, the main effect of *Choice* (χ2(1) = 10.79, p = .001) and the interaction between *Dilemma Type* and *Choice* (χ2(1) = 4.95, p = .026) were significant (Figure 6.4.C). Post-hoc contrasts showed that choosing to turn in the unbalanced situations was judged more acceptable than letting the car drive straight ahead in autonomous mode (p < .001), while in the balanced situations there were no differences between the two choices.

Arousal. For arousal, the main effect of *Choice* (χ2(1) = 4.51, p = .033) and both the interaction effects *Legal Liability* × *Choice* (χ2(2) = 9.46, p = .008) and *Dilemma Type* × *Choice*: (χ2(1) = 8.34, p = .003) were significant (Figure 6.4.B and 6.4.D). As demonstrated by the post-hoc test, only participants in the NI condition showed increased arousal for the choice of letting the car continue in autonomous mode (p = .002), while in the other two conditions a similar level of activation was reported for both choices. The decision to let the car go straight in the unbalanced situations was judged more activating than making the same choice in the balanced ones (p = .036). Moreover, only in the unbalanced situations, it elicited more arousal than turning (p = .002).

Valence. Finally, the best model for valence included only the main effects (χ2(1) = 3.96, p = .046). The factor *Choice* (χ2(1) = 4.74, p = .029) significantly influenced the self-reported valence: letting the vehicle continue on the main road was rated as more unpleasant than choosing to intervene and turn the car. *Dilemma Type* (χ2(1) = 3.99, p = .045) was also significant, showing that the unbalanced situations elicited a greater unpleasantness compared to the balanced ones.

General descriptive statistics for responsibility, acceptability, arousal and valence are reported in Table 6.2.

Exploratory analysis. Regarding the exploratory analysis, the gender and the presence of prior knowledge of AVs did not affect results on responsibility and acceptability: all the prior significant main effects and interactions remained significant, and the only new one (acceptability: *Legal Liability* × *Prior Knowledge of AVs*, χ2(2) = 6.18, p = .045) did not reach the significance threshold in any of the post-hoc contrasts (all p > .05). Instead, there is an influence of the *Gender* factor on the perceived arousal and valence: in addition to all the previous results that have been confirmed, a new main effect of *Gender* (arousal: χ2(1) = 7.08, p = .007; valence: χ2(1) = 10.28, p = .001) and a *Gender* × *Choice* interaction (arousal: χ2(1) = 5.18, p = .022) were found. Females experienced more arousal and unpleasantness than males when faced with dilemma situations. Post-hoc contrasts showed that female participants were more activated than males when they chose to take control of the car and turn (p =.006). In addition, males reported increased arousal for the choice of letting the car continue autonomously on the main road compared to the choice of turning (p =.002), while females showed high arousal scores for both the choices.

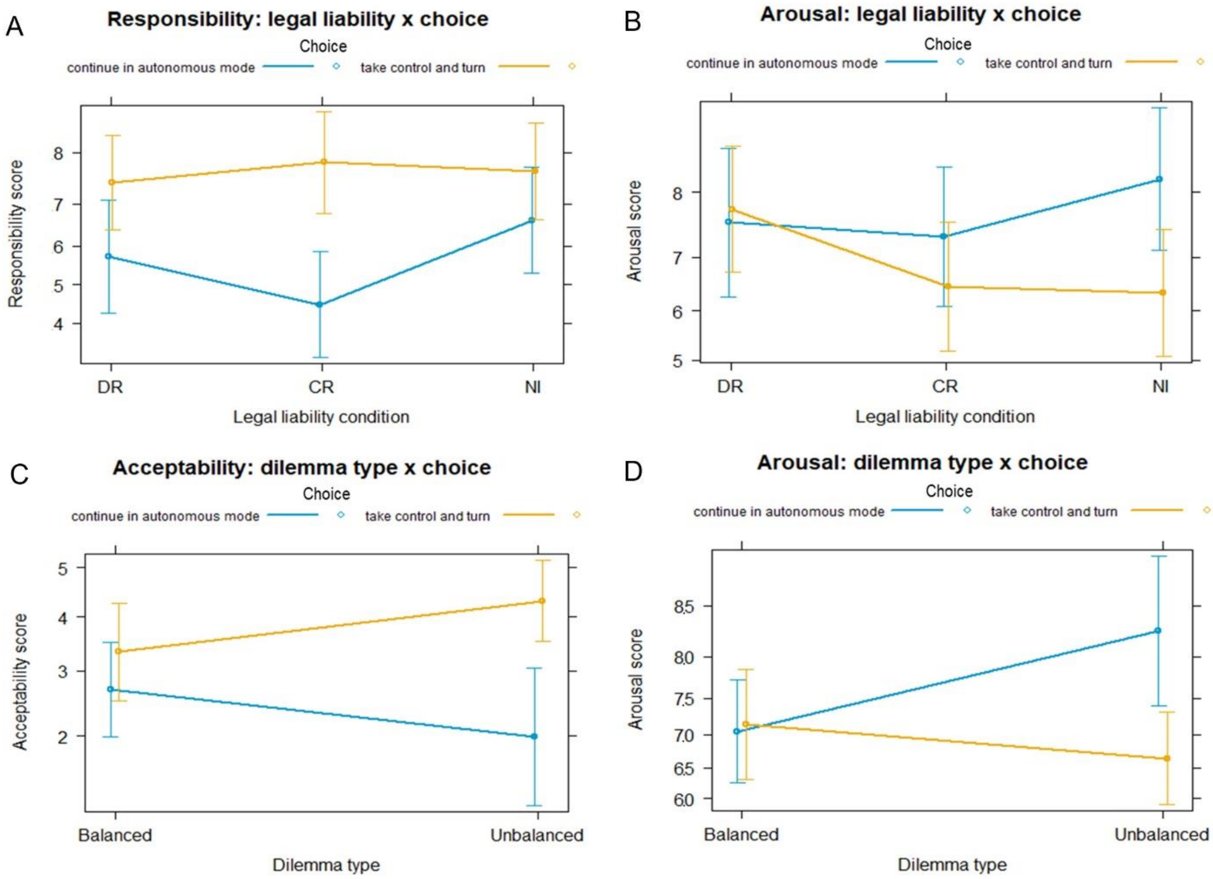


Figure 6.4. Effects of type of Choice on post-trial evaluation scores as a function of Legal Liability condition (a and b) and Dilemma Type (c and d). Error bars indicate 95% confidence intervals.

Table 6.2. Means and standard deviations for all the post-trial variables as a function of Legal Liability, Choice and Dilemma Type. In case of significant main effect, the values are indicated in bold.

	Legal Liability condition			Choice		Dilemma Type	
DV	DR (M ± SD)	CR (M ± SD)	NI (M ± SD)	Continue autonomously (M ± SD)	Turn (M ± SD)	Balanced (M ± SD)	Unbalanced (M ± SD)
Responsibility	6.77±2.73	6.27±2.68	7.04±2.37	**5.55±2.86**	**7.41±2.16**	6.17±2.82	7.21±2.27
Acceptability	4.40±2.63	3.55±2.31	3.98±2.56	**3.37±2.2**	**4.37±2.64**	3.63±2.39	4.33±2.6
Arousal	6.77±1.92	6.25±1.79	6.51±2.07	**6.68±1.65**	**6.40±2.09**	6.59±1.83	6.42±2.05
Valence	2.92±1.81	2.4±1.43	2.22±1.32	**2.32±1.38**	**2.64±1.65**	**2.56±1.62**	**2.46±1.5**

Questionnaire Results

Regarding the interest in purchasing an AV, the best model was the one with all the main effects (*Gender*, *Legal Liability, Prior Knowledge of AVs* and *Time*) and the interaction *Gender* × *Legal Liability* (χ2(2) = 8.9, p = .011). The main effect of *Time* was significant: Compared to before the virtual simulation, all participants showed less interest in purchasing an AV after the experiment (χ2(1) = 6.93, p = .008). Also the interaction between *Gender* and *Legal Liability* was significant, but none of the post-hoc contrasts with Tukey correction reached the significance threshold (all p > .05).

The items and the general descriptive statistics of the post-experiment questionnaire are reported in Table 6.3. For questions Q1, Q5, Q6, and Q8, no model among those tested showed a significant increase of the goodness of fit during the model selection procedure (all p > .05). For Q2 and Q4, the best model included only the main effect of *Gender* (Q2: χ2(1) = 3.88, p = .048; Q4: χ2(1) = 4.81, p = .028). *Gender* was significant in both cases, showing that females found less morally acceptable than males the idea that a machine can autonomously decide who to save and kill (χ2(1) = 3.91, p = .047). At the same time, women are less likely than men to use a fully autonomous car that replaces the driver in any situation (χ2(1) = 4.84, p = .027). For Q3, the full model was the best one (χ2(2) = 9.47, p = .008). The interactions *Gender* × *Prior Knowledge of AVs* (χ2(2) = 8, p = .018) and *Gender* × *Legal Liability* × *Prior Knowledge of AVs* (χ2(2) = 9.07, p = .01) both reached the significance threshold, but the post-hoc comparisons did not lead to significant results. Finally, for Q7 the best model was the one with the main effects of *Gender* and *Legal Liability* (χ2(2) = 6.02, p = .049). The factor *Legal Liability* was significant (χ2(2) = 6.13, p = .046). From the post-hoc test emerged that participants in the CR condition were less in agreement than those in the DR condition that it is right to legally prosecute a driver who lets the car decide autonomously in extreme accident situations (p = .043).

Table 6.3. Means and standard deviations for the items of the post-experiment questionnaire as a function of Gender, Legal Liability and Prior Knowledge of AVs. (The response scale ranged from 1 -totally disagree- to 9 -totally agree-). In case of significant main effect, the values are indicated in bold.

	Gender		Legal Liability condition			Prior Knowledge of AVs	
Question	M (M ± SD)	F (M ± SD)	DR (M ± SD)	CR (M ± SD)	NI (M ± SD)	Yes (M ± SD)	No (M ± SD)
Q1. During the experiment, I felt present in the virtual environment as if I were really there.	5.48±2.06	5.4±2.37	5.4±2.44	5.63±2.03	5.31±2.12	5.53±2.18	5±2.21
Q2. I would use a fully self-driving car capable of replacing the driver in all driving tasks and in any situation.	**4.53±2.56**	**3.11±2.87**	4.4±2.92	3.22±2.56	4.22±2.77	4.25±2.83	2.3±1.63
Q3. I would use a self-driving car only if it were one that gives the driver the opportunity to take back control during critical situations.	7.17±2.17	6.7±2.25	7.4±1.94	7±2.02	6.54±2.59	7.16±1.97	6±3.16
Q4. I believe it is morally acceptable that a car can autonomously choose who to save and who to kill in an unavoidable accident situation.	**3.94±2.64**	**2.22±1.45**	3.22±2.18	2.95±2.12	3.54±2.84	3.48±2.44	1.9±1.44
Q5. I find it right that a self-driving car gives the driver the ability to regain control during critical situations.	8.12±1.15	8.11±1.42	8.36±0.95	8.13±1.42	7.86±1.35	8.12±1.23	8.1±1.44
Q6. If during a critical situation the driver chooses to take back control of the self-driving car and hits someone, it is right that the legal responsibility will rest entirely with the driver.	7.17±1.95	7.29±2.14	7.36±2.46	6.27±2.76	8.04±1.09	7.17±2.07	7.5±1.77
Q7. If during a critical situation the driver chooses NOT to take back control of the self-driving car and the car hits someone, it is right that the legal responsibility will rest entirely with the driver.	5.48±2.81	6.59±2.63	**6.45±2.78**	**5.09±2.68**	**6.27±2.78**	5.71±2.79	7.2±2.44
Q8. If during a critical situation the driver is not given the opportunity to regain control and the self-driving car hits someone, it is right that the legal responsibility will rest entirely with the driver.	1.89±1.31	2.92±2.4	2.22±2.13	1.81±1.62	2.9±1.79	2.25±1.85	2.7±2.16

6.5. Discussion

The present study explored how people react to moral dilemmas in a semi-autonomous driving context under different legal liability frameworks. Participants were randomly assigned to one of three experimental conditions: in one (DR), they were informed that the driver is always legally responsible for the behavior of the car, in another (CR) that the legal liability lies with the manufacturing company when the driver has not the control of the vehicle, and in the last (NI) they did not receive any information. After each dilemma, the participants' experience was assessed, measuring the perceived moral responsibility, acceptability, arousal, and emotional valence related to the choice. The experiment was conducted using desktop VR to increase ecological validity and obtain more realistic measures of the emotional experience.

Regarding the first research question, the legal liability manipulation did not affect the participants' behavior, similar to the finding of Pletti et al. (2015) with classic moral dilemmas. It is not so surprising that participants in the DR and NI conditions intervened when they disagreed with the default choice of the driving system, being always legally responsible for the behavior of the car (DR) or lacking any information on the legal aspects (NI). On the other hand, it is interesting to notice that the participants in the CR condition acted similarly since leaving the car in autonomous mode would have avoided any legal issues for them. This result suggests two possible conclusions: either people do not use information regarding the legal consequences to decide in extreme accident situations, or acting according to what is considered more morally acceptable is more important than avoiding legal repercussions. The latter point implies that it is critical for a driver to be able to interfere in case of disagreement with the semi-autonomous driving system, even at the cost of legal consequences. The analyses of the post-experiment judgments of participants are in favor of this interpretation (items Q2, Q3, Q5), showing how participants find it right to have the possibility of regaining control during critical situations and how it would be rather unwilling to use an AV where this is not possible. It is interesting to note that recent ethical guidelines in the automotive sector also highlighted the importance of human agency and recommended the insertion of an "override function," allowing the driver of maintaining personal autonomy in AVs at least up to level 4 (Luetge, 2017; Lütge et al., 2021).

The second research question concerns the role of legal responsibility on emotional experience. Even if considering the legal consequences did not change the pattern of choices, this does not mean that it did not affect the participant's experience. Indeed a significant *Legal Liability* × *Choice* interaction effect was found for both the perceived moral responsibility and arousal. Participants in the CR condition felt less morally responsible for the choice to leave the car in autonomous mode than the decision to regain control and turn. In contrast, in the other two conditions, the participants reported similar levels of moral responsibility for both options. Therefore, knowing that the legal liability lies with the manufacturing company when the driver has no control of the vehicle, even if it did not change the participants' behavior, led to a moral deresponsabilization for accidents that occurred in autonomous driving mode. This result is partially in contrast with a previous focus group study on ethical issues in automated mobility, in which participants stated in consensus that they would morally feel responsible for killing someone in an accident, even if they would not be held accountable legally (Dogan et al., 2021). However, thinking abstractly about how responsible one can feel in an unavoidable accident versus experiencing it firsthand in a realistic virtual scenario can lead to different results. Previous works have illustrated that mental processes are dependent on the extent to which participants are immersed in a situation similar to driving (Kochupillai et al., 2020; Madary & Metzinger, 2016). In addition, differences have already been found between more and less realistic versions of moral dilemmas (e.g., textual vs. VR modality: Francis et al., 2016; Patil et al., 2014; with vs. without time constraints: Grasso et al., 2020; Lucifora & Grasso, 2020), highlighting the importance to study moral judgment and the related emotional experience not only with classic paradigm but also in a more ecological way.

Regarding the arousal result, only participants in the NI condition showed a significant difference between the two possible options, reporting a greater emotional activation for the choice to let the car continue autonomously compared to the decision of regaining control and turning. If an unavoidable collision in itself is a stressful event for a driver, allowing a vehicle to manage it autonomously represents an at least unprecedented and ambiguous situation. However, it is possible that knowing who is the legally responsible part in this event (be it the driver himself or the vehicle manufacturers) can mitigate the uncertainty of this situation, thus reducing the arousal associated with it. Similarly, the work of Pletti et al. (2015) showed that the consideration of the legal consequences in a classic decision-making paradigm was associated with a dampened emotional experience. In this light, it is understandable that the NI condition, the only one that has no information about legal liability and so without a "reference point," showed increased emotional activation.

Finally, regarding the third research question, a significant effect of the type of dilemma on both drivers' behavior and emotional experience was found. In the "unbalanced" situations, participants exhibited an increased tendency to intervene and regain manual control of the vehicle compared to the "balanced" ones, showing that the number and the age of the potential victims played a more crucial role than their gender or observance of road rules. These findings integrate the ones of previous works that highlighted a strong preference for maximizing the number of saved lives and keeping alive the younger potential target (Awad et al., 2018; Benvegnu et al., 2021; Bergmann et al., 2018; Faulhaber et al., 2019; Kallioinen et al., 2019; Skulmowski et al., 2014; Wintersberger et al., 2017), and a weaker (but still present) tendency to save a woman instead of man (Awad et al., 2018; Skulmowski et al., 2014) or a person compliant with road rules when compared to a non-compliant one (Awad et al., 2018; S. Li et al., 2019). In addition, longer RTs were found in balanced dilemmas than in unbalanced ones, suggesting that those choices were more difficult to make and required additional effort. The behavioral results are consistent with the subjective experience of the participants: the significant *Dilemma Type* × *Choice* interaction showed that in unbalanced dilemmas only, the two options exhibited a clear distinction in their level of moral acceptability and arousal. In particular, only in those dilemmas participants found less acceptable and more activating to let the car continue in autonomous mode on the main road, and it was precisely in unbalanced dilemmas that participants were more likely (and faster) to retake the control and turning. In other words, I suggest that participants acted in a way that minimizes the emotional activation and maximizes the moral acceptability associated with the output. Similarly, the work of (Pletti et al., 2016) highlighted how in moral dilemmas resolution, participants tended to choose the option that minimized the intensity of negative emotions.

From the exploratory analysis conducted to assess possible differences related to gender or prior knowledge about AVs, females reported an overall increased negative valence and arousal compared to males. This result is in accordance with previous work on moral dilemmas (Lotto et al., 2014). The significant *Gender* × *Choice* interaction showed that female participants were more emotionally activated than male ones for the choice of retaking the control of the vehicle. In addition, while males decreased the perceived arousal when choosing to reassume the control instead of letting the car continue on autonomous mode, females showed high activation levels for both options. These results suggest that, overall, the experience of driving a semi-autonomous vehicle during extreme accidents was more unpleasant and arousing for women, in line with their hostility toward AVs that emerged from the post-experiment questionnaire. Regarding this last point, female participants showed to be more opposed than men to the idea of a machine making life and death decisions and, consistently, they reported to be less likely to use a fully autonomous vehicle, a trend also reported in previous studies (Charness et al., 2018; Hulse et al., 2018; Othman, 2021). Beyond gender, the legal liability condition also showed an effect on the answers to the post-experimental questionnaire. Particularly, participants in the CR condition resulted to be less in agreement than those in the DR condition on the right to legally prosecute a driver in case of lack of human intervention. Interestingly, after only ten minutes of virtual simulation under a specific legal frame, participants showed an opinion in agreement with their condition. Finally, after the experiment, participants showed a significant reduction in the intention to buy an AV (compared to the measurement made before starting the virtual simulation), probably due to being continuously exposed to unpleasant accident situations involving semi-autonomous vehicles.

Some limitations of the present study should be acknowledged. The first and most important was the choice of adopting a desktop VR instead of a more immersive system (e.g., head-mounted display or CAVE) which could have allowed us to obtain even more ecological measures of the driver's emotional experience. Indeed, although not many studies have directly compared immersive and non-immersive VR in the moral dilemmas field (Pan & Slater, 2011; Sütfeld, Ehinger, et al., 2019), one of these highlighted a more intense experience of panic in immersive VR simulation (Pan & Slater, 2011). Second, emotion was measured only from a subjective and conscious perspective, while the inclusion of physiological measures could have helped to obtain a more complete understanding of the emotional experience of participants. Finally, in the present work, the focus was on situations with two possible options and a clear output (the death of one of the two possible targets); however, real-world accidents are often complex, with more possible resolutions and not always lead to the death of the people involved. Future studies should assess the role of different legal liability frameworks in more ecological driving situations.

To summarize, to my knowledge, the present study was the first one to assess the effect of different legal liability frameworks on moral decision-making during semi-autonomous crashes, assessing the drivers' behavior and emotional experience in a VR simulation. The main results showed that the adoption of a legal framework that attributes accountability to manufacturing companies did not make the participants more prone to avoid retaking control of the vehicle, but it diminished the perceived moral responsibility associate with this choice. In addition, participants who received no information on legal consequences showed increased activation for the choice of letting the car in autonomous mode. Finally, female participants experienced more unpleasantness and arousal than male ones, and they were also more opposed to the idea of a machine making moral decisions and to the adoption of AVs. Taken together, these findings suggest that legal liability could be not an influential factor for deciding in those extreme situations, but it affects drivers' emotional reactions, which is a relevant aspect that should be considered by policymakers and other parties involved in the delicate world of AVs' legislation.

7. Study 3 – I Think One Way but Act Another: Differences between Textual and VR Moral Dilemmas in the Driving Context

7.1. Introduction

Moral dilemmas such as the Trolley and the Footbridge dilemmas (Foot, 1967; Thomson, 1985) have been widely employed as a paradigm to study moral decision-making. As described in chapter 1, in Trolley-type dilemmas killing one individual is a foreseen but unintended consequence of saving others and most people choose to kill in order to maximize the number of saved lives (utilitarian option). Instead, in Footbridge-type dilemmas, where the sacrifice of one person is literally the means to save more lives, people generally decide not to kill (deontological option) (Greene et al., 2001, 2004; Petrinovich et al., 1993; Petrinovich & O'Neill, 1996).

According to Greene et al.'s dual-process theory (Greene et al., 2001, 2004), this pattern of findings is due to the competition between cognitive and emotional processes: in Footbridge-type dilemmas, a strong negative emotional response elicited by the idea to kill a man leads to the decision of not killing, instead, in Trolley-type dilemmas, cognitive processes prevail and lead to the individual endorsing the utilitarian option.

However, these results and a great part of the literature on the topic are based on moral dilemmas presented in text format, which ensures strong experimental control but diminishes realism and limits the subject's emotional involvement (Patil et al., 2014). Furthermore, the adoption of self-reported surveys to investigate moral decision-making increase the risk of getting answers tainted by social desirability bias (Tan et al., 2021).

In recent years, the introduction of VR technology has finally made it possible to fill this gap and study moral dilemmas resolution in more ecological and engaging situations, while maintaining control over variables and parameters. In addition, the use of VR simulations seems to mitigate the effect of the social desirability bias; in fact, it has been shown that people tend to respond to situations presented in VR as if they were real (Rovira, 2009; Slater et al., 2006). Until now, almost all studies that compared text and VR versions of Trolley and Footbridge dilemmas have failed to confirm Greene's dual-process theory (Francis et al., 2016, 2017, 2018; Patil et al., 2014), finding instead an increase in utilitarian responses and physiological arousal in VR dilemmas, whereas it was expected that the augmented emotional involvement of VR simulations should have led to an increment in deontological choices.

Based on these findings, it was hypothesized that moral judgment and moral action might be distinct constructs (Francis et al., 2016, 2017; Patil & Silani, 2014). Greene’s dual-process theory (Greene et al., 2001, 2004) seems to remain valid in the "arid" textual versions of the dilemmas but not in the virtual ones. It is possible that the greater “contextual richness” offered by VR makes the outcome of an imminent decision more vivid and salient. In this light, letting people die (and likely witness their deaths) could be more emotionally aversive than sacrificing just one to save them, leading to utilitarian resolution as a way to minimize this distress (Francis et al., 2016; Patil et al., 2014). This interpretation fits nicely with Cushman's version of the dual-process theory (Cushman, 2013; see chapter 1), which focuses on the opposition not between emotion and cognition but between processes that assign value to the action (e.g., the negative value assigned to the act of killing a man) or to the outcome (e.g., the positive value assigned to the representation of saving more lives). According to the model, utilitarian decisions are more guided by outcome-based value representations (Cushman, 2013).

In the present study, the comparison between textual and VR versions of the dilemmas was further explored. In particular, modified versions of the Trolley dilemma applied to the driving context, in which participants play the role of drivers during inevitable collisions, were investigated. In fact, the introduction of the first autonomous cars created a renovated interest in these types of dilemmas, as a means to collect data about how people decide during extreme road accident situations. However, until now, only one study directly compared textual and virtual "traffic" dilemmas, but its findings did not replicate the previously described effects, making further investigation of this topic sorely needed.

Specifically, the VR moral dilemmas from Study 1 (see Chapter 5) were used and adapted to a text-based version. In these dilemmas, participants faced accident situations either in human driving mode (in which they could actually choose how to behave) or in autonomous driving mode (in which the machine decided for them). Furthermore, there was a third option in addition to the classic two possible resolutions: sacrifice yourself by crashing into a tree to save all the pedestrians present. After each dilemma, the perceived responsibility, acceptability, and emotional reactions (arousal and valence) related to the decision were assessed. Based on the above, the following hypotheses were formulated:

H.1 Participants in the condition with the dilemmas in text format tend to choose the self-sacrifice option to a greater extent than participants in the VR condition, as the text-form should be more prone to social desirability bias and lead to less focus on the output of the decision.

H.2 Participants in the VR condition experience more arousal and unpleasantness when facing moral dilemmas than participants in the text condition, as VR simulations are expected to be more emotionally engaging.

H.3 Only in the VR condition are there differences in emotional reactions between autonomous and human driving modes, as VR should allow participants to experience situations more realistically, making it easier to discriminate the emotions associated with them.

7.2.Methods

In the present study, in addition to the data collected *ex novo* for the text condition, data from Study 1 (VR condition) were reused. For this reason, only the new information regarding the text condition was reported in the Material and Procedure subsections. Similarly, in the Results section, only new analyses (comparisons between the two conditions and separate analyses for text condition) were reported. For detailed information and results of the VR condition, see chapter 5.

7.2.1. Experimental Design

A 2 (method) × 2 (driving mode) × 4 (dilemma type) mixed-participants design was employed. The levels of the first two variables (method and driving mode) were, respectively, VR and text and human driving and autonomous driving. The dilemma type variable consisted of 4 levels: two moral dilemmas and two non-moral control situations: three vs one (Numerosity dilemma), elder vs children (Age dilemma), one vs empty road, and elder vs empty road (non-moral dilemmas). As previously reported in the method section of Study 1, the car in the autonomous driving mode always chooses to kill the single person in the Numerosity dilemma, the elder in the Age dilemma, and to pick the empty road in the non-moral dilemmas. In the VR condition only, two training trials (one in human and one in autonomous driving mode) were employed at the start of the simulation to help participants to familiarize with the technology and the task. The results of the training trials were not included in the analysis. The participants in both VR and text conditions were exposed to all eight resulting trials, presented in random order.

7.2.2. Participants

36 participants took part in the study (15 in the VR condition and 21 in the text one; 16 males; age M = 24.39, SD = 2.27). Participants were recruited from the University of Padua. They voluntarily choose to participate in the experiment. For the VR condition, the analysis of the Responsibility scores was performed on only 11 participants, due to a possible bias linked to social desirability. For the same reason, the scores of Responsibility and Acceptability in textual condition were analyzed on 15 participants and not on the whole sample (for more information, see "Social Desirability Control Check" subsection). To participate a native-level understanding of Italian, having an age between 20 and 35 years and having possessed a driving license for at least one year were required. Exclusion criteria included having a history of migraines or motion sickness, having experienced previous car-related trauma, having had a seizure or history of epilepsy, and presenting vision problems that cannot be corrected with the use of glasses.

7.2.3. Material

For the text condition, the dilemmas and the questionnaires were prepared and administered using Google Forms. The mean number of words was fully balanced between driving mode and dilemma type (M words in moral human and autonomous driving scenarios = 50 and 49 and in non-moral human and autonomous driving scenarios = 50 and 49, H (3) = 7, P = .071). The textual description of each dilemma was accompanied by a graphical representation (in first-person view, as in the VR condition) (Figure 7.1.). For a detailed description of the VR simulation, see chapter 5.

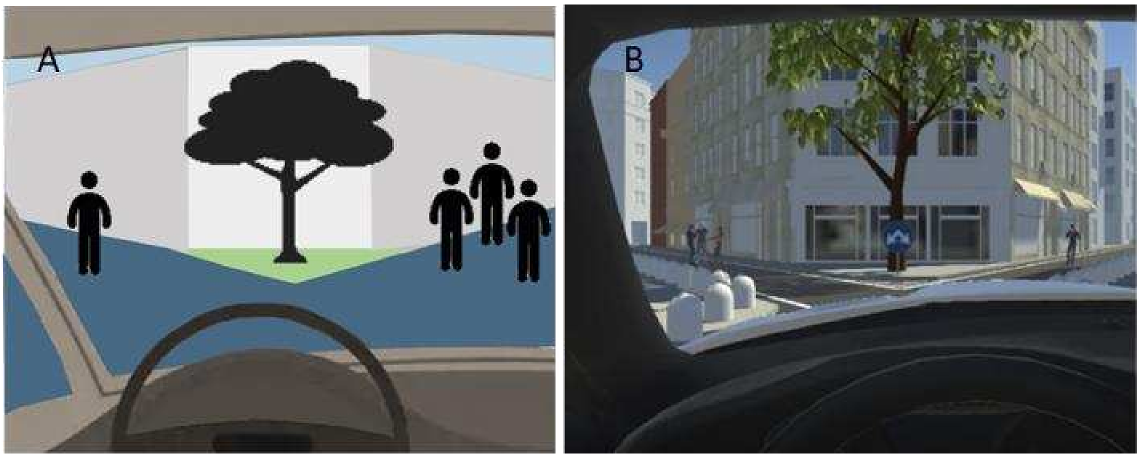


Figure 7.1. a) The graphical representation of the Numerosity dilemma in human driving mode used in text condition and b) The same dilemma in the VR condition.

7.2.4. Measures

For both VR and text conditions, the following measures were considered: the choice made in the dilemmas in human driving mode, (i.e., to turn right, left, or let the car go straight); the perceived responsibility, acceptability, arousal, and affective valence related to the decision in each dilemma (using a 9-point Likert scale to measure the first two variables and the SAM (Lang et al., 2008) for arousal and valence); demographic information (age and gender); the MC-SDS-b (Manganelli-Rattazzi et al., 2000).

7.2.5. Procedure

Participants in the text condition took part in the experiment via a link. After the compilation of the informed consent and the assessment of inclusion and exclusion criteria, they filled a brief demographic questionnaire and then read the instructions for the task: they will have to read some short descriptions of road scenarios in which they will play the role of driver. In some cases, the vehicle will be a traditional car, and they will have to choose the direction -left, right or none- that the car will go towards. In other cases, the vehicle will be a self-driving car: in this case, the car's decision will be reported (Table 7.1.). After every single choice (made by them or by the vehicle itself), a series of questions will be presented. At the end of the task, participants filled the MC-SDS-b questionnaire. See Chapter 5 for a description of the VR condition procedure.

Table 7.1. Textual description of the Numerosity dilemma in human and autonomous driving modes.

Human driving mode	Autonomous driving mode
Your car has a brake system failure while you are at high-speed approaching a junction. You cannot stop the car, but you can swerve. If you continue straight, you will crash into a tree. If you turn left, you will hit a person. If you turn right, you will hit three people. What do you do?	Your self-driving car has a brake system failure while at high speed near a junction. It can't stop, but it can swerve. If it continues straight, it will crash into a tree. If it turns left, it will hit a person. If it turns right, it will hit three people.
□ Go straight ahead: crashing into the tree will kill you, but all the other people will be safe. □ Turn left: you will kill a man, but you and the three people in the street on the right will be safe. □ Turn right: you will kill three people, but you and the person in the street on the left will be safe.	The car chooses to turn left: it will kill a man, but you and the three people in the road to the right will be saved.

7.3.Results

7.3.1. Choice

To compare choices in VR and text conditions, a series of Fisher's Exact Tests were performed on dilemmas in human driving mode. In case of significance, post-hoc comparisons with Benjamini-Hochberg were conducted.

6 trials of the VR condition (out of 120) ended in an unintended crash due to technical problems or (clearly reported) errors of the participants and were therefore removed from the analyses.

The two conditions showed a significantly different pattern of choices (P < .001). The post-hoc test revealed that participants in the text condition decided more frequently to sacrifice themselves than the ones in the VR condition, instead of opting for the utilitarian resolution (P < .001).

The same analyses were performed on the Numerosity dilemma and Age dilemma, separately. In both cases, a significant difference between VR and text conditions emerged (Numerosity dilemma: P < .01; Age dilemma: P = .04). However, after post-hoc tests, the difference remained significant in the Numerosity dilemma only, showing that participants in the text condition sacrificed themselves more often than participants who performed the task in VR (P < .01) (Figure 7.2.).

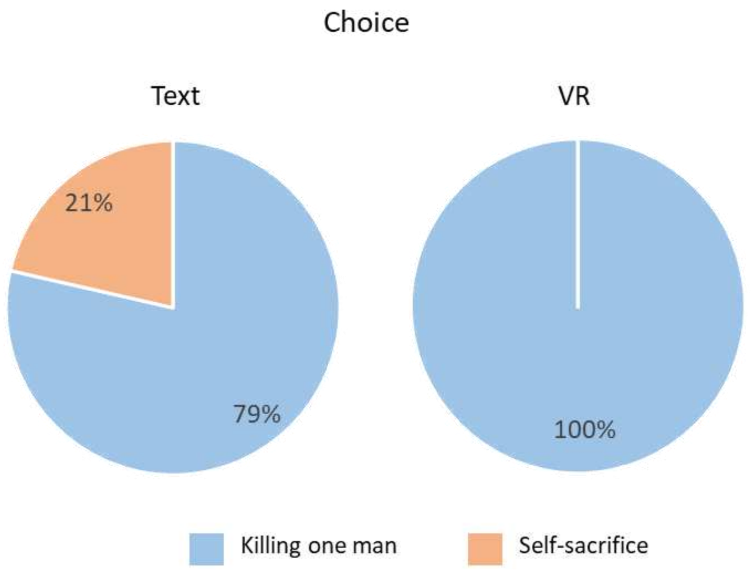


Figure 7.2. Percentages of choice in Numerosity dilemma in text and VR conditions. The "Killing three people" option was not reported because it was never chosen under either condition.

7.3.2. Self-Reports Results

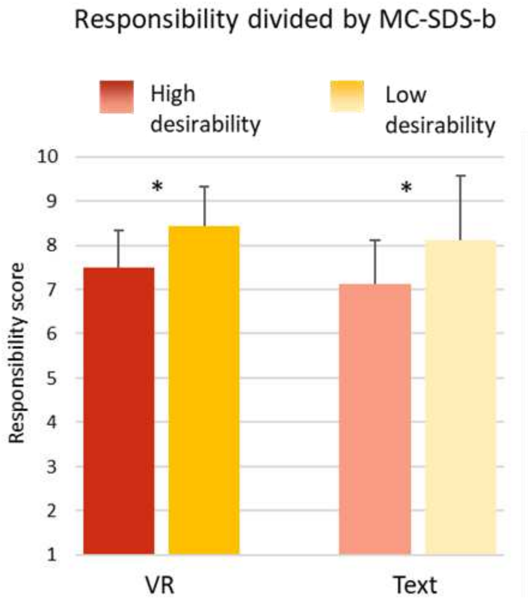
Social Desirability Control Check

To verify that participants in the text and VR conditions exhibited a similar baseline level of social desirability, a Mann-Whitney U test on the MC-SDS-b score was performed. No significant differences emerged (W = 197.5, P = 0.18).

As reported in chapter 5, the self-reported assessment of the perceived responsibility and acceptability in life/death decisions could be biased by social desirability. In particular, people with a high tendency to present themselves in a favorable light (high social desirability) could minimize their responsibility in socially condemned acts such as killing someone, to offer a better image of themselves. For this reason, an initial check on these two variables was performed in both VR and text conditions, dividing participants into two groups based on their scores on the MC-SDS-b ("high desirability" group, score> 5 -with 5 being the average score-, and "low desirability" group, score ≤ 5).

A series of Mann-Whitney U test tests were conducted to assess possible differences between high and low desirability groups, separately for VR and text conditions. When facing moral dilemmas in human driving modality, participants in the high desirability subset reported a significantly lower responsibility compared to participants in the low desirability one, in both VR (W = 26, P = .03, mean ranks = 5.5, 8.42) and text conditions (W = 28, P = .02, mean ranks = 6.25, 8.12) (Figure 7.3.).

No differences emerged in Acceptability scores or considering the distinction between Numerosity and Age dilemmas. To avoid biased results, the analyses of the responsibility scores presented in the following section were implemented considering the low desirability subset only (VR: 11 participants; text: 15 participants) and not in the complete sample.



M	7.50	8.42	7.12	8.12
SD	0.84	0.90	0.99	1.45

Figure 7.3. Mean and standard deviation of Responsibility in moral dilemmas in human driving mode, in high and low desirability groups, respectively in VR and text condition.

Post-Trial Evaluations Results

A first analysis was conducted to verify the presence of general differences between VR and text conditions in the four post-trials variables (Responsibility, Acceptability, Arousal, and Valence). To this aim, a weighted score of the variables was created separately for VR and text conditions by subtracting the score of the non-moral dilemmas from the score of the moral ones. Separate Mann-Whitney U tests were conducted to compare the weighted score of each variable between the two conditions. No significant differences were found.

A second analysis was run within each condition (text and VR) to investigate: 1) the impact of the driving mode (human vs autonomous) and the presence/absence of a moral dilemma (Numerosity + Age vs non-moral dilemmas) on post-trial evaluations (Fig. 7.4.); 2) the presence of differences in post-trial evaluations between the two types of moral dilemma (Numerosity vs Age). For this purpose, for each post-trial variable a Friedman rank sum test was performed, followed by Wilcoxon signed-rank tests for paired samples (with Benjamini-Hochberg correction) in case of significance. To have meaningful comparisons, only trials with typical results were included in these analyses.

Below are the results of the text condition (for a detailed description of the results of the VR condition, see chapter 5. For a schematic comparison of VR and text conditions, see Table 7.2.).

Concerning the impact of driving modalities and presence/absence of a moral dilemma situations, all the four variables were significant (Responsibility: $\chi2(3) = 60.82$, $P < .001$; Acceptability: $\chi2(3) = 89.73$, $P < .001$; Arousal: $\chi2(3) = 75.45$, $P < .001$; Valence: $\chi2(3) = 93.95$, $P < .001$).

Responsibility. From Wilcoxon comparisons emerged an effect of the driving modality: increased sense of responsibility in human driving mode compared to the autonomous mode in both moral ($V = 2$, $P < .001$, mean ranks = 8.11, 4.13) and non-moral decisions ($V = 0$, $P < .001$, mean ranks = 8.43, 3.8).

Acceptability. The only significant effect on acceptability score was the moral content, with higher acceptability for actions that did not involve killing a person in both driving modalities (Human driving: $V = 1.5$, $P < .001$, mean ranks = 8.66, 3.88; autonomous driving: $V = 0$, $P < .001$, mean ranks = 8.28, 4.57).

Arousal. No effect of the driving modality but only of the presence/absence of a moral situation was found: participants experienced more arousal when facing a moral situation compared to non-moral one in both human ($V = 768.5$, $P < .001$, mean ranks = 7.32, 3.64) and autonomous driving mode ($V = 741$, $P < .001$, mean ranks = 6.69, 3.23).

Valence. Only an effect of the moral content was found: participants experienced lower pleasantness when a person was killed in both human ($V = 2$, $P < .001$, mean ranks = 2.72, 7.83) and autonomous driving modality ($V = 0$, $P < .001$, mean ranks = 2.92, 7.85) compared to non-moral situations.

Regarding the differences between Numerosity and Age dilemmas, only the Friedman test for Responsibility was significant (χ2(3) = 37.95, P < .001). However, no differences between the two types of moral dilemma reach the significance threshold in the Wilcoxon comparisons analysis.

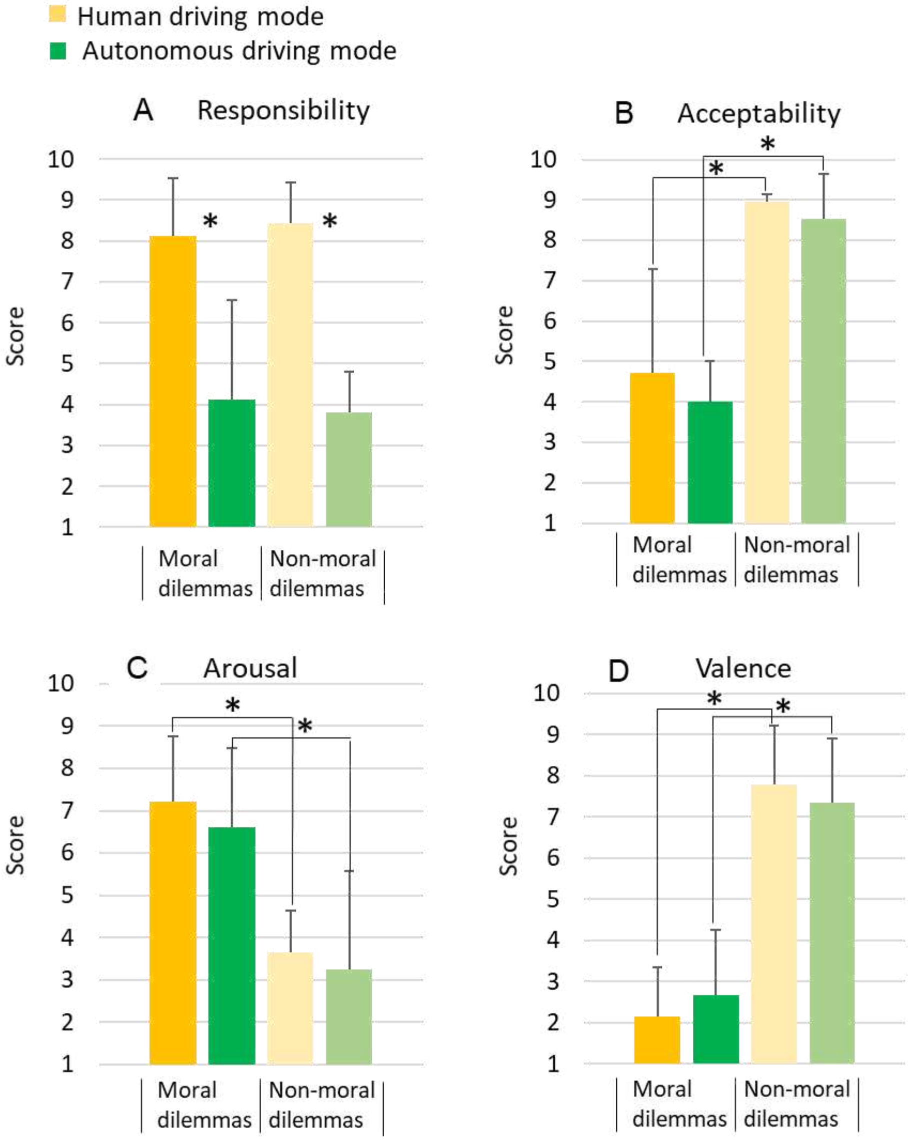


Figure 7.4. Bar graph by dilemma type and driving mode of a) responsibility, b) acceptability, c) arousal, and d) valence scores in the text condition. No distinction was made between Numerosity and Age dilemmas.

Table 7.2. Significant comparisons of the post-trial evaluation results in VR and text conditions. Cells in which only one of the two conditions was significant were highlighted in yellow.

	Moral dilemmas	Non-moral dilemmas	Human driving mode	Autonomous driving mode
	Human driving mode vs Autonomous driving mode	Human driving mode vs Autonomous driving mode	Moral dilemmas vs non-moral dilemmas	Moral dilemmas vs non-moral dilemmas
Responsibility	VR / text	VR / text		VR
Acceptability			VR / text	VR / text
Arousal	VR		VR / text	VR / text
Valence	VR	VR	VR / text	VR / text

7.4. Discussion

In the present study, VR and text versions of moral dilemmas in the driving context were compared. Differently from previous works, participants in both conditions faced the dilemmas in two different driving modalities: an autonomous driving mode, in which the car chose for them, and a human one, in which they could decide how to behave. Furthermore, participants had the opportunity not only to choose which potential victim/s kill to save the other/s, but they could decide to sacrifice themselves to save everyone else. Participants' experience was collected after each dilemma, assessing the responsibility, acceptability, arousal, and emotional valence related to the choice.

The first hypothesis of the study was confirmed: in the text condition, participants chose the self-sacrifice option significantly more often than those in the VR condition. This finding could be due to the tendency to respond to questionnaires and surveys in a way to present themselves in a favorable light (social desirability bias), a major concern in research on ethics and moral decision-making (Tan et al., 2021). Indeed, sacrificing oneself to save all other parties involved seems an overly altruistic response, reflecting a willingness to respond in line with social norms instead of one's own true beliefs. Other survey studies already found a preference for resolutions that involve the death of the drivers; for example, in the work of Bonnefon (2016) participants approved that AVs sacrifice their drivers/passengers for the greater good (however, they also reported that they would themselves prefer to buy an AVs that protect the passengers at all costs). Importantly, in the present study, participants in VR and text conditions did not differ in terms of individual social desirability score (as measured by the Marlowe-Crowne Social

Desirability Scale -brief form-), suggesting that the bias found in responses is related to the way dilemmas were presented (VR vs. text) rather than individual characteristics.

An alternative explanation to the behavioral results could be the difficulty to mentally simulate the scenario and feeling "present" in the situation described in the text condition. Indeed, while in the VR condition the participants were immersed in the dilemma scenario and could avoid any imaginative effort, the use of written stimuli forced the participants in the textual condition to create a mental representation of the dilemma and rely on it to make a decision. To help them in the process, the description of each dilemma was accompanied by a first-person graphical representation, however, it is not comparable to the direct experience provided by an immersive and interactive VR simulation. This interpretation fits nicely with Cushman's theory (2013) and with the results of previous works that compared textual and virtual presentation modes. Indeed, Francis et al. (2016) and Patil et al. (2014) suggested that the contextual saliency of the VR versions of moral dilemmas increased the negative value assigned to the outcome-based representations for not acting and letting the people to be killed, increasing utilitarian resolutions. Instead, in the textual modality, the reliance on imagination and the absence of salient features probably led to attributing a more negative value to the action of harming and less to the outcome of not acting (Francis et al., 2016; Patil et al., 2014). It is possible that a similar effect also occurred in the present study: in the "contextually enriched" VR simulation, seeing the imminent collision with the tree made the outcome of the self-sacrifice option more salient, increasing the probability of turning and killing an innocent pedestrian.

Regarding the second hypothesis, no general difference in emotional reactions between VR and text formats in moral dilemmas resolution was found. However, it is important to note that previous works comparing the VR and textual versions of the dilemmas measured emotion at the physiological level (arousal dimension only, through electrodermal activity and heart rate) (Francis et al., 2016, 2018, 2019; Navarrete et al., 2012; Patil et al., 2014), while in the present study arousal and valence were assessed using a brief self-report after each dilemma. It is possible that the difference between the two formats, at least for this type of situation, concerns only unconscious components of emotion.

Finally, the third research hypothesis was confirmed. Differences in emotional valence and arousal between autonomous and human driving modes emerged only in the VR condition. In extreme accidents involving the death of someone, the experience of being the actual driver was perceived as more intense and unpleasant than being in the same situation but with a machine making that difficult decision instead of the driver, as in the autonomous driving mode. In text condition instead, the only differences were between moral and non-moral (control) situations, while the two driving modalities elicited a similar level of valence and arousal. Again, these results could be due to the difficulty of feeling "present" in the situation described. It is not simple to perceive the emotional difference between driving modes while reading a text comfortably sitting in a room, with no time pressure and no real actions to perform other than a click of the mouse. Indeed, previous studies have already suggested that mental processes depend on the level of immersion in a situation similar to driving (Kochupillai et al., 2020; Madary & Metzinger, 2016). Also consistent with this interpretation is the fact that participants in the text condition did not show a differentiation in the perceived sense of responsibility between moral and non-moral dilemmas in autonomous driving mode. This "flattening" may have resulted from a less ecological and emotional but more conceptual assessment of the situation, from which perhaps emerged the conclusion "If I'm not the one driving, I'm not responsible, regardless of whether or not the car hits someone."

The present work shares some important limitations with the two previous studies. The main one is the absence of physiological measurements of emotion. Integrating subjective assessment with measures of the unconscious components of emotional processes could have helped to gain a clearer view of the participants' experience during dilemma resolution. Second, although the number of VR trials/textual stimuli was kept low, repetition of similar situations may have induced habituation in participants or, in any case, may have exerted an influence on their responses.

In conclusion, the present study highlights how using VR to study moral dilemma situations in driving contexts allows for minimizing the effect of social desirability and for obtaining more realistic behavioral responses and evaluations of emotional aspects. The adoption of ecological methods of administration, like virtual simulations, could be the key to obtaining better insight into the true moral preferences and reactions of drivers.

8. General discussion

As described in previous chapters, the Trolley and the Footbridge dilemmas (Foot, 1967; Thomson, 1985) have been extensively employed as a paradigm to investigate moral decision-making (chapter 1). Their use, until a few years ago, was almost exclusively through the presentation of textual stimuli, and they have been criticized for not being ecological and suffering from low external validity (Bauman et al., 2014). In recent times, two factors have led to a renewed interest in these dilemmas. First, VR technologies, which have made it possible to create more realistic forms of dilemmas and to test moral judgment theories in more engaging situations (e.g., Francis et al., 2016; Patil et al., 2014) (chapter 2). Second, the introduction of the first autonomous driving systems, whereby the Trolley dilemma (suitably adapted to the driving context) has become a means of collecting user preferences to create a bottom-up framework for the "decisional system" of AVs (chapter 3). To understand which are the typical choices of the drivers in inevitable road accidents, different situations have been tested, manipulating factors such as the number, age, gender, and position of the potential victims (Bergmann et al., 2018; Faulhaber et al., 2019; Frison et al., 2016; Ju et al., 2019, 2016; Sütfeld et al., 2017; Sütfeld, Ehinger, et al., 2019). However, there are still many fundamental aspects in this area that still need to be investigated.

The present Ph.D. thesis is part of these research lines and aims to investigate some of these aspects, measuring the behavior and emotional reactions in realistic VR versions of Trolley-type dilemmas applied in the driving context. The three studies described in this project had both practical and theoretical purposes. On the one hand, they aimed to offer some insight on possible effects related to the introduction of autonomous-driving technology, testing the influence of different driving modalities (human vs. autonomous; study 1) and legal liability frameworks (responsibility on the driver vs. on the company vs. no information; study 2) on the choice and emotional experience. On the other hand, they aimed to contribute to the discussion on the applicability of classic moral judgment theories to these more "concrete" and realistic versions of the dilemmas (VR dilemmas vs. text dilemmas; study 3).

In study 1 (chapter 5), the main objective was to verify whether the experience of being the actual driver making a choice in a Trolley-like accident situation was different compared to the experience of being in an AV facing the same dilemma, with no possibility of interfering with the decision of the car. In fact, almost all of the studies that have applied virtual moral dilemmas to the driving context have focused on how a human driver chooses (Bergmann et al., 2018; Faulhaber et al., 2019; Ju et al., 2019; S. Li et al., 2019; Lucifora et al., 2021; Sütfeld, Ehinger, et al., 2019), assuming that these preferences can be applied to autonomous driving. However, it is not certain that the same choice made by an AV is equally acceptable and has the same impact on the driver. To test this, participants in Study 1 faced a series of moral dilemmas in immersive VR, alternating between human and autonomous driving modes in random order. After each dilemma, their experience in terms of perceived responsibility, acceptability, emotional arousal, and valence was measured with a short self-report administered directly in VR, in order to not diminish the sense of presence. Results revealed that less unpleasantness and arousal were reported when participants faced accidents in autonomous driving mode than for crashes in which they were the actual drivers, probably because without any possibility to interfere with the vehicle's decisions they felt less responsible for the tragic consequences. Indeed, the self-reported score of responsibility was in line with this interpretation, showing a significant decrease in the autonomous driving mode. Interestingly, the pattern described for emotional valence is reversed in non-moral dilemmas (i.e., control situations in which it was possible to turn into an empty road, avoiding killing someone), suggesting that people find it less unpleasant to let the car act in their place only in extreme driving situations with very negative outcomes, but in neutral and simple conditions they prefer to be the ones to decide.

Overall, Study 1 outlines two interesting points on possible effects related to the introduction of autonomous-driving technology. First, it showed that although the acceptability assigned to decisions did not change significantly, the perceived sense of responsibility and emotional experience were instead strongly influenced by the driving mode. Knowing that people are generally skeptical about AVs (Edmonds, 2019) and averse to the concept of machines making moral decisions (Bigman & Gray, 2018), the research in this area could begin to also consider how they feel (and not just what they decide) when interacting with autonomous systems. This might help to better understand such distrust. Second, the pattern between moral and non-moral dilemmas showed an opposite direction compared to the current state of design of semi-autonomous systems (SAE level 3): these vehicles are programmed to act autonomously under normal driving conditions, but the driver must regain control in the event of a critical situation (SAE, 2021), whereas the participants in study 1 seem to prefer exactly the opposite.

Study 2 (chapter 6), on the other hand, focused on investigating whether the choice and the emotional experience in accidents involving semi-autonomous vehicles were influenced by legal liability, a factor that is still underexplored by studies in this field. To this end, participants were randomly assigned to one of three conditions: in one, they were informed that the driver is always legally responsible for the behavior of the vehicle, in another that the liability lies with the manufacturing company when the driver has not the control of the car, and in the last, they did not receive any information. They then faced a series of Trolley-type dilemmas in desktop VR in which they had to choose between letting the semi-autonomous car run over one or more potential victims on the main road or saving them by assuming control of the vehicle and turning onto a secondary road where another person stands. As in study 1, the participant's experience was measured with a brief self-report after each dilemma.

The main results highlighted two interesting points for the debate over AVs legislation. First, the emotional experience was influenced by the legal liability frameworks: participants in the "accountability on the company" condition reported to feel less morally responsible for crashes in autonomous driving mode, while the ones who received no information on legal consequences showed increased emotional activation. This latter finding pointed out how having a clear understanding of who is legally responsible could help make these situations less stressful.

Second, the manipulation of legal responsibility, while influencing the participants' experience, did not affect their behavioral choices. Particularly, adopting a legal framework that attributes liability to manufacturing companies did not make the participants more prone to avoid retaking control of the car, although leaving the vehicle in autonomous mode would have avoided any legal issues for them. This suggests the importance for a driver to have the possibility to interfere when the course of action decided by the semi-autonomous car is at odds with their preference.

It is important to note that this finding is not in contradiction with the previously discussed result of Study 1 (i.e., participants find it a less stressful experience to let the car act in their place in moral dilemma situations than to decide themselves). In fact, in study 1, the car in autonomous driving mode was programmed to always choose the utilitarian resolution (kill one life to save more), the same decision made by all the participants in the human driving mode. In that experiment, there was no disagreement between the human driver and the vehicle (and, in any case, there was no possibility of interfering, being the car completely autonomous). In study 2, if participants did not choose to take control and intervene, the semi-autonomous vehicles would run over the "preferred" target (or, at least, the target known from the literature to be the one that is most often saved). Taken together, these results suggest that when the preferences of the vehicle and driver are aligned, it is less activating and unpleasant to let the car in autonomous mode run over the victim, but in the event of no agreement the driver prefers to be able to intervene, even at cost of becoming legally responsible for a person's death. The importance of finding a way of maintaining personal autonomy in self-driving systems is highlighted also by recent ethical guidelines, which have recommended the inclusion of an “override function” to allow drivers to maintain some degree of agency at least up to SAE level 4 (Luetge, 2017; Lütge et al., 2021).

Finally, study 3 (chapter 7) aimed to verify whether the method adopted (text or VR) influences the decision and emotional reactions during unavoidable accident situations. Indeed, previous studies with classic Footbridge and Trolley dilemmas found important differences related to the presentation method (Francis et al., 2016; Patil et al., 2014), suggesting that moral judgment and moral action could be distinct constructs. In study 3, the VR versions of the dilemmas used in study 1 were compared with text transpositions of the same dilemmas. Results showed that participants in the text condition chose more often than those in the VR condition to sacrifice themselves to save anyone else. This discrepancy in the behavioral output complements and extends the findings of previous work. Francis et al. (2016) and Patil et al. (2014) argue that the increase in arousal and utilitarian responses found in their VR versions of the dilemmas is due to the augmented “contextual richness” offered by this media. Consequently, experiencing, and not only abstractly imagining, the situation represented in the dilemma, would lead to a greater focus on the tragic consequences (i.e., the death of a greater number of people) and to assign a less negative value to the action itself necessary to avoid this outcome (i.e., killing a man) (Francis et al., 2016; Patil et al., 2014). Similarly, it is possible that participants in Study 3 who faced dilemmas in VR found the outcomes of their inaction (i.e., dying by crashing into a tree) more salient than the action itself of turning and killing a single pedestrian. This interpretation is more in line with the dual-process model proposed by Cushman (2013) than with the version of Greene et al. (2001, 2004) (see chapter 1). Cushman's model (2013) in fact advocates the presence of contraposition between action-based and outcome-based value representations and argues that the latter plays a predominant role in the Trolley dilemma, favoring the choice of the utilitarian option. In this light, the more activating and engaging a virtual scenario is, the stronger the negative value assigned to outcomes should be, increasing the probability of choosing the option that with the least negative outcome (i.e., saving more lives in Trolley dilemma, saving themselves in the modify Trolley problem of study 3). In contrast, Greene's model (Greene et al., 2001, 2004) although it continues to effectively explain moral judgment in the textual forms of dilemmas, does not seem applicable to moral action in VR scenarios. Indeed, according to this model, a greater role of emotional processes should correspond to a greater preference for choosing not to kill. An expectation that is at odds with the results described above. On this basis, Study 3 fits into the dialogue about the applicability of moral judgment theories to more realistic dilemma situations, supporting a distinction between moral judgment and moral action.

From a more practical and less theoretical perspective, Study 3 also highlighted that to study differences in emotional experience between different driving modes, VR is better suited than text-based methods. In fact, many of the differences found in study 1 between autonomous and human driving modes were not significant with text-based versions of the same dilemmas, in line with previous studies suggesting that mental processes depend on the level of immersion in a situation similar to driving (Kochupillai et al., 2020; Madary & Metzinger, 2016).

Taken together, the results of the three studies open the way for new areas of investigation in this research field. As mentioned earlier, they underscore the relevance of considering emotional aspects and not just behavioral output. Knowing the impact on the driver of a certain driving mode or legal framework provides information about the possible critical issues that could result from their adoption. For example, the tendency to feel less morally responsible for accidents in autonomous driving mode (study 1), especially when legal liability is assigned to a third party (study 2), is an issue that should not be underestimated and should be further investigated in future studies. It would be interesting to test whether such moral deresponsibility persists even when giving the user the opportunity to choose *a priori* how the AV algorithm should decide in such situations and then experience the programmed behavior in a virtual environment.

In addition, although these studies have shown the effectiveness of virtual simulations in studying moral decision-making in the driving context compared to more traditional methods (study 3), there is still a lack of systematic investigation into the differences between desktop VR and immersive VR in this area. From a methodological standpoint, the use of desktop VR would have the advantage of achieving substantial sample sizes. The system implemented in study 2, for example, allowed the experiment to be conducted directly on the participant's computer, accessed via a link. It was also possible to collect data from multiple users performing the task simultaneously. In this way, the use of such systems would allow for the same degree of ubiquity as studies using online surveys, but with a greater degree of realism. On the other hand, immersive VR may be better suited to emotionally engage the user. Further research is therefore needed in this area, in order to understand which system to choose according to specific research purposes.

The studies presented in this thesis share some important limitations with other related studies in the literature. As described in more detail in the discussions of the individual studies, the very choice to use modified versions of the Trolley dilemma can be considered a limitation. Such dilemmas, even if made more realistic using VR remain fairly unlikely situations. Although in study 1 and 3 an attempt was made to break out of the dichotomy of the dilemma by offering a third option to the participant, in a real accident situation the possibilities for action are many more. In addition, in a real accident, it is often unclear to drivers what the outcomes of their actions will be. The decision to use abundantly studied situations such as the Trolley dilemma allowed for control of complexity and the ability to compare with similar studies, but at the expense of generalizability of the results to real-world situations. In addition, repetition of similar dilemma situations across multiple trials may have influenced participants' emotional reactions, despite technical precautions taken to avoid habituation. Finally, in all three studies presented in this thesis, emotional processes were measured through self-report scales. These methods assume that individuals are able to assess their emotional state. However, not all are equally aware of their emotional responses, or accurate in assessing them. As a result, self-reports provide only an approximate measure of the actual emotional reactions felt by participants (Larsen & Fredrickson, 1999) and give only information about those emotional states that have reached awareness.

To conclude, this dissertation, despite its limitations, showed how VR can be effectively used to study behavior and emotional experience in extreme accident situations, shedding some light into the moral decision-making research field.

References

Aquinas, T. (1947). *Summa Theologiae. II-II, Q. 64, art. 7. (Original work published in 1265–1272.)*. Benzinger Bros. Edition.

Awad, E., Dsouza, S., Bonnefon, J. F., Shariff, A., & Rahwan, I. (2020). Crowdsourcing moral machines. *Communications of the ACM*, *63*(3), 48–55. https://doi.org/10.1145/3339904

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J. F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., Bonnefon, J. F., & Rahwan, I. (2020). Drivers are blamed more than their automated cars when both make mistakes. *Nature Human Behaviour*, *4*(2), 134–143. https://doi.org/10.1038/s41562-019-0762-8

Bamodu, O., & Ye, X. M. (2013). Virtual reality and virtual reality system components. *Advanced Materials Research*, *765–767*, 1169–1172. https://doi.org/10.4028/www.scientific.net/AMR.765-767.1169

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv Preprint ArXiv:1406.5823.*

Bauman, C. W., Mcgraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554. https://doi.org/10.1111/spc3.12131

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience*, *8*(11), 1458–1463. https://doi.org/10.1038/nn1584

Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D., & Knight, R. T. (2003). The regulatory function of self-conscious emotion: insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology*, *85*(4), 594–604.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, *30*(5), 961–981. https://doi.org/doi: 10.1287/opre.30.5.961

Benvegnu, G., Pluchino, P., & Garnberini, L. (2021). Virtual morality: Using virtual reality to study moral behavior in extreme accident situations. *Proceedings - 2021 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2021*, 316–325. https://doi.org/10.1109/VR50410.2021.00054

Bergmann, L. T., Schlicht, L., Meixner, C., König, P., Pipa, G., Boshammer, S., & Stephan, A. (2018). Autonomous vehicles require socio-political acceptance—an empirical and philosophical perspective on the problem of moral decision making.

Frontiers in Behavioral Neuroscience, *12*(31), 1–12. https://doi.org/10.3389/fnbeh.2018.00031

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*(March), 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Bimbraw, K. (2015). Autonomous Cars: Past, Present and Future. *2015 12th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 191–198. https://www.scitepress.org/Papers/2015/55405/55405.pdf

Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication, 3*(2), JCMC324.

Blascovich, J., Loomis, J. M., Beall, A. C., Swinth, K. R., Hoyt, C. L., Bailenson, N., & Bailenson, J. N. (2002). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, *12*(12), 752–762. https://doi.org/10.1038/nrn3122

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Bono, R., Alarcón, R., & Blanca, M. J. (2021). Report Quality of Generalized Linear Mixed Models in Psychology: A Systematic Review. *Frontiers in Psychology*, *12*(April), 1–15. https://doi.org/10.3389/fpsyg.2021.666182

Borg, J. S., Hynes, C., Horn, J. Van, & Grafton, S. (2006). Consequences, Action, and Intention as Factors in Moral Judgements. *Journal of Cognitive Neuroscience*, *18*(5), 803–817.

Brooks, M., Kristensen, K., van Benthem, KJ Magnusson, A., Berg, C., Nielsen, A., Skaug, H., Maechler, M., & Bolker, B. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–400. journal.r-project.org/archive/2017/RJ-2017-066/index.html

Campbell, M., Egerstedt, M., How, J. P., & Murray, R. M. (2010). Autonomous driving in urban environments: Approaches, lessons and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1928), 4649–4672. https://doi.org/10.1098/rsta.2010.0110

Charness, N., Yoon, J. S., Souders, D., & Stothart, C. (2018). *Predictors of Attitudes Toward Autonomous Vehicles : The Roles of Age , Gender , Prior Knowledge , and Personality*. *9*(December), 1–9. https://doi.org/10.3389/fpsyg.2018.02589

Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. A traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics*, *21*(4), 529–538. https://doi.org/10.1109/TVCG.2015.2391853

Choe, S. Y., & Min, K. H. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making*, *6*(7), 580–592.

Ciaramelli, E., Muccioli, M., Làdavas, E., & Di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *2*(2), 84–92. https://doi.org/10.1093/scan/nsm001

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*(6625), 604–608.

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: the design and implementation of the CAVE. *Proc ACM SIGGRAPH 93 Conf Comput Graphics*, 135–142.

Cunningham, M. L., Ledger, S. A., & Regan, M. (2018). A survey of public opinion on automated vehicles in Australia and New Zealand. *28th ARRB International Conference–Next Generation Connectivity*.

Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*(1), 2–7. https://doi.org/10.1037/a0025071

Cushman, F., Young, L., & Greene, J. D. (2010). Our multi-system moral psychology: Towards a consensus view. In J. Doris, G. Harman, S. Nichols, J. Prinz, W. Sinnott-Armstrong, & S. Stich (Eds.), *The Oxford Handbook of Moral Psychology* (pp. 47–72). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199582143.003.0003

Cushman, F., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment Testing Three Principles of Harm. *Psychological Science*, *17*(12), 1082–1089.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, *41*(2), 81–94.

de Waal, F. (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals.* Harvard Univerthesity Press.

Dogan, E., Barbier, C., & Peyrard, E. (2021). Public perception of ethical issues concerning automated mobility. *European Conference on Cognitive Ergonomics (ECCE)*, 1–6. https://doi.org/10.1145/3452853.3452877

Edmonds, E. (2019). Three in Four Americans Remain Afraid of Fully Self-Driving Vehicles. *American Automobile Association*.

Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., Stephan, A., Pipa, G., & König, P. (2019). Human Decisions in Moral Dilemmas are Largely Described by Utilitarianism: Virtual Car Driving Study Provides Guidelines for Autonomous Driving Vehicles. *Science and Engineering Ethics*, *25*(2), 399–418.

https://doi.org/10.1007/s11948-018-0020-x

Feng, Y., Duives, D. C., & Hoogendoorn, S. P. (2022). Wayfinding behaviour in a multi-level building: A comparative study of HMD VR and Desktop VR. *Advanced Engineering Informatics*, *51*(June 2021), 101475. https://doi.org/10.1016/j.aei.2021.101475

Fleetwood, J. (2017). Public health, ethics, and autonomous vehicles. *American Journal of Public Health*, *107*(4), 532–537. https://doi.org/10.2105/AJPH.2016.303628

Foot, P. (1967). The Problem of Abortion and the Doctrine of the Double Effect. *Oxford Review*, *5*, 5–15.

Fox, J. (1987). Effect displays for generalized linear models. *Sociol. Methodol.*, *17*, 347–361. https://doi.org/10.2307/271037

Francis, K. B., Gummerum, M., Ganis, G., Howard, I. S., & Terbeck, S. (2018). Virtual morality in the helping professions: Simulated action and resilience. *British Journal of Psychology*, *109*(3), 442–465. https://doi.org/10.1111/bjop.12276

Francis, K. B., Gummerum, M., Ganis, G., Howard, I. S., & Terbeck, S. (2019). Alcohol, empathy, and morality: acute effects of alcohol consumption on affective empathy and moral decision-making. *Psychopharmacology*, *236*(12), 3477–3496. https://doi.org/10.1007/s00213-019-05314-z

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PLoS ONE*, *11*(10), 1–22. https://doi.org/10.1371/journal.pone.0164374

Francis, K. B., Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., & Howard, I. S. (2017). Simulating Moral Actions: An Investigation of Personal Force in Virtual Moral Dilemmas. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-13909-9

Friedman, D., Pizarro, R., Or-Berkers, K., Neyret, S., Pan, X., & Slater, M. (2014). A method for generating an illusion of backwards time travel using immersive virtual reality - an exploratory study. *Frontiers in Psychology*, *5*(943). https://doi.org/10.3389/fpsyg.2014.00943

Frison, A.-K. K., Wintersberger, P., & Riener, A. (2016). First Person Trolley Problem: Evaluation of Drivers' Ethical Decisions in a Driving Simulator. *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 117–122. https://doi.org/10.1145/3004323.3004336

Fujimi, T., & Fujimura, K. (2020). Testing public interventions for flash flood evacuation through environmental and social cues: The merit of virtual reality experiments. *International Journal of Disaster Risk Reduction*, *50*(February), 101690. https://doi.org/10.1016/j.ijdrr.2020.101690

Gamberini, L., Cottone, P., Spagnolli, A., Varotto, D., & Mantovani, G. (2003). Responding to a fire emergency in a virtual environment: Different patterns of action for different situations. *Ergonomics*, *46*(8), 842–858.

https://doi.org/10.1080/00140130310001111266

Gamberini, L., & Spagnolli, A. (2015). An Action-Based Approach to Presence: Foundations and Methods. In M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, & R. Schaevitz (Eds.), *Immersed in Media* (pp. 101–114). Springer.

Gao, P., Hensley, R., & Zielke, A. (2014). A road map to the future for the auto industry | McKinsey &amp; Company. *McKinsey Quarterly*, *October*, 1–11. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/a-road-map-to-the-future-for-the-auto-industry

Geistfeld, M. A. (2017). A roadmap for autonomous vehicles: State tort liability, automobile insurance, and federal safety regulation. *Calif. L. Rev.*, *105*, 1611.

Goodall, N. J. (2014). *Machine Ethics and Automated Vehicles*. 93–102. https://doi.org/10.1007/978-3-319-05990-7_9

Grasso, G. M., Lucifora, C., Perconti, P., & Plebe, A. (2020). Integrating human acceptable morality in autonomous vehicles. *Advances in Intelligent Systems and Computing*, *1131 AISC*, 41–45. https://doi.org/10.1007/978-3-030-39512-4_7

Greene, J. D., & Haidt, J. (2002). How (and Where) Does Moral Judgment Work? *Trends in Cognitive Sciences*, *6*(12), 517–523.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The Neural Basis of Cognitive Conflict and Control in Moral Judgment. *Neuron*, *44*(2), 389–400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, *31*(1), 191–221. https://doi.org/10.1111/j.1559-1816.2001.tb02489.x

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, *13*(1), 1–6. https://doi.org/10.1016/j.tics.2008.09.006

Hulse, L. M., Xie, H., & Galea, E. R. (2018). Perceptions of autonomous vehicles : Relationships with road users , risk , gender and age. *Safety Science*, *102*(August 2017), 1–13. https://doi.org/10.1016/j.ssci.2017.10.001

Jeurissen, D., Sack, A. T., Roebroeck, A., Russ, B. E., & Pascual-Leone, A. (2014). TMS affects moral judgment, showing the role of DLPFC and TPJ in cognitive and emotional processing. *Frontiers in Neuroscience*, *8*(8 FEB), 1–9. https://doi.org/10.3389/fnins.2014.00018

Ju, U., Kang, J., & Wallraven, C. (2019). To brake or not to brake? Personality traits predict decision-making in an accident situation. *Frontiers in Psychology*, *10*(134), 1–11. https://doi.org/10.3389/fpsyg.2019.00134

Ju, U., Kang, J., & Wallraven, C. (2016). Personality differences predict decision-making in an accident situation in virtual driving. *Proceedings - IEEE Virtual Reality*, *2016-July*, 77–82. https://doi.org/10.1109/VR.2016.7504690

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *The American Psychologist*, *58*(9), 697–720.

Kallioinen, N., Pershina, M., Zeiser, J., Nosrat Nezami, F., Pipa, G., Stephan, A., König, P., Nezami, F. N., Stephan, A., Pipa, G., König, P., Nosrat Nezami, F., Pipa, G., Stephan, A., & König, P. (2019). Moral Judgements on the Actions of Self-Driving Cars and Human Drivers in Dilemma Situations From Different Perspectives. *Frontiers in Psychology*, *10*(2415), 1–29. https://doi.org/10.3389/fpsyg.2019.02415

Kang, S., Chanenson, J., Ghate, P., Cowal, P., Weaver, M., & Krum, D. M. (2019). Advancing ethical decision making in virtual reality. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*, 1008–1009. https://doi.org/10.1109/VR.2019.8798151

Kant, I. (1959). *Foundation of the metaphysics of morals. (Original work published in 1785).* BobbsMerrill.

Kardong-Edgren, S. (Suzie), Farra, S. L., Alinier, G., & Young, H. M. (2019). A Call to Unify Definitions of Virtual Reality. *Clinical Simulation in Nursing*, *31*, 28–34. https://doi.org/10.1016/j.ecns.2019.02.006

Kinateder, M., Ronchi, E., Nilsson, D., Kobes, M., Müller, M., Pauli, P., & Mühlberger, A. (2014). Virtual Reality for Fire Evacuation Research. *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems Pp.*, *2*, 313–321. https://doi.org/10.15439/2014F94

Kochupillai, M., Lütge, C., & Poszler, F. (2020). Programming Away Human Rights and Responsibilities? “The Moral Machine Experiment” and the Need for a More “Humane” AV Future. *NanoEthics*, *14*(3), 285–299. https://doi.org/10.1007/s11569-020-00374-4

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911. https://doi.org/10.1038/nature05631

Kuehne, M., Heimrath, K., Heinze, H. J., & Zaehle, T. (2015). Transcranial Direct Current Stimulation of the Left Dorsolateral Prefrontal Cortex Shifts Preference of Moral Judgments. *PloS One*, *10*(5).

Kyriakidis, M., Happee, R., & Winter, J. C. F. De. (2015). Public opinion on automated driving : Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Psychology and Behaviour*, *32*, 127–140. https://doi.org/10.1016/j.trf.2015.04.014

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. In *Technical Report*. University of Florida.

Larsen, R. J., & Fredrickson, B. L. (1999). *Measurement issues in emotion research. Well-being: Foundations of hedonic psychology* (D. Kahneman, E. Diener, & N. Schwarz (Eds.)). Russell Sage.

Leder, J., Horlitz, T., Puschmann, P., Wittstock, V., & Schütz, A. (2019). Comparing immersive virtual reality and powerpoint as methods for delivering safety training: Impacts on risk perception, learning, and decision making. *Safety Science*, *111*, 271–286. https://doi.org/10.1016/j.ssci.2018.07.021

Length, R. V. (2020). *Estimated Marginal Means, aka Least-Squares Means.*

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences*, *14*(10), 457–463. https://doi.org/10.1016/j.tics.2010.08.001

Li, J., Zhao, X., Cho, M. J., Ju, W., & Malle, B. F. (2016). From Trolley to Autonomous Vehicle: Perceptions of Responsibility and Moral Norms in Traffic Accidents with Self-Driving Cars. *SAE Technical Papers*, *2016-April*(April). https://doi.org/10.4271/2016-01-0164

Li, S., Zhang, J., Li, P., Wang, Y., & Wang, Q. (2019). Influencing factors of driving decision-making under the moral dilemma. *IEEE Access*, *7*, 104132–104142. https://doi.org/10.1109/ACCESS.2019.2932043

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, *34*, 199–249.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2), JCMC321.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, *92*(368), 805–824.

Lotto, L., Manfrinati, A., & Sarlo, M. (2014). *A New Set of Moral Dilemmas : Norms for Moral Acceptability , Decision Times , and Emotional Salience*. *65*(May 2013), 57–65. https://doi.org/10.1002/bdm

Lucifora, C., Grasso, G. M., Perconti, P., & Plebe, A. (2021). Moral reasoning and automatic risk reaction during driving. *Cognition, Technology and Work*, *VanRullen 2017*. https://doi.org/10.1007/s10111-021-00675-y

Lucifora, C., & Grasso, M. (2020). *Moral dilemmas in self-driving cars*. *11*, 238–250. https://doi.org/10.4453/rifp.2020.0015

Luetge, C. (2017). The German Ethics Code for Automated and Connected Driving. *Philosophy and Technology*, *30*(4), 547–558. https://doi.org/10.1007/s13347-017-0284-0

Lütge, C., Poszler, F., Acosta, A. J., Danks, D., Gottehrer, G., Mihet-Popa, L., & Naseer, A. (2021). AI4people: Ethical guidelines for the automotive sector-fundamental requirements and practical recommendations. *International Journal of Technoethics*, *12*(1), 101–125. https://doi.org/10.4018/IJT.20210101.oa2

Madary, M., & Metzinger, T. K. (2016). Real virtuality: A code of ethical conduct. Recommendations for good scientific practice and the consumers of VR-technology. *Frontiers Robotics AI*, *3*(FEB), 1–23. https://doi.org/10.3389/frobt.2016.00003

Manfrinati, A., Lotto, L., Sarlo, M., & Palomba, D. (2013). Moral dilemmas and moral principles : When emotion and cognition unite. *Cognition and Emotion*, *27*(7), 1276–1291. https://doi.org/10.1080/02699931.2013.785388

Manganelli-Rattazzi, A. M., Canova, L., & Marcorin, R. (2000). La desiderabilità sociale. Un'analisi di forme brevi della scala di Marlowe e Crowne [Social desirability. An analysis of short forms of the Marlowe-Crowne Social Desirability Scale]. *Testing, Psychometrics, Methodology in Applied Psychology*, *7*(1), 5–17.

Martirosov, S., Bureš, M., & Zítka, T. (2021). Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality. *Virtual Reality*, *0123456789*. https://doi.org/10.1007/s10055-021-00507-4

Mayer, M. M., Bell, R., & Buchner, A. (2021). Self-protective and self-sacrificing preferences of pedestrians and passengers in moral dilemmas involving autonomous vehicles. *Plos One*, *16*(12), e0261673. https://doi.org/10.1371/journal.pone.0261673

McDonald, M. M., Defever, A. M., & Navarrete, C. D. (2017). Killing for the greater good: Action aversion and the emotional inhibition of harm in moral dilemmas. *Evolution and Human Behavior*, *38*(6), 770–778. https://doi.org/10.1016/j.evolhumbehav.2017.06.001

McManus, R. M., & Rutchick, A. M. (2019). Autonomous Vehicles and the Attribution of Moral Responsibility. *Social Psychological and Personality Science*, *10*(3), 345–352. https://doi.org/10.1177/1948550618755875

Mellers, B., Schwarz, A., & Ritov, I. (1999). Emotion-Based Choice (1999) Mellers, Schwarz etc.pdf. In *The American Psychological Association* (Vol. 128, Issue 3, pp. 332–345).

Miller, R., & Cushman, F. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social and Personality Psychology Compass*, *7*(10), 707–718. https://doi.org/10.1111/spc3.12066

Miller, R., Hannikainen, I. A., & Cushman, F. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion*, *14*(3), 573–587. https://doi.org/10.1037/a0035361

Moll, J., & de Oliveira-Souza, R. (2007). Response to Greene: Moral sentiments and reason: friends or foes? *Trends in Cognitive Sciences*, *11*(8), 323–324. https://doi.org/10.1016/j.tics.2007.06.004

Moll, J., de Oliveira-Souza, R., & Zahn, R. (2008). The neural basis of moral cognition: Sentiments, concepts, and values. In A. Kingstone & M. B. Miller (Eds.), *The year in cognitive neuroscience 2008* (pp. 161–180). Blackwell Publishing.

National Highway Traffic Safety Administration. (2017). *Automatic vehicle control* *systems–investigation of Tesla accident*.

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: emotion and action in a simulated three-dimensional "trolley problem." *Emotion*, *12*(2), 364–370. https://doi.org/10.1037/a0025561

Niforatos, E., Palma, A., Gluszny Solarwinds, R., & Liarokapis, F. (2020). Would you do it?: Enacting Moral Dilemmas in Virtual Reality for Understanding Ethical Decision-Making. *Conference: 2020 CHI Conference on Human Factors in Computing Systems (CHI 2020)*, *March*. https://doi.org/10.1145/3313831.3376788

Othman, K. (2021). Public acceptance and perception of autonomous vehicles: a comprehensive review. In *AI and Ethics* (Issue 0123456789). Springer International Publishing. https://doi.org/10.1007/s43681-021-00041-8

Pallavicini, F., Pepe, A., & Minissi, M. E. (2019). Gaming in Virtual Reality: What Changes in Terms of Usability, Emotional Response and Sense of Presence Compared to Non-Immersive Video Games? *Simulation and Gaming*, *50*(2), 136–159. https://doi.org/10.1177/1046878119831420

Pan, X., & Slater, M. (2011). Confronting a Moral Dilemma in Virtual Reality: A Pilot Study. *25th BCS Conference on Human-Computer Interaction*, 46–51. citeulike-article-id:13326879%5Cnhttp://portal.acm.org/citation.cfm?id=2305326

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107. https://doi.org/10.1080/17470919.2013.870091

Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2014.00501

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, *17*(3), 145–171. https://doi.org/10.1016/0162-3095(96)00041-6

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*(3), 467–478. https://doi.org/doi.org/10.1016/0162-3095(96)00041-6

Pletti, C., Lotto, L., Tasso, A., & Sarlo, M. (2016). Will I regret it? Anticipated negative emotions modulate choices in moral dilemmas. *Frontiers in Psychology*, *7*(DEC), 1–15. https://doi.org/10.3389/fpsyg.2016.01918

Pletti, C., Sarlo, M., Palomba, D., Rumiati, R., & Lotto, L. (2015). Evaluation of the legal consequences of action affects neural activity and emotional experience during the resolution of moral dilemmas. *Brain and Cognition*, *94*, 24–31. https://doi.org/10.1016/j.bandc.2015.01.004

Pöllänen, E., Read, G. J. M., Lane, B. R., Thompson, J., & Salmon, P. M. (2020). Who is to blame for crashes involving autonomous vehicles? Exploring blame attribution across the road transport system. *Ergonomics*, *63*(5), 525–537.

https://doi.org/10.1080/00140139.2020.1744064

Pulijala, Y., Ma, M., Pears, M., Peebles, D., & Ayoub, A. (2018). Effectiveness of Immersive Virtual Reality in Surgical Training—A Randomized Control Trial. *Journal of Oral and Maxillofacial Surgery*, *76*(5), 1065–1072. https://doi.org/10.1016/j.joms.2017.10.002

R Core Team. (2015). *RStudio: integrated development for R*.

Ramirez, E. J. E. J. (2018). Ecological and ethical issues in virtual reality research: A call for increased scrutiny. *Philosophical Psychology*, *32*(2), 211–233. https://doi.org/10.1080/09515089.2018.1532073

Rhim, J., Lee, G. bbeum, & Lee, J. H. (2020). Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of Korea and Canada. *Computers in Human Behavior*, *102*(April 2019), 39–56. https://doi.org/10.1016/j.chb.2019.08.010

Richardson, N., Doubek, F., Kuhn, K., & Stumpf, A. (2017). Assessing Truck Drivers' and Fleet Managers' Opinions Towards Highly Automated Driving. *Advances in Human Aspects of Transportation*, 473–484.

Riva, G. (2002). Virtual reality for health care: The status of research. *Cyberpsychology & Behavior*, *5*(3), 219-225.

Ronchi, E., Kinateder, M., Müller, M., Jost, M., Nehfischer, M., Pauli, P., & Mühlberger, A. (2015). Evacuation travel paths in virtual reality experiments for tunnel safety analysis. *Fire Safety Journal*, *71*, 257–267. https://doi.org/10.1016/j.firesaf.2014.11.005

Rovira, A. (2009). The use of virtual reality in the study of people's responses to violent incidents. *Frontiers in Behavioral Neuroscience*, *3*(59), 1–10. https://doi.org/10.3389/neuro.08.059.2009

Sacks, R., Perlman, A., & Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics*, *31*(9), 1005–1017. https://doi.org/10.1080/01446193.2013.828844

SAE. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (No. J3016_202104)*. SAE International.

Sarlo, M., Lotto, L., Manfrinati, A., Rumiati, R., Gallicchio, G., & Palomba, D. (2012). Temporal dynamics of cognitive-emotional interplay in moral decision-making. *Journal of Cognitive Neuroscience*, *24*(4), 1018–1029. https://doi.org/10.1162/jocn_a_00146

Sarlo, M., Lotto, L., Rumiati, R., & Palomba, D. (2014). If it makes you feel bad, don't do it! Egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas. *Physiology and Behavior*, *130*, 127–134. https://doi.org/10.1016/j.physbeh.2014.04.002

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in 'theory of mind'. *Neuroimage*, *19*(4), 1835–1842.

Schroeder, R. (1996). *Possible Worlds: The Social Dynamic of Virtual Reality Technologies.* Westview Press.

Shendarkar, A., Vasudevan, K., Lee, S., & Son, Y. J. (2008). Crowd simulation for emergency response using BDI agents based on immersive virtual reality. *Simulation Modelling Practice and Theory*, *16*(9), 1415–1429. https://doi.org/10.1016/j.simpat.2008.07.004

Sherman, W. R., & Craig, A. B. (2003). Understanding Virtual Reality: Interface, Application, and Design. *Presence*, *12*, 441–442.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, *8*(426), 1–16. https://doi.org/10.3389/fnbeh.2014.00426

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE*, *1*(1), e39. https://doi.org/10.1371/journal.pone.0000039

Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., Watson, Z., Breen, G., Schwarz, R., Steptoe, W., Szostak, D., Halan, S., Fox, D., & Silver, J. (2020). The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality*, *1*(March), 1–13. https://doi.org/10.3389/frvir.2020.00001

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing Our Lives with Immersive Virtual Reality. *Frontiers in Robotics and AI*, *3*(December), 1–47. https://doi.org/10.3389/frobt.2016.00074

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior.* Harvard University Press.

Spagnolli, A., Masotina, M., Furlan, M., Pluchino, P., Martinelli, M., & Gamberini, L. (2021). Sharing the Space With the "Victim" Can Increase Help Rates. A Study With Virtual Reality. *Frontiers in Psychology*, *12*(September). https://doi.org/10.3389/fpsyg.2021.729077

Sütfeld, L. R., Ehinger, B. V., König, P., & Pipa, G. (2019). How does the method change what we measure ? Comparing virtual reality and text-based surveys for the assessment of moral decisions in traffic dilemmas. *PLoS ONE*, *14*(10), e0223108.

Sütfeld, L. R., Gast, R., König, P., & Pipa, G. (2017). Using Virtual Reality to Assess Ethical Decisions in Road Traffic Scenarios: Applicability of Value-of-Life-Based Models and Influences of Time Pressure. *Frontiers in Behavioral Neuroscience*, *11*(122), 1–13. https://doi.org/10.3389/fnbeh.2018.00128

Sütfeld, L. R., König, P., & Pipa, G. (2019). Towards a Framework for Ethical Decision Making in Automated Vehicles. In *PsyArXiv*.

https://doi.org/doi.org/10.31234/osf.io/4duca

Tan, H. C., Ho, J. A., Teoh, G. C., & Ng, S. I. (2021). Is social desirability bias important for effective ethics research? A review of literature. *Asian Journal of Business Ethics*. https://doi.org/10.1007/s13520-021-00128-9

Tarnanas, I., & Manos, G. C. (2001). Using virtual reality to teach special populations how to cope in crisis: the case of a virtual earthquake. *Studies in Health Technology and Informatics*, *81*, 495—501.

Tassy, S., Oullier, O., Duclos, Y., Coulon, O., Mancini, J., Deruelle, C., Attarian, S., Felician, O., & Wicker, B. (2012). Disrupting the right prefrontal cortex alters moral judgement. *Social Cognitive and Affective Neuroscience*, *7*(3), 282–288. https://doi.org/10.1093/scan/nsr008

Thomson, J. J. (1985). The Trolley Problem. *Yale Law Journal*, *94*, 1395–1415. https://doi.org/10.2307/796133

Thomson, J. J. (1986). *Rights, restitution, and risk: Essays in moral theory*. Harvard University Press.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

Van Arem, B., Van Driel, C. J. G., & Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems*, *7*(4), 429–436. https://doi.org/10.1109/TITS.2006.884615

Vienna Convention on Road Traffic. (1968). https://treaties.un.org/Pages/ViewDetailsIII.aspx?src=TREATY&mtdsg_no=XI-B-19&chapter=11&Temp=mtdsg3&lang=en

Wilson, H., & Theodorou, A. (2019). Slam the brakes: Perceptions of moral decisions in driving dilemmas. *CEUR Workshop Proceedings*, *2419*.

Wintersberger, P., Prison, A. K., Riener, A., & Hasirlioglu, S. (2017). The experience of ethics: Evaluation of self harm risks in automated vehicles. *IEEE Intelligent Vehicles Symposium, Proceedings*, *2017-June*(Iv), 1–7. https://doi.org/10.1109/IVS.2017.7995749

Witmer, B. G., Jerome, C. J., & Singer, M. J. (2005). *The factor structure of the Presence Questionnaire*. *14*(3), 298–312.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240.

Yokoi, R., & Nakayachi, K. (2020). Trust in Autonomous Cars: Exploring the Role of Shared Moral Values, Reasoning, and Emotion in Safety-Critical Decisions. *Human Factors*. https://doi.org/10.1177/0018720820933041

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, *107*(15), 6753–6758.

Zackova, E., & Romportl, J. (2018). What might matter in autonomous cars adoption: first person versus third person scenarios. *ArXiv Preprint ArXiv:1810.07460.*