Virtual morality in the helping professions: simulated action and resilience

Article

Accepted Version

Virtual morality in the helping professions:

Simulated action and resilience

Francis, K. B. [1*], Gummerum, M. [2], Ganis, G. [2], Howard, I. S. [3], & Terbeck, S [2]

[1]University of Reading, Department of Philosophy, Reading, UK
[2]University of Plymouth, School of Psychology, Plymouth, UK
[3]University of Plymouth, Centre for Robotics and Neural Systems, Plymouth, UK

*Corresponding author: Kathryn B. Francis (University of Reading, UK)
(email: kathrynbfrancis@googlemail.com)

ABSTRACT

Recent advances in virtual technologies have allowed the investigation of simulated moral actions in aversive moral dilemmas. Previous studies have employed diverse populations in order to explore these actions, with little research considering the significance of occupation on moral decision-making. For the first time, in this study we have investigated simulated moral actions in Virtual Reality made by professionally trained paramedics and fire service incident commanders who are frequently faced with and must respond to moral dilemmas. We found that specially trained individuals showed distinct empathic and related personality trait scores and that these declined with years of experience working in the profession. Supporting the theory that these professionals develop resilience in moral conflict, reduced emotional arousal was observed during virtual simulations of a distressing dilemma. Furthermore, trained professionals demonstrated less regret following the execution of a moral action in virtual reality when compared to untrained control populations. We showed that, contrary to previous research, trained individuals made the same moral judgments and moral actions as untrained individuals, though showing less arousal and regret. In the face of increasing concerns regarding empathy decline in healthcare professionals, we suggest that the nature of this decline is complex and likely reflects the development of a necessary emotional resilience to distressing events.

Introduction

The moral ideologies of utilitarianism and deontology are often pitted against one another in "thought problems" examining the moral principles underlying moral decision-making (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). In the "trolley problem", two contrasting dilemmas generate a moral conundrum. In one, the *switch* dilemma, a runaway trolley threatens the lives of five people if it is not diverted on to a track where it will kill only one person (Foot, 1978). In another, the *footbridge* dilemma, the same trolley approaches five people on the tracks but this time, a bystander must be pushed from a footbridge in front of the trolley to stop it (Thomson, 1976). Factually, these dilemmas are identical in their trade-off of one life for many (Thomson, 1976). However, when asked to rate the moral acceptability of the proposed utilitarian solutions (i.e., killing one person to save five), individuals tend to approve the solution in the *switch* dilemma but refuse to do so in the *footbridge* variant (Greene et al., 2001; Thomson, 1976).

Perhaps the most widely discussed model, seeking to interpret these distinct responses, is Greene's dual process model of moral judgment (Greene et al., 2001). Greene argues that in "personal" dilemmas such as the *footbridge*, an immediate aversive response is triggered by emotional systems, resulting in refusals to endorse harm (a non-utilitarian response). In "impersonal" dilemmas such as the *switch* dilemma, on the other hand, the absence of this emotional response results in a controlled utilitarian calculation (Greene et al., 2001).

Beyond these traditional investigations of moral decision-making, research has begun to investigate how moral judgments might relay into actions, with evidence suggesting that they may be partially distinct (e.g., Tassy, Oullier, Mancini, & Wicker, 2013). Recently, advances in Virtual Reality (VR) technologies have allowed the investigation of simulated

moral actions in behavioural paradigms (e.g., Francis et al., 2016; Francis et al., 2017; Patil, Cogoni, Zangrando, Chittaro, & Silani, 2014). Although these novel approaches have revealed mixed findings (Navarrete, McDonald, Mott, & Asher, 2012; Skulmowski, Bunge, Kaspar, & Pipa, 2014), recent virtual paradigms have found an increased preference for utilitarian actions in VR, when compared to text-based judgments. This has been observed in impersonal (i.e., switch dilemma; Pan & Slater, 2011; Patil et al., 2014) and personal moral dilemmas (i.e., footbridge dilemma; Francis et al., 2016; Francis et al., 2017; McDonald, Defever, & Navarrete, 2017). Whilst these scenarios remain hypothetical in nature, they can offer insights into how immediate and emotive responses drive moral actions, as opposed to deliberated judgments about what is right and wrong (Francis et al., 2016; Rovira, Swapp, Spanlang, & Slater, 2009; Slater et al., 2006).

Occupation and Moral Decision-Making

Importantly, research investigating both moral judgments and moral actions has thus far, employed generalised populations in order to shed light on moral decision-making (e.g., Cushman, Young, & Greene, 2010; Francis et al., 2016; Francis et al., 2017; Greene, 2015; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Patil et al., 2014; Shenhav & Greene, 2014). Recently research has begun to investigate the significance of occupation on moral decision-making; more specifically, occupations that are directly connected to moral decision-making including those in healthcare settings and the military (e.g., Colangeli et al., 2015; Grinberg, Hristova, & Kadreva, 2016; Ransohoff, 2011).

This line of investigation is significant to the field of morality for several reasons. Firstly, specialised training in medical and emergency services involves frequent and direct exposure to emotionally salient and aversive situations (Grinberg et al., 2016). Secondly, recent research has suggested that "helping professions" adopt a "rescue personality"

(Wagner, Martin, & McFee, 2009) which is often associated with emotional resilience and coping strategies. For example, emergency service professionals have been found to have reduced aspects of empathy (e.g., Neumann et al., 2011; Williams et al., 2012) and down-regulation of responses to the pain of others (Decety, Yang, & Cheng, 2010) . Lastly, with VR offering immersive approaches to investigating realistic moral actions, emergency and healthcare services have begun to utilise virtual simulations in their training programmes (e.g., Colangeli et al., 2015). If moral judgments and moral actions in VR are at least partially distinct (e.g., Tassy et al., 2013), this is significant for occupations involved in regular moral decision-making, who begin to utilise VR in their training and assessment.

Morality and the Helping Professions

Grinberg et al. (2016) first investigated moral judgments in midwives and firefighters revealing a strong effect of occupation on moral judgments made in traditional text-based moral dilemmas; less utilitarian judgments were made by helping professions overall with fewest given by midwives. The authors initially note that in these professions, decision-making can impact human life and subsequently, training might lessen emotional arousal resulting in more utilitarian judgments. However, they later argue that their reverse finding suggests that "…these professions share high moral values… respect more strongly individual rights" (Grinberg et al., 2016, p. 709).

The idea that fewer utilitarian judgments can be observed in "helping professions" has been further supported in research examining moral judgments made by medical doctors and public health professionals in text-based vignettes (Ransohoff, 2011). Medical doctors are professionally obliged to protect the rights of their patients while public health professionals are more likely to be concerned with the bigger picture of public health (Greene, 2014). These moral ideals were supported in the finding that public health professionals gave more

utilitarian judgments than both medical doctors and controls suggesting that they give "…priority to the greater good" (Greene, 2014, p. 130).

In line with this research, Colangeli et al. (2015) utilised traditional text-based moral dilemmas and found that military pilots endorsed fewer utilitarian solutions compared to non-pilots when asked to judge the moral acceptability of the proposed utilitarian solution. When asked whether they would do the proposed utilitarian action, no difference was found between pilots and non-pilots. The authors argued that pilots' ratings of moral acceptability may result from their previous experiences leading to increased awareness and sensitivity when faced with emotionally aversive dilemmas (Colangeli et al., 2015).

Despite the recent focus on investigating the significance of occupation on moral judgment and separate lines of investigation examining the discrepancy between judgment and action, research has yet to investigate the impact of specialised training on simulated moral actions and moral inconsistency (the disparity between moral judgment and action). This is particularly relevant given the increased incorporation of VR technologies in healthcare and military training (Colangeli et al., 2015; Francis et al., 2016) and the need to assess real actions "…in strong time pressure conditions" (Colangeli et al., 2015, p. 76). As such, in the present research, we investigated both moral judgments and moral actions made by non-specialised controls and helping professions that are both experienced and in training. In the following section, we present specific characteristics of trained professionals that may alter moral judgment and moral action, followed by specific hypotheses for the present investigation.

Specialised Training and Moral Inconsistency

Arguably, the professional code of conduct under which helping professions operate can bias moral judgment outside of professional working environments (Grinberg et al.,

2016); personal moral orientations may change through training and experience to fall in line with professional codes of conduct. For example, strong moral convictions in healthcare professionals have been found to become less prominent in experienced individuals, perhaps reflecting this process (Grinberg et al., 2016).

Evidence that helping professions demonstrate resilience to aversive events comes from a multidisciplinary body of research. For example, in extensive interviews with firefighters, Alexander (2016, p. 162) noted that "…they instinctively understood the danger of upward counterfactual thinking" or imagining alternative circumstances that may have been better than reality (e.g., Roese, 1997). Firefighters also produced fewer expressions of regret, taking strength from "…professional pride, sense of duty" (Alexander, 2016, p. 212). This association between occupations involving moral decision-making and resilience, has also been found in research with physicians; traits associated with resilience support functioning in professions that are both "…demanding and stressful" (Eley et al., 2013, p. 7).

Exposure, Empathy, and Arousal

Investigations seeking to understand the origins of resilience in helping professions have highlighted the role of empathy arguing that lower aspects of empathy may "…serve as an adaptive function" providing resilience in emotionally aversive emergencies (Williams et al., 2012, p. 9). In recent research, empathy, assessed using generalised and specialised measures (e.g., Bellini, Baime, & Shea, 2002; Bellini & Shea, 2005), was found to decline in medical students and residents (Hojat et al., 2004; Neumann et al., 2011). However, this does not implicate all aspects of empathy as, arguably, specialist training allows professionals to empathise with a patient's circumstances "…without becoming emotionally entangled" (Williams et al., 2012, p. 9). Further, this decline has not been observed across all studies, with some research showing that empathy, as assessed using specialised measures, remains

stable throughout healthcare training across professions (e.g., Thomas et al., 2007; Williams et al., 2014).

Given that "empathy" encompasses a variety of components and is frequently used to refer to "…a heterogeneous collection of related phenomena" (Decety & Cowell, 2015, p. 2), here we investigate aspects of empathy using the Perspective Taking, Empathic Concern, and Personal Distress facets of the Interpersonal Reactivity Index (Davis, 1983). Although seen as a distinct conceptualisation to empathy (e.g., Batson, 2009; Decety & Moriguchi, 2007), Personal Distress is included as a measure of "…self-oriented, egoistic" reactions (Decety & Moriguchi, 2007, p. 17) with previous research indicating that it results in fewer helping behaviours compared to empathy (Carrera et al., 2013).

Specific Hypotheses

If helping professions are frequently exposed to aversive situations, we might expect moral judgments and moral actions to be more utilitarian in nature when compared to non-specialised populations as a result of successful emotion management (e.g., Grinberg et al., 2016). Alternatively, based on previous research and if exposure to single-life saving incidences is prevalent, these individuals could prioritise individual rights, resulting in non-utilitarian moral judgments and non-utilitarian moral actions (e.g., Greene, 2014; Grinberg et al., 2016). However, if the hypothetical nature of these trolley-type problems is extensive, it might be that decisions made in response to them, bear little resemblance to the real-life decisions made by professionals (Grinberg et al., 2016) resulting in no difference in moral responses between untrained and trained populations.

In terms of moral inconsistency, specialised training may alter personal moral orientations with both moral judgments and moral actions deriving from the same underlying moral principles or codes of conduct prescribed by the profession. This would subsequently

result in less moral inconsistency in trained professionals. Alternatively, if moral judgment and action are driven by distinct mechanisms and training does not alter perspectives, moral inconsistency may remain universal and unaffected by training.

In terms of resilience, if trained professionals do demonstrate forms of resilience, then post-hoc evaluations of simulated moral actions may differ from normal populations in that they reflect this resilience rather than regret.

Further, if "empathy" is lowered in trained professionals, then empathy scores and associated traits are likely to be significantly lower in helping professions when compared to non-specialised individuals and this should be related to years of experience. On the hand, given the inconsistency in these findings in previous work (e.g., Williams et al., 2014), empathy may remain stable across groups. However, given that empathy can encompass a variety of features, it is also important to formulate hypotheses at the component-level. As such, following evidence of empathy decline as reported in previous research (e.g., Neumann et al., 2011), we might expect to see a decline in Empathic Concern specifically (e.g., Bellini & Shea, 2005; Stratton, Saunders, & Elam, 2008) in helping professions. Alternatively, both affective (Empathic Concern) and cognitive empathy (Perspective Taking) may remain unaffected (e.g., Thomas et al., 2007; Williams et al., 2014) and we may instead see a decline in Personal Distress, which often hinders pro-social behaviours (Carrera et al., 2013).

With trained professionals demonstrating emotional resilience (e.g., Neumann et al., 2011; Williams et al., 2012) and an ability to down-regulate responses to the pain of others (Decety et al., 2010), repeated exposure to emotionally aversive events is also likely to affect arousal during moral decision-making. In previous VR paradigms, research has found increased arousal in virtual moral dilemmas as compared to text-based counterparts (Francis et al., 2016; Patil et al., 2014) indicating the value of virtual paradigms in triggering realistic

physiological responses to aversive situations (Rovira et al., 2009). Trained professionals, having had repeated exposure to aversive incidents, may prove emotionally resilient in virtual simulations showing little or no arousal increase and this should be related to years of experience in the profession.

Given the findings of previous research exploring moral actions in VR (Francis et al., 2016; Francis et al., 2017), we also assess both pro- and anti-social personality traits that have previously been associated with moral action and empathy (e.g., Francis et al., 2016; Tassy, Deruelle, Mancini, Leistedt, & Wicker, 2013).

Method

Participants

Sixty participants were recruited in total from three populations. Participants were sampled in order to gain a non-specialised group (control), a specialised group in training (paramedic practitioners) and an experienced specialised group (incident commanders from the fire service)

Twenty control participants comprising 17 females and three males, (M_{age} = 19.40, SD = 1.67 years, age range: 18 – 25 years) were recruited from a university participant pool and completed the study for course credit. Twenty paramedic practitioners comprising 12 females and eight males (M_{age} = 24.05, SD = 7.67 years, age range: 18 – 50 years) were recruited from a school of health professions and completed the study on a voluntary basis. Paramedic practitioners were currently completing the first year of a BSc Paramedic Practitioner course and were predominantly in training at a university. Twenty incident commanders comprising two females and 18 males (M_{age} = 39.85, SD = 9.37 years, age range: 22 – 61 years) were recruited from the Devon and Somerset Fire and Rescue Service

and completed the study on a voluntary basis at the Fire Training Centre at Exeter Airport.

Incident commanders were based at different stations across the counties of Cornwall, Devon

and Somerset. Although having various roles within the fire service rank structure, all were

trained in incident command.[1]

Measures

Participants were asked to fill out an electronic questionnaire comprising four self-

report questionnaires.

The *Interpersonal Reactivity Index (IRI)* (Davis, 1983) is an inventory designed to

measure dispositional empathy. It contains four subscales to measure Perspective Taking (7

items; α = .80), Empathic Concern (7 items; α = .73), and Personal Distress (7 items; α = .79).

The Fantasy Seeking subscale was not included in the present research given recent evidence

suggesting that it does not represent modern conceptualisations of empathy (Baldner &

McGinley, 2014; Batson, 2009). The subscales included here comprised 21 items with

responses given on a 5-point Likert scale (from A = *Does not describe me well* to E =

Describes me very well). The scale contains items such as "*I sometimes find it difficult to see*

things from the other guy's point of view".

The *Levenson Psychopathy Scale (LSRP)* (Levenson, Kiehl, & Fitzpatrick, 1995) is a

self-report measure of Psychopathy intended for research purposes. It has a two-factor

structure assessing both primary (i.e., selfishness) (16 items; α = .78) and secondary

psychopathic traits (i.e., impulsivity) (10 items; α = .55) in non-institutionalised populations.

The scale contains 26 items total, rated on a 4-point Likert scale (from 1 = *strongly disagree*

---

[1] Gender effects were not examined here as they were neither of interest a priori nor did the gender composition
of each group allow such a comparative analysis.

to 4 = *strongly agree*). The scale includes items such as "*Looking out for myself is my top priority*".

The *Hexaco-IP-PR* (Ashton & Lee, 2009) is a personality inventory designed to assess six dimensions of personality. In the present research, the subscales of Honesty-Humility (Items 10; α = .66), Emotionality (Items 10; α = .88), and Agreeableness (Items 10; α = .79) were included given the association between Honesty-Humility and Psychopathy (e.g., Djeriouat & Tremoliere, 2014), and the relevance of Emotionality and Agreeableness to the current investigation; Emotionality is linked to self-oriented aspects of empathy and Agreeableness is associated with being cooperative (Ashton & Lee, 2007, 2009). The subscales combined comprise 30 items with responses given on a 5-point Likert scale (from 1 = *strongly disagree* to 5 = *strongly agree*). The inventory contains items such as "*I wouldn't use flattery to get a raise or a promotion at work, even if I thought I would succeed*".

The *Self-Importance of Moral Identity Scale* (Aquino & Reed, 2002) provides a measure of moral identity. It contains two subscales that assess symbolization (i.e., public dimension of moral identity) (5 items; α = .64) and internalization (i.e., private dimension of moral identity) (5 items; α = .80). The inventory contains 10 items with responses given on a 5-point Likert scale (from A = *Does not describe me well* to E = *Describes me very well*). The scale contains items such as "*It would make me feel good to be a person who has these characteristics*". This measure was included as individuals with psychopathic traits often demonstrate a reduced sense of moral identity (Glenn, Koleva, Iyer, Graham, & Ditto, 2010).

Moral Decision-Making Measures

In all groups, participants completed a non-moral and moral action and a non-moral and moral judgment task. [2] In the action tasks, virtual scenarios were presented via a head-mounted display (Oculus Rift 1) (see Figure 1). In both the non-moral and moral virtual tasks, verbal instructions informed participants that the task involved the use of a joystick and that they would be given a chance to interact with a virtual object. Participants first completed a non-moral virtual task that required them to interact with a virtual object using a joystick, pushing it after hearing a tone. This task allowed a baseline measure of heart rate change to be collected (see 'physiological measures'). It also allowed us to determine whether increased arousal in the subsequent moral virtual task was attributable to the modality of VR or the moral content of the scenario itself. In the moral virtual task, an audio-visual virtual version of the *footbridge* dilemma was presented to participants. The dilemma played out as described by Foot (1978) but in real-time with the participant standing on a bridge behind a large virtual character. In the scenario, the participant viewed the scene in first person view. During the scenario, a trolley car (modern train railcar) approached from behind and travelled towards five virtual humans standing on the tracks in front of the participant. Participants had to decide whether they wanted to push the large person off the bridge to stop the trolley car's progress or to allow the trolley car to continue and kill the five people standing on the tracks. Audio instructions were played within the virtual environment:

- 30 seconds: "*Look behind you, a train is coming.*"

- 55 seconds: "*Hey I am too far away but if you want to save the people you could push the large person on to the tracks and derail the train. If you're going to push them, do it now, but it is your choice.*"

---

[2] Note that the labels "moral action" and "moral judgment" encompass both utilitarian and non-utilitarian decisions given that both are guided by moral principles.

As in the original study (see Francis et al., 2016), participants were then given a maximum of ten seconds to respond in the dilemma by either choosing to push the person off the bridge, using the joystick or by choosing to do nothing.

Figure 1. Incident commander wearing the Oculus Rift in the setup adopted at the Fire Training Centre at Exeter Airport. Photograph taken by the author: June, 2016.

For the judgment task, participants were first given a non-moral sample vignette to read which comprised a set of instructions displayed in the format of the pending dilemmas. Given that participants completed both the moral action and moral judgment tasks, we could not present the *footbridge* dilemma in the moral judgment task to prevent carryover effects (e.g., Bartels, Bauman, Cushman, Pizarro, & McGraw, 2015). As such, in the moral judgment task, participants were given a vignette describing a validated comparable dilemma (see Appendix S1) as opposed to the *footbridge* dilemma. This was embedded in five additional distractor dilemmas (Greene et al., 2001). These dilemmas were classified as personal and impersonal and were selected from those originally used in Greene, Morelli, Lowenburg, Nystrom, & Cohen (2008). All dilemmas were presented electronically in a random order.

After each dilemma, participants were asked a judgment question ("*Is it morally acceptable to [specific to the scenario]*?"). After a response was given, a second action-choice question was displayed asking ("*Would you do it?*"). Participants were given ten seconds to respond to each question, matching the virtual moral action task. Participants responded by selecting "Yes" (Y) or "No" (N).

Physiological Measures

Heart rate was recorded and heart rate change was measured using the equipment and procedures adopted in Francis et al. (2016). A Cateye-PL-6000 heart rate monitor was attached to participants via an ear lobe clip prior to completion of the electronic questionnaire to ensure that the device was working correctly. As outlined in previous research, heart rate change (bpm) can be both gradual and abrupt (Francis et al., 2016) and as such, heart rate readings were taken at the onset and offset of all non-moral and moral tasks. The time duration between these onset and offset task markers was dependent on the task type (non-moral; moral) and was determined by reading speed in the judgment tasks specifically (for a full description of this sampling procedure see Francis et al., 2016).

Heart rate sampling was completed with the primary aims of assessing whether arousal was modality or moral specific and in this research specifically, to determine whether trained professionals demonstrate emotional resilience to arousing moral scenarios. Although previous research has investigated arousal as a predictor of moral actions in virtual dilemmas (Navarrete et al., 2012; Patil et al., 2014), in the present virtual moral action task, heart rate change also incorporates the time in which participants witness the consequences of their actions (observing the death of virtual characters) and as such, records arousal beyond decision-making (Francis et al., 2016). As such, we do not investigate predictions of moral actions from arousal in the present study.

Procedure

All participants first completed the electronic questionnaire comprising the self-report trait questionnaires. All groups completed the action and judgment tasks with the order of completion counterbalanced. In the judgment tasks, both the non-moral and moral vignettes were presented on a computer running E-Prime software. In the action tasks, all participants first completed a questionnaire assessing their previous game-related and VR experience (weekly hours of video game playing, number of times playing games annually, and VR experience (from 1 = *no exposure* to 7 = *extensive exposure*)). In addition to this, all individuals were asked to record their direct experience working with any sector of the emergency services (in years). Both the non-moral and moral virtual action tasks were presented to participants via the Oculus Rift head-mounted VR system, generating a fully immersive and motion-tracked, 3-D environment. The headset uses a 7-inch display with colour depth of 24 bits per pixel creating a virtual environment with a 110° diagonal field of view and resolution of 640 x 800 pixels per eye. Crucially, a head-tracker running at 250 Hz allows head orientation tracking. During the virtual tasks, participants interacted with the scenario using a joystick and wore a pair of Sennheiser headphones to promote full immersion.

Having completed both the judgment and action tasks, a post-questionnaire was given to participants targeting the virtual moral action task specifically. Individuals were also asked "Did you take the right action?" and could respond with "Yes" or "No". This additional post-question was included to allow assessment of post-hoc expressions of regret and resilience. Participants were also asked "Did you feel coerced into carrying out these actions in Virtual Reality?" and could response with "Yes" or "No".

Results

Order Effects

Across groups, no order effects on moral responses were found based on task presentation when referencing the judgment question ($p >.490$) or action-choice question ($p >.598$).

Pre-Questionnaire Responses

For the moral action task, endorsing a utilitarian outcome (pushing the person in VR) was not associated with prior gaming experience ($ps >.165$) or VR experience ($ps >.262$) across groups.

Moral Responses

In order to compare moral judgments and moral actions, simulated moral actions in the virtual version of the footbridge dilemma were compared to the moral judgments made in response to the text-based counterpart. In all groups, the proportion of utilitarian responses was higher when simulated action was required in VR compared to when judgment was required in the text-based counterpart (see Table 1).

Table 1.

Percentage of Utilitarian Responses in Moral Judgment and Action Tasks

Group	Moral judgment task		Moral action task
	Judgment question	Action-choice question	
Control	10%	5%	60%
Paramedic	30%	30%	70%
Fire service	25%	25%	60%

Given that responses to the moral judgment task and moral action task were binary, Generalised Estimating Equations (GEE) were performed using a binary logistic model with

task (judgment task; action task) as within-subjects factor and group (control; paramedics; fire service) as between-subjects factor (see Figure 2). Two analyses were carried out, the first using the judgment question in the judgment task and the second using the action-choice question in the judgment task[3].

Judgment question. When selecting the judgment question, analysis revealed a main effect of task, (Wald $X^2[1] = 30.50$, $p < .001$), with a greater proportion of utilitarian responses overall in action tasks compared to judgment tasks. There was no main effect of group ($p = .373$) and no interaction ($p = .473$).

Action-choice question. When selecting the action-choice question, analysis revealed a main effect of task, (Wald $X^2[1] = 24.91$, $p < .001$), with a greater proportion of utilitarian responses overall in action tasks compared to judgment tasks. There was no main effect of group ($p = .217$) and no interaction ($p = .288$).

---

[3] This analysis was designed to compare moral actions in VR and moral judgments in text-based vignettes, hence the judgment and action-choice questions were referenced in separate analyses (as both derived from the same text-based moral dilemma). In order to determine if there were differences in responses to questions, an additional GEE analysis incorporating both the judgment and action-choice questions, revealed a main effect of task (Wald $X^2[1] = 33.36$, $p < .001$), with a greater proportion of utilitarian responses overall in the action task as compared to both questions in the judgment task ($ps < .001$) but no difference between moral responses to the judgment question and action-choice question ($p = 1.000$).
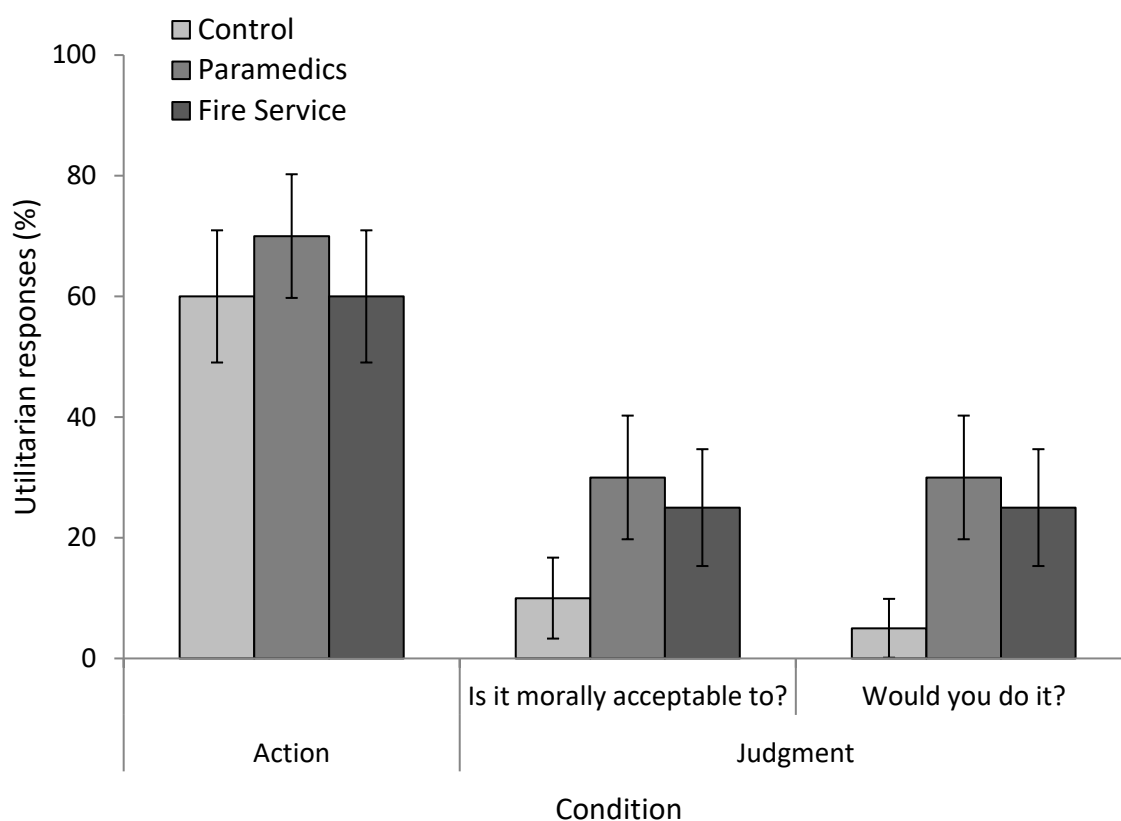
Figure 2. Utilitarian responses (%) in the moral action task (VR *footbridge*) and the moral judgment task (text-based *footbridge* counterpart) in the control, paramedic, and fire service groups. In the judgment task, participants were asked whether the action was morally acceptable and whether they would do it. A greater number of utilitarian outcomes were endorsed in the moral action task. Error bars represent +- 1 SE$_p$.

Heart Rate Analyses

In all groups, changes in heart rate were calculated as the difference in heart rate readings (bpm) between the onset and offset task markers. These heart rate changes were calculated for both judgment and actions tasks and for both moral and non-moral task types resulting in heart rate change readings for non-moral (judgment; action) and moral (judgment; action) tasks. Mean heart rate change increased in the moral action task for the control group and paramedic practitioners but decreased for incident command officers in the fire service (see Table 2).

Table 2

Mean Heart Rate Change across Judgment and Action Tasks

	Judgment task		Action task	
	Non-moral	Moral	Non-moral	Moral
Group	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Control	-0.45 (1.15)	0.80 (2.50)	-0.95 (2.21)	3.10 (3.21)
Paramedics	-0.50 (0.67)	-0.05 (1.43)	-0.25 (1.74)	1.75 (3.52)
Fire service	-0.05 (1.15)	-0.10 (1.68)	0.05 (1.54)	-0.20 (3.40)

We conducted a mixed ANOVA with task (judgment task; action task) and type (non-moral task; moral task) as within-subjects factors and group (control; paramedics; fire service) as the between-subjects factor. Analysis revealed a main effect of task, ($F(1, 57) = 5.30$, $p = .025$, $\eta_p^2 = .09$), a main effect of type, ($F(1, 57) = 14.99$, $p < .001$, $\eta_p^2 = .21$), a significant interaction of type x task, ($F(2, 57) = 7.17$, $p = .010$, $\eta_p^2 = .11$), and a significant interaction of type x group, ($F(2, 57) = 6.35$, $p = .003$, $\eta_p^2 = .18$) (see Figure 3). There was no main effect of group ($p = .136$). The three-way interaction of type x task x condition approached marginal significance, ($F(2. 57) = 2.83$, $p = .067$, $\eta_p^2 = .09$). In this instance, statistically significant two-way interactions were followed up.
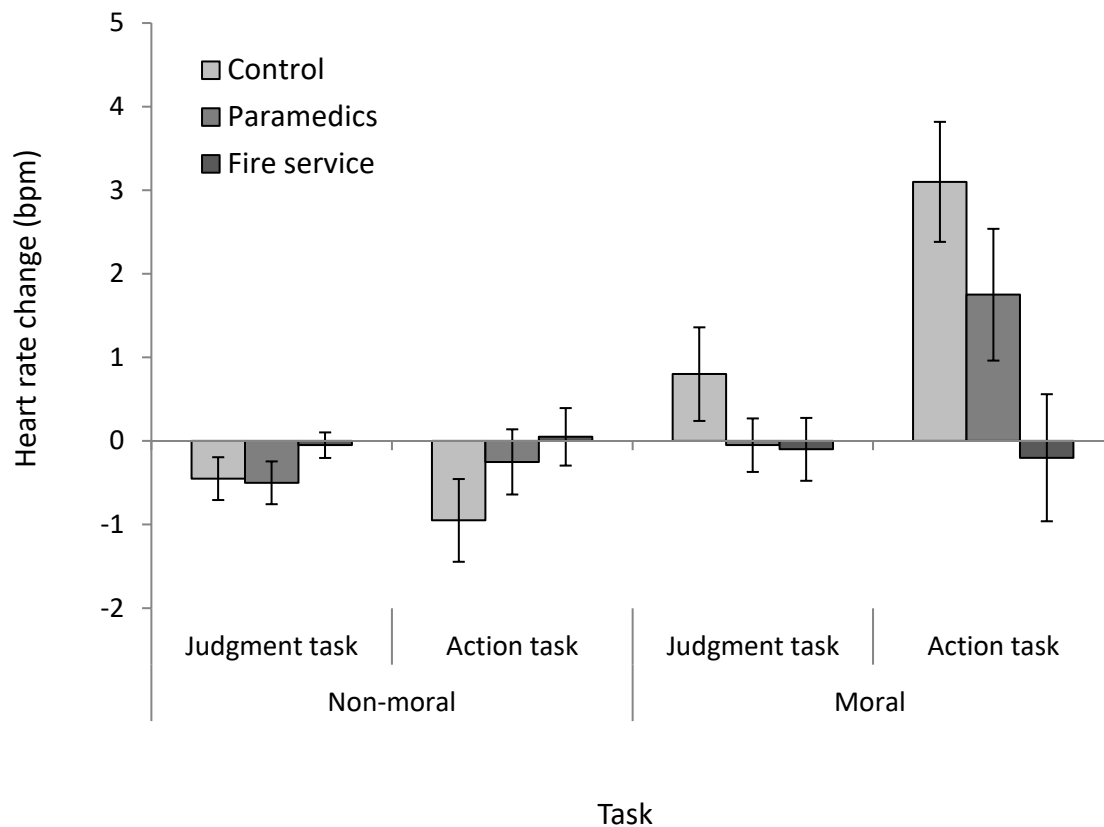
Figure 3. Mean heart rate change (bpm) for non-moral and moral task type in judgment and action tasks by group. Increased heart rate changes were observed in moral tasks for the control group and paramedics only. Error bars represent +- 1 SE

To further investigate the interaction of type x task, simple effects analyses were performed comparing heart rate changes in control and moral tasks within both judgment and action tasks. In the judgment task, heart rate change was marginally significantly higher in the moral task compared to the non-moral task across groups, $(F(1, 57) = 3.92, p = .052, \eta_p^2 = .06)$. In the action task, heart rate change was significantly higher in the moral task compared to the non-moral task, $(F(1, 57) = 14.26, p < .001, \eta_p^2 = .20)$. For the moral tasks only, there was a significant difference in heart rate change between the judgment and action task across groups, $(F(1, 57) = 7.86, p = .007, \eta_p^2 = .12)$ with heart rate change highest in the virtual moral action task.

To further investigate the interaction of type x group, simple effects analyses were performed comparing heart rate change in non-moral and moral tasks across groups. This analysis suggested that for the non-moral tasks, heart rate changes were not significantly different between the three groups, ($p = .159$). However for the moral tasks, analysis suggested that heart rate changes were different between groups, ($F(2, 57) = 5.28$, $p = .008$, $\eta_p^2 = .16$). Specifically, in the control group, heart rate changes were significantly greater in the moral tasks than in the non-moral tasks, ($F(1, 57) = 22.75$, $p = < .001$, $\eta_p^2 = .29$) and this was also the case for the paramedic practitioners, ($F(1, 57) = 4.86$, $p = .032$, $\eta_p^2 = .08$). Analysis suggested that there was no significant difference in heart rate changes between non-moral and moral tasks for the incident command officers in the fire service group, ($p = .788$).

Personality Traits

In order to assess any differences between personality traits across the groups, a one-way ANOVA was used to compare traits. Several significant differences between personality traits were found between groups and follow-up comparisons with Bonferroni corrections were performed to determine between which groups these differences were present (see Table 3).

Table 3

Personality Trait Differences

Measure	Subscale	Control M (*SD*)	Paramedics M (*SD*)	Fire service M (*SD*)
1. LPS				
	Primary	30.60 (*7.43*)	29.20 (*5.55*)	29.20 (*5.15*)
	Secondary	20.85 (*3.28*)	21.45 (*3.00*)	18.60 (*3.80*)
2. HEXACO				
	H	3.31 (*0.53*)	3.73 (*0.60*)	3.73 (*0.39*)
	Em	3.32 (*0.83*)	3.23 (*0.63*)	2.64 (*0.61*)
	A	3.23 (*0.71*)	3.14 (*0.69*)	3.16 (*0.56*)
	C	3.65 (*0.47*)	3.56 (*0.48*)	3.78 (*0.49*)
3. IRI				
	PT	17.90 (*3.84*)	17.20 (*4.96*)	16.75 (*4.64*)
	EC	19.30 (*4.74*)	19.30 (*3.25*)	17.70 (*4.29*)
	PD	11.15 (*3.73*)	8.35 (*4.37*)	5.45 (*3.07*)
4. Moral Identity				
	Symbolization	2.92 (*0.62*)	3.04 (*0.85*)	3.13 (*0.53*)
	Internalization	4.36 (*0.44*)	4.37 (*0.51*)	4.25 (*0.43*)

Note. H = Honesty-Humility, Em = Emotionality, Ex = Extraversion, A = Agreeableness, C = Conscientiousness, O = Openness to experience. PT = Perspective Taking, EC = Empathic Concern, PD = Personal Distress.
------ p =.05 ------ p <.05 —— p <.001

No significant differences between the control group, the fire service and the paramedics were found for Moral Identity in either Symbolization or Internalization (ps > .621), the HEXACO subscales Agreeableness and Conscientiousness (ps > .368), primary Psychopathy (p = .707), or the IRI subscales Perspective Taking and Empathic Concern (ps > .376).

Years of Experience

Heart rate. Given that trained professionals are frequently exposed to emotionally aversive situations, bivariate correlations were carried out to determine whether heart rate change in the virtual moral task and heart rate change in the text-based moral judgement task were associated with years of direct experience working with any sector of the emergency services (see Figure 4). Years of experience was not correlated with heart rate change in the judgment moral task ($p = .566$) or heart rate change in either the non-moral judgment or action tasks ($ps > .227$). Years of experience had a small negative correlation with heart rate change in the moral action task, ($r(58) = -.28$, $p = .033$) and when entered into a univariate linear regression, was found to explain 7.6% of the variance in the model, ($R^2 = .076$, $F(1,58) = 4.76$, $p = .033$) when predicting heart rate change ($\beta = -.28$, $p = .033$).
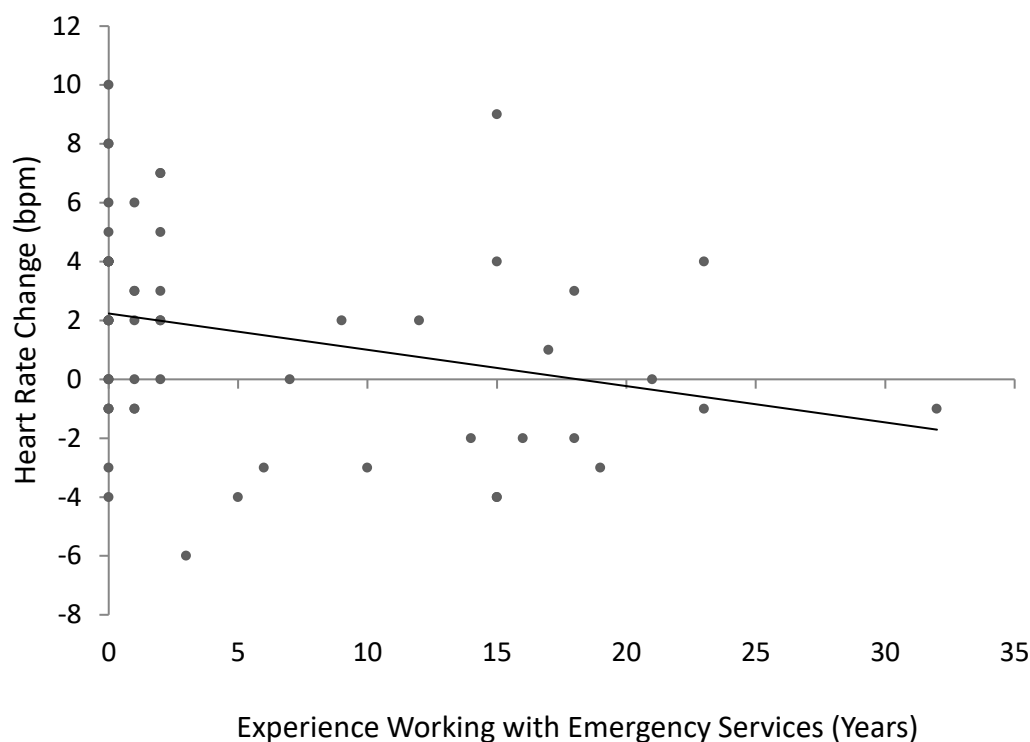

Figure 4. Heart rate change (bpm) in the virtual moral action task plotted against direct experience working with any sector of the emergency services (years). Years of experience negatively predicted heart rate change in the virtual moral task. Linear regression trendline: $R^2 = .08$.

Personality traits. Given that previous research has found empathy decline in trained professionals (Neumann et al., 2011) arguably as an adaptive function (Williams et al., 2012), personality traits were also correlated with years of direct experience working with any sector of the emergency services. Years of experience were negatively correlated with secondary Psychopathy ($r(58) = -.44$, $p < .001$), Emotionality ($r(58) = -.32$, $p = .011$) and Personal Distress ($r(58) = -.40$, $p = .001$) (see Figure 5). A multivariate regression with these traits entered as dependent variables and years of experience entered as the predictor variable, found that experience negatively predicted all traits, $\Lambda_{pillai} = .320$, $F(3,56) = 8.77$, $p < .001$ (see Table 4). [4]

---

[4] Given theories regarding the relationship between empathic components (and related traits) and emotional resilience in experienced professionals working in the emergency sector, we also performed mediation analyses. These revealed that heart rate change in the virtual moral task did not mediate the relationship between years of experience and the above traits ($ps > .085$).
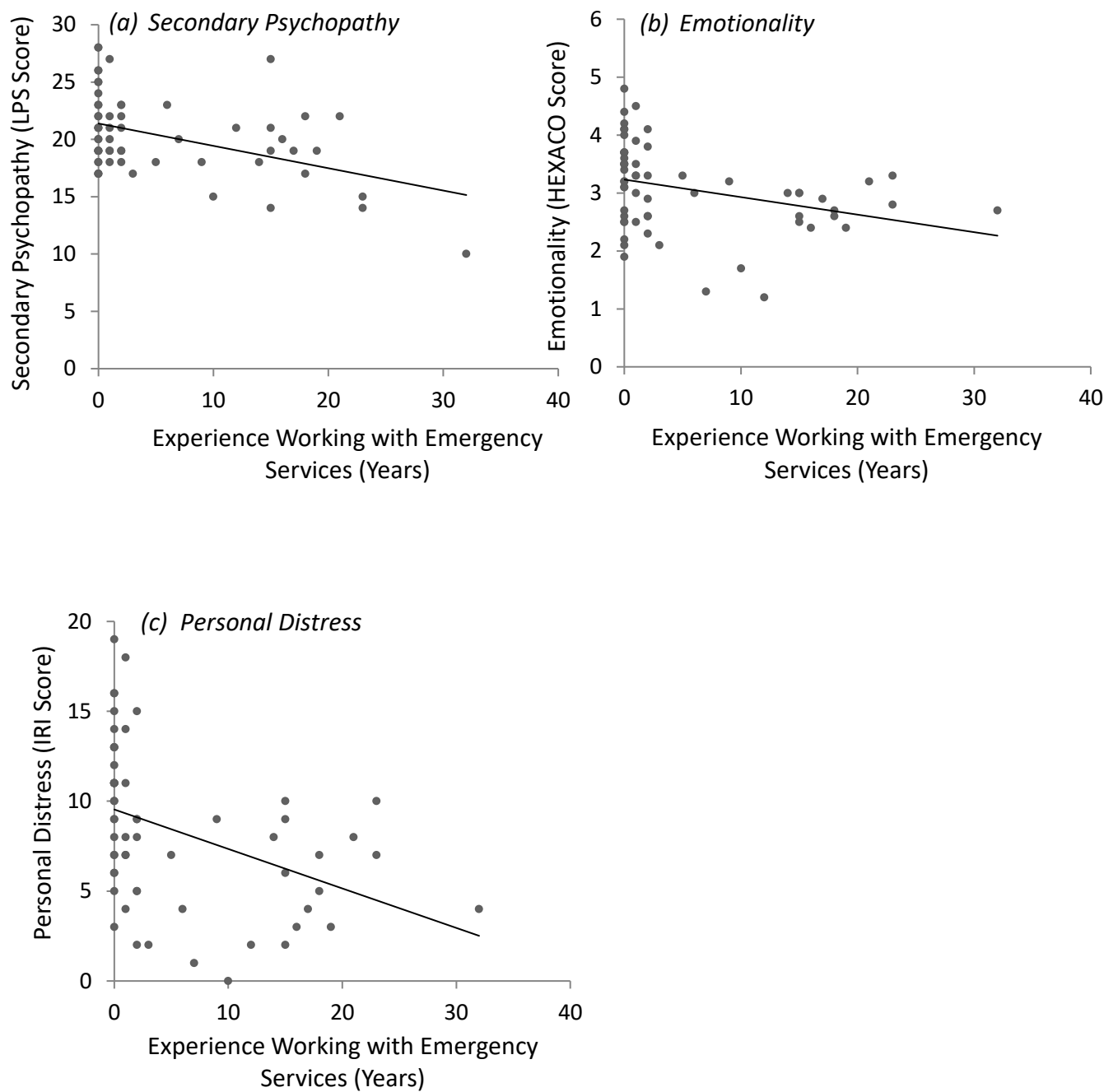
Figure 5. Personality traits including *(a)* Secondary Psychopathy *(b)* Emotionality *(c)* Personal Distress plotted against direct experience working with any sector of the emergency services (years). Linear regression trendlines for *(a)* $R^2 = .20$ *(b)* $R^2 = .11$ *(c)* $R^2 = .24$ *(d)* $R^2 = .16$. LPS = Levenson Psychopathy Scale; IRI = Interpersonal Reactivity Index.

Table 4

Univariate Outputs of Multivariate Regression with Year of Experience as Predictor of Traits

Dependent Variable	Included	B (*SE*)	β	R^2
Secondary Psychopathy				
	Constant	21.39 (*0.51*)		
	Experience	-0.20 (*0.05*)	-.44**	.20
Emotionality				
	Constant	3.23 (*0.11*)		
	Experience	-0.03(*0.01*)	-.32*	.11
PD				
	Constant	9.54 (*0.64*)		
	Experience	-0.22(*0.07*)	-.40*	.16

Note: Experience = years of direct experience working with any sector of the emergency services. PD = Personal Distress. *p <.05. **p <.001. (*SE)* = standard error.

Post-Action Judgments

In order to assess post-hoc evaluations of simulated utilitarian actions in VR (or omission of simulated utilitarian actions), responses to the question "Did you take the right action?" were compared across groups. This was done with the intention of examining resilience versus regret in post-hoc evaluations of moral actions. A three-way chi-square test was performed to examine the relation between group (control; paramedics; fire service) and post response (right action: yes; no) based on response type in VR (endorsed; refused). Fisher's Exact Tests (FET) are reported for instances in which cells have expected counts less than five. There was a significant association between group and whether or not participants reported that they had taken the right action. This was only the case when a utilitarian action had been endorsed in VR, ($\chi^2(2) = 14.79, p < .001$; FET).

In order to interpret this finding, follow-up tests using two-way chi square tests were performed with Bonferroni corrections. Of the participants in the control group who pushed in VR, only 33.33% stated that they had taken the right action. Of the paramedic practitioners who pushed in VR, 100% stated that this was the right action, significantly more than in the control group, ($p < .001$; FET). Similarly, of the incident commanders who pushed in VR, 83.33% stated that this was the right action, again, significantly more than in the control group, ($\chi^2(1) = 6.17$, $p = .013$) (see Figure 6). There was no significant difference in post responses between paramedics and incident commanders ($p = .203$; FET).
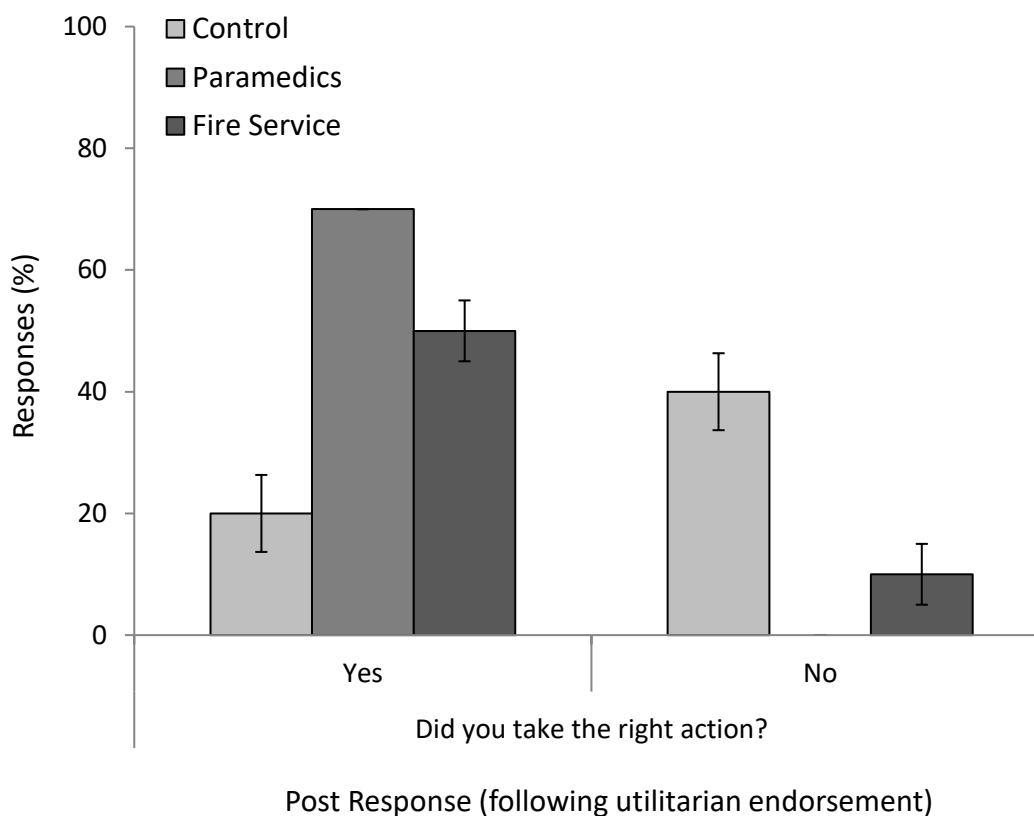

Figure 6. Post responses (%) to the question "Did you take the right action?" following the endorsement of a utilitarian action (pushing the man) in VR by group. A greater number of "Yes" responses were given in both the paramedic and fire service groups. Error bars represent +- 1 SE_p.

Coercion

The majority of the control group (70%), fire service (70%), and paramedic practitioners (80%) reported that they did not feel coerced in VR. In order to see if coercion predicted action responses in the virtual footbridge dilemma, univariate logistic regressions were performed with group (control; paramedics; fire service) as the selection variable, coerced (yes; no) as the categorical predictor and VR action response as the categorical outcome (non-utilitarian action; utilitarian action). Coercion did not predict VR action responses in the control group ($p = .187$), in the paramedic group ($p = .999$) or in the fire service group ($p = .691$).

Discussion

Overall, greater utilitarian endorsement was observed when simulated action was required in VR and this was the case across untrained individuals, specialist individuals in-training, and specialist trained individuals. While moral judgments and moral actions were not significantly different between controls and individuals working in helping professions, there were differences between these populations with regards to resilience, in terms of arousal in virtual simulations, and in post-hoc judgments of actions.

Moral Actions versus Judgments

In terms of moral judgment and moral action, participants endorsed the utilitarian response (pushing the person off the bridge) when action was required in VR significantly more than when judgment was required in the traditional text-based version of the footbridge dilemma across all groups, regardless of specialised training. Our results fall in line with previous VR research (Francis et al., 2016; Francis et al., 2017; McDonald et al., 2017; Pan & Slater, 2011; Patil et al., 2014) with VR moral dilemmas offering insights into the role of

immediate moral actions as opposed to judgments made in text-based moral dilemmas. In previous VR research, this preference for simulating utilitarian actions in VR has been interpreted through two accounts. In the contextual saliency account, it is theorised that the physical features in virtual scenarios enable participants to "see" victims, placing greater negative emphasis on witnessing victims die than on generating harmful actions (Francis et al., 2016; Francis et al., 2017; Patil et al., 2014). In frame of reference of accounts, it is theorised that actions are driven by egocentric perspectives that overlap with self-interested motives, as people consider the self-relevant consequences of their own actions (Francis et al., 2017; Tassy et al., 2013). Conversely, judgments rely on allocentric evaluations that are influenced by cultural norms (Tassy et al., 2013).

For the helping professions specifically, we had hypothesised that they may demonstrate reduced moral inconsistency as a result of professional codes of conduct influencing and driving both moral judgments and moral actions. However, the present findings suggest that moral inconsistency remains unaffected by professional training, supporting the theory that moral judgment and action are driven by distinct mechanisms (e.g., Tassy et al., 2013). Further, with trained professionals often exposed to aversive situations, we theorised that this might lead to greater utilitarian endorsements while previous research argued that exposure to single-life incidences would lead to fewer utilitarian endorsements (e.g., Grinberg et al., 2016). However, in the present study, moral actions and moral judgments were not significantly different between groups, contradicting previous research (Grinberg et al., 2016) that has argued that trolley-type dilemmas can have significance "…in the real world and not just in the lab" (Greene, 2014, p. 130). Our findings here do not directly support this claim and might suggest that the hypothetical nature of these trolley-type problems limit their application to real-life decision-making.

However, we propose an alternative interpretation of the present finding. Hypothetical moral dilemmas were not originally developed with the intention of revealing insights into real-life decisions but were designed, in their experimental simplicity, to allow moral conflicts to play out in artificial contexts with anonymous agents making them impervious to individual experience (Christensen & Gomila, 2012; Hauser, Cushman, Young, Kang‑Xing Jin, & Mikhail, 2007). Our finding that specialised training does not alter moral decisions in trolley-type problems, whether text-based or virtual, supports the experimental control available in these hypothetical scenarios. They remain unaffected by personal experiences, as was originally intended in their formulation, allowing moral psychologists to understand the "…foundational psychological processes that underlie human moral cognition" (Christensen & Gomila, 2012, p. 1250).

Previous research has acknowledged that endorsing a utilitarian or supposedly "utilitarian" moral decision can derive from distinct motivational forces including morally-driven intentions and intentions driven by personality traits (e.g., Bartels et al., 2015). Importantly, while moral judgments and moral actions were not significantly different across control participants and specialised professionals, underlying factors contributing to moral decision-making did differ between these groups.

Specialised Training and Personality Traits

With specialised professionals often employing a "rescue personality" (Wagner et al., 2009) associated with emotional resilience, we predicted that there would be differences in empathy and associated traits between untrained individuals and helping professionals. This was supported, with several personality trait differences found between untrained controls, paramedic practitioners, and incident commanders in the fire service. We outlined contrasting hypotheses predicting either an associated decline or consistency in these traits with years of

experience working in the profession. The former hypothesis was supported with negative associations found between several personality traits and years of experience working with the emergency services. Secondary psychopathy, a trait associated with impulsivity and quick-temperedness was negatively predicted by years of experience. In terms of HEXACO traits, Emotionality, a self-oriented trait associated with fearfulness and empathic sensitivity (Ashton & Lee, 2005), was negatively predicted by years of experience. Further and in line with this, Personal Distress, a trait "…negatively related to measures of social functioning" (Davis, 1983, p. 116) and associated with feeling anxiety when observing the negative experiences of others, also declined with years of experience working in the emergency services.

The finding that length of time working in helping professions results in a significant decline in traits associated with fearfulness and personal distress, might reflect the emotional resilience that is formed in response to repeated exposure to aversive, dangerous, and emotionally volatile situations (e.g., Eley et al., 2013; Grinberg et al., 2016) and suggests that helping professionals gain "…increasing confidence in facing emergencies" (Bellini & Shea, 2005, p. 167). Whether our findings support reports of empathy decline in healthcare professionals (e.g., Neumann et al., 2011), is less clear. While lowered Personal Distress and Emotionality scores were observed, both Empathic Concern and Perspective Taking scores in both paramedic practitioners and experienced incident command officers were stable. Given that the latter components fall within conceptualisations of "empathy" and are arguably "…two behaviours particularly important for medicine" (Bellini & Shea, 2005, p. 167), our findings suggest that caring for another's welfare and putting oneself into the mind of another, are not affected through healthcare and emergency training.

Importantly, with healthcare research disputing the adaptive function of lowered empathic responses versus "...the need to incorporate, promote and instil empathy" (Williams

et al., 2012, p. 10), we are reminded that in this complex issue "…empathic shades of gray are needed" as trained professionals learn to "…moderate the degree to which they harden their hearts" (Newton, 2013, p. 9). Future research should begin to distinguish empathic components when investigating the supposed "empathy decline" in healthcare and military settings.

Exposure and Arousal

Heart rate change was assessed to determine whether arousal was triggered by modality (VR versus text) or moral content and also as a means of examining emotional resilience in trained professionals. With both incident commanders and paramedics frequently working in critical emergency situations in which they have to make "…quick and potentially life-saving decisions…with limited medical back-up" (Williams et al., 2012, p. 9), we theorised that arousal might be lower in these populations and influenced by experience. In the present study, greater heart rate change was observed for moral tasks compared to control tasks for both action and judgment tasks suggesting that the VR modality alone was not responsible for increased arousal but moral content, supporting previous findings (Francis et al., 2016). However, this higher arousal in moral as compared to control tasks was only found for the untrained control group and paramedic practitioners in-training and not in the trained incident commanders from the fire service, suggesting that trained professionals may demonstrate emotional resilience to arousing situations.

Further, years of experience in the emergency services did negatively predict a proportion of variation in heart rate change in the virtual moral task. Whilst this provides some support for a resilience hypothesis, this association was small. Importantly, this relationship was not found for heart rate change in the text-based judgment task, providing

further evidence that utilising VR paradigms in the moral domain can offer valuable insights into realistic emotional responses (Parsons, 2015; Rovira et al., 2009).

Specialised Training and Regret

Previous research has suggested that helping professions also demonstrate resilience in decision-making, exhibiting reduced expressions of regret (Alexander, 2016) and resistance to imagining alternatives that may have been better than reality (e.g., Roese, 1997). We theorised that this resilience might surface in post-hoc judgments about virtual moral actions with trained professionals expressing less regret than untrained populations. In support of this, it was found that when utilitarian actions were endorsed in VR, untrained individuals in the control group, expressed feelings of regret reporting that they had taken the wrong action. Conversely, the majority of paramedic practitioners and incident commanders who had endorsed the utilitarian action in VR reported that they had taken the right action. This suggests that specialised training in the helping professions provides post-hoc resilience following utilitarian moral action. Future research might consider extending this investigation of resilience following moral action in trained professionals by incorporating counterfactual thinking and moral emotion measures.

Limitations and Implications

As with previous research, gaming and VR experience did not predict utilitarian responses in the virtual moral task for any group supporting previous arguments that responses in VR tasks are not akin to those of a gaming environment (Francis et al., 2016; Francis et al., 2017). As an alternative interpretation of the present findings, it might be argued that greater utilitarian actions in VR were elicited as a result of coercion induced by auditory instructions. To address this, in the present study, we asked participants whether they felt coerced into carrying out simulated actions in VR. Across the groups, the majority

of participants reported that they had not felt coerced. Further, coercion did not predict utilitarian moral actions. As such, it is unlikely that the greater utilitarian actions observed across the present VR studies were as a result of coercion.

It might be argued that the incorporation of a single hypothetical dilemma in the present study limits the reliability with which we can assert that helping professions produced the same moral actions and moral judgments as untrained controls. In previous studies that found a significant difference in healthcare professionals' moral judgments (e.g., Grinberg et al., 2016; Ransohoff, 2011), multiple moral dilemmas were incorporated. In the present experiment, additional text-based moral dilemmas were incorporated within the judgment task but as distracter dilemmas. However, to address the above consideration, we later analysed moral judgments across all text-based moral dilemmas comparing them between groups (see Appendix S2). Overall, greater utilitarian responses were observed for personal over impersonal dilemmas across all groups but no significant differences between groups in their judgments and no interactions between condition, dilemma type, or question were found. We were unable to carry out a similar analysis for moral actions given the incorporation of a single virtual moral simulation. However, this follow-up analysis does suggest that moral judgments remained the same across untrained individuals, specialists in-training, and trained specialists, when incorporating multiple moral dilemmas.

It is also important to note the potential limitations of our relative measure of moral judgment, that measures utilitarian and deontological inclinations on a single spectrum. It is possible that trained professionals have a heightened concern not only regarding harm avoidance but also regarding utility maximisation; this would result in a null effect when using the relative moral judgment measure as the two tendencies cancel each other out. As such, future work might consider utilising process dissociation approaches, rather than relative measures, to reveal if these professionals do in fact have elevated levels of both

utilitarian and deontological inclinations (Conway & Gawronski, 2013). Further and in line with these considerations regarding methodology, the types of dilemmas included in the present study involve scenarios that pit the welfare of two different sets of strangers, which are not the typical kinds of situations that these professionals might encounter in their duties. These individuals are more likely to encounter dilemmas that pit their own welfare against the welfare of other individuals (Patil et al., 2017; Rand & Epstein, 2014). As such, it is possible that training influences their responses to altruistic dilemmas as opposed to sacrificial moral dilemmas. There is recent work showing that judgment-behaviour discrepancy is also observed in such altruistic dilemmas (Patil et al., 2017) and so future work might explore if helping professionals exhibit the same discrepancy as controls in these altruistic contexts.

In addition, in the present study, while the differences between paramedic practitioners and incident command officers are investigated, we do assess years of experience in the helping professions broadly without distinguishing years of experience as a paramedic or years of experience in the fire service. This approach to studying helping professions has been criticised; "…assuming homogeneity among professions may be unwise" (Wagner et al., 2009, p. 7). For example, in the present study, the relationship between years of experience working in the helping professions and changes in arousal in the virtual moral action task were small. Arguably, decreased arousal responses may result from working in the fire service specifically and may not reflect the effect of working in helping professions more broadly. As such, future research should explore emergency service groups independently prior to identifying disparities and similarities between them (Wagner et al., 2009). For example, future research examining the effect of years of experience working in the emergency services on emotional resilience, might consider recruiting both experienced professionals and professionals in training from the same helping profession.

Lastly, existing research examining moral decision-making in specialist occupations has largely adopted the theory that exposure and experience may produce either utilitarian or deontological ideals (e.g., Grinberg et al., 2016; Ransohoff, 2011). However, the pattern of responses given by professionals in the present research and similarity of responses to non-specialist populations, supports the theory that individuals do not assume one moral label but instead adopt a "…particularist approach to morals that takes the details of each case into account" (Christensen et al., 2014, p. 16). Future research should consider this particularist approach when investigating moral decision-making in specialist occupations.

References

Alexander, L. E. (2016). *When the bells go down: Resilience and vulnerability in firefighters.* University of Hertfordshire. Retrieved from http://uhra.herts.ac.uk/handle/2299/17096?show=full

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*, 1423-1440.

Ashton, M. C., & Lee, K. (2005). Honesty-humility, the big five, and the five-factor model. *Journal of Personality, 73*, 1321-1353. doi: 10.1111/j.1467-6494.2005.00351.x

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*, 150-166. doi: 10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340-345. doi: 10.1080/00223890902935878

Baldner, C., & McGinley, J. J. (2014). Correlational and exploratory factor analyses (EFA) of commonly used empathy questionnaires: New insights. *Motivation and Emotion, 38*, 727-744.

Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2015). Moral judgment and decision making. In G. Keren & G. Wu (Eds.), *The wiley blackwell handbook of judgment and decision making*. Chichester, England: Wiley Online Library.

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomona. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3-15). Cambridge, MA: MIT Press.

Bellini, L. M., Baime, M., & Shea, J. A. (2002). Variation of mood and empathy during internship. *JAMA, 287*(23), 3143-3146.

Bellini, L. M., & Shea, J. A. (2005). Mood change and empathy decline persist during three years of internal medicine training. *Academic Medicine, 80*, 164-167.

Carrera, P., Oceja, L., Caballero, A., Muñoz, D., López-Pérez, B., & Ambrona, T. (2013). I feel so sorry! Tapping the joint influence of empathy and personal distress on helping behavior. *Motivation and Emotion, 37*, 335-345.

Christensen, J. F., Flexas, A., Calabrese, M., Gut, N. K., & Gomila, A. (2014). Moral judgment reloaded: a moral dilemma validation study. *Frontiers in Psychology, 5*(607), 1-18. doi: 10.3389/fpsyg.2014.00607

Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: a principled review. *Neuroscience and Biobehavioral Reviews, 36*, 1249-1264. doi: 10.1016/j.neubiorev.2012.02.008

Colangeli, S., Boccia, M., Piccardi, L., Verde, P., Cordellieri, P., Ferlazzo, F., & Giannini, A. M. (2015). Moral decision-making in military pilots versus military personnel. *Italian Journal of Aerospace Medicine, 13*, 72-84.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *Journal of Personality and Social Psychology*, *104*(2), 216-235.

Cushman, F., Young, L., & Greene, J. (2010). Our multi-system moral psychology: Towards a consensus view. In J. M. Doris (Ed.), *The moral psychology handbook* (pp. 47-71). Oxford, England: Oxford University Press.

Davis, M. H. (1983). Measuring individual-differences in empathy - Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113-126. doi: 10.1037//0022-3514.44.1.113

Decety, J., & Cowell, J. M. (2015). Empathy, justice, and moral behavior. *AJOB Neuroscience, 6*, 3-14. doi: 10.1080/21507740.2015.1047055

Decety, J., & Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: implications for intervention across different clinical conditions. *Biopsychosocial Medicine, 1*(22), 1-22. doi: 10.1186/1751-0759-1-22

Decety, J., Yang, C. Y., & Cheng, Y. (2010). Physicians down-regulate their pain empathy response: an event-related brain potential study. *Neuroimage, 50*, 1676-1682. doi: 10.1016/j.neuroimage.2010.01.025

Djeriouat, H., & Tremoliere, B. (2014). The Dark Triad of personality and utilitarian moral judgment: The mediating role of Honesty/Humility and Harm/Care. *Personality and Individual Differences, 67*, 11-16. doi: 10.1016/j.paid.2013.12.026

Eley, D. S., Cloninger, C. R., Walters, L., Laurence, C., Synnott, R., & Wilkinson, D. (2013). The relationship between resilience and personality traits in doctors: implications for enhancing well being. *PeerJ, 1*.

Foot, P. (1978). *Virtues and vices and other essays in moral philosophy*. Berkeley, CA: University of California Press.

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. doi: 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*. doi: 10.1038/s41598-017-13909-9

Glenn, A. L., Koleva, S., Iyer, R., Graham, J., & Ditto, P. H. (2010). Moral identity in psychopathy. *Judgment and Decision Making, 5*, 497-505.

Greene, J. D. (2014). *Moral tribes: emotion, reason and the gap between us and them*. London, England: Atlantic Books Ltd.

Greene, J. D. (2015). The rise of moral cognition. *Cognition, 135*, 39-42. doi: 10.1016/j.cognition.2014.11.018

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*, 1144-1154. doi: 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 bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389-400. doi: 10.1016/j.neuron.2004.09.027

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*, 2105-2108. doi: 10.1126/science.1062872

Grinberg, M., Hristova, E., & Kadreva, V. (2016). *Moral judgments: Studying people with helping professions*. Unpublished manuscript. Department of Cognitive Science and Psychology. New Bulgarian University. Retrieved from https://mindmodeling.org/cogsci2016/papers/0132/paper0132.pdf

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind and Language, 22*(1), 1-21.

Hojat, M., Mangione, S., Nasca, T. J., Rattner, S., Erdmann, J. B., Gonnella, J. S., & Magee, M. (2004). An empirical study of decline in empathy in medical school. *Medical Education, 38*(9), 934-941. doi:10.1111/j.1365-2929.2004.01911.x

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*, 151-158.

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,* 770-778. doi: 10.1016/j.evolhumbehav.2017.06.001

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*, 364-370. doi: 10.1037/a0025561

Neumann, M., Edelhäuser, F., Tauschel, D., Fischer, M. R., Wirtz, M., Woopen, C., . . . Scheffer, C. (2011). Empathy decline and its reasons: a systematic review of studies with medical students and residents. *Academic Medicine, 86*, 996-1009.

Newton, B. W. (2013). Walking a fine line: is it possible to remain an empathic physician and have a hardened heart? *Frontiers in Human Neuroscience, 7*(233), 1-12. doi: 10.3389/fnhum.2013.00233

Pan, X., & Slater, M. (2011). *Confronting a moral dilemma in virtual reality: a pilot study.* Paper presented at the Proceedings of the 25th BCS Conference on Human-Computer Interaction.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience, 9*(660), 1-19. doi: 10.3389/fnhum.2015.00660

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*, 94-107. doi: 10.1080/17470919.2013.870091

Patil, I., Zanon, M., Novembre, G., Zangrando, N., Chittaro, L., & Silani, G. (2017). Neuroanatomical basis of concern-based altruism in virtual environment. *Neuropsychologia.* doi: 10.1016/j.neuropsychologia.2017.02.015

Rand, D. G., & Epstein, Z. G. (2014). Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PLoS ONE*, *9*(10). doi: 10.1371/journal.pone.0109687

Ransohoff, K. (2011). *Patients on the trolley track*. Undergraduate thesis. Harvard University, Cambridge, MA.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133-148.

Rovira, A., Swapp, D., Spanlang, B., & Slater, M. (2009). The use of virtual reality in the study of people's responses to violent incidents. *Frontiers in Behavioral Neuroscience, 3*(59). doi:10.3389/neuro.08.059.2009

Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *Journal of Neuroscience, 34*, 4741-4749. doi:10.1523/Jneurosci.3390-13.2014

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. doi:10.3389/fnbeh.2014.00426

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., . . . Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *Plos One, 1*(1), 1-10.

Stratton, T. D., Saunders, J. A., & Elam, C. L. (2008). Changes in medical students' emotional intelligence: an exploratory study. *Teaching and Learning in Medicine, 20*(3), 279-284. doi:10.1080/10401330802199625

Tassy, S., Deruelle, C., Mancini, J., Leistedt, S., & Wicker, B. (2013). High levels of psychopathic traits alters moral choice but not moral judgment. *Frontiers in Human Neuroscience, 7*(229), 1-6. doi:10.3389/Fnhum.2013.00229

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology, 4*(250), 1-8. doi:1 0.3389/Fpsyg.2013.00250

Thomas, M. R., Dyrbye, L. N., Huntington, J. L., Lawson, K. L., Novotny, P. J., Sloan, J. A., & Shanafelt, T. D. (2007). How do distress and well-being relate to medical student empathy? A multicenter study. *Journal of General Internal Medicine, 22*(2), 177-183.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *Monist, 59*, 204-217.

Wagner, S. L., Martin, C. A., & McFee, J. A. (2009). Investigating the" Rescue Personality". *Traumatology, 15*, 5-12.

Williams, B., Boyle, M., Brightwell, R., Devenish, S., Hartley, P., McCall, M., . . . Webb, V. (2012). Paramedic empathy levels: results from seven Australian universities. *International Journal of Emergency Services, 1*, 111-121.

Williams, B., Brown, T., McKenna, L., Boyle, M. J., Palermo, C., Nestel, D., . . . Russo, V. (2014). Empathy levels among health professional students: a cross-sectional study at two universities in Australia. *Advances in Medical Education and Practice, 5*, 107.