teorema
Vol. XLI/1, 2022, pp. 7-27
ISNN 0210-1602
[BIBLID 0210-1602 (2022) 41:1; pp. 7-27

Hit by the Virtual Trolley: When is Experimental Ethics Unethical?

Jon Rueda

RESUMEN

El dilema del tranvía es uno de los marcos de investigación más vivos de la ética experimental. En la última década, la neurociencia social y la psicología moral experimental han ido más allá de los estudios con dilemas morales meramente hipotéticos basados en textos. En este artículo, presento los fundamentos para investigar la conducta moral en escenarios más realistas a través de la Realidad Virtual y resumo las evidencias generadas por los experimentos con tranvías virtuales. A continuación, analizo el argumento de Ramirez y LaBarge (2020), quienes afirman que la simulación virtual de la versión de la pasarela peatonal del problema del tranvía es una práctica de investigación poco ética, y planteo algunas objeciones al respecto. Por último, ofrezco algunas reflexiones sobre los medios y los fines de dilemas del tipo del tranvía y otros dilemas sacrificiales en la ética experimental.

PALABRAS CLAVE: *dilema del tranvía, ética de la investigación, ética experimental, realidad virtual.*

ABSTRACT

The trolley problem is one of the liveliest research frameworks in experimental ethics. In the last decade, social neuroscience and experimental moral psychology have gone beyond the studies with mere text-based hypothetical moral dilemmas. In this article, I present the rationale behind testing the actual behaviour in more realistic scenarios through Virtual Reality and summarize the body of evidence raised by the experiments with virtual trolley scenarios. Then, I approach the argument of Ramirez and LaBarge (2020), who claim that the virtual simulation of the Footbridge version of the trolley dilemma is an unethical research practice, and I raise some objections to it. Finally, I provide some reflections about the means and ends of trolley-like scenarios and other sacrificial dilemmas in experimental ethics.

KEYWORDS: *Experimental Ethics, Research Ethics, Trolley Problem*, *Virtual Reality.*

INTRODUCTION

For better or worse, the trolley problem has been one of the flagships of experimental ethics in the last two decades. The empirical study of morality has been particularly buoyant at a famous spotlight — a runaway trolley that threatens the life of anyone on the tracks. The landmark dilemma was originally proposed as a thought experiment by Philippa Foot (1967) and goes as follows. A driver of an out-of-control trolley faces the

decision of steering from one track where it is going to kill five workers into another track where, conversely, it will kill one worker. In another usual description, a bystander near the rails can turn a lever to divert the runaway trolley from the five trapped innocent people to another track in which there is one trapped innocent person. This is the so-called Switch case of the trolley dilemma — also known as Side-track or Spur version.

The trolley problem has a great exploratory potential for ethics and moral psychology. Due to its protean character, it can be reformulated in a plethora of versions and subversions. When one or more of the characteristics of the scenario are modified, it allows us to test whether our moral judgements and corresponding justifications change according to the new variables. For instance, Judith Jarvis Thomson (1985) proposed a famous variant in which the only way to stop the trolley is by pushing a large person from a footbridge to the railway to save the five trapped workers. Although the action has the same consequences as in the original version (kill one to save five), fewer people would find shoving the large person morally permissible — 12% of respondents according to the large-scale survey of Marc Hauser and colleagues (2007). This variant is often referred to as the Footbridge or Bridge dilemma or, as originally named, the Fat Man version.

The use of these trolley scenarios has a consolidated pedigree in the empirical study of human morality. Moral philosophers, jurists, experimental moral psychologists, neurocognitive scientists, anthropologists and sociologists of morality, or cross-cultural researchers have investigated moral cognition applying these types of dilemmas. Therefore, the trolley problem can be considered, as Fiery Cushman (2013), p. 274, said, as a "lingua franca" that has facilitated the common understanding and investigation between several domains of the science of morality. Moreover, these trolley studies have led to the emergence of conspicuous empirically ingrained theories of moral judgement, such as dual-process theories [Greene et al. (2001); Cushman (2013)], universal moral grammar theory [Mikhail (2007)], or two-dimensional model of utilitarianism [Kahane et al. (2018)].

However, the so-called "trolleology" has suffered from a significant limitation. Most of the research has focused on vignette-based studies on hypothetical moral judgement. These experiments may seem based on artificial constructs that lack ecological validity and whose results may be inaccurately extrapolated to real-life situations. To overcome these limitations, some researchers have defended the recreation of trolley dilemmas through Virtual Reality (VR) technologies to test the actual behaviour of participants. During the last decade, both the Switch version [Navarrete

et al. (2012), Patil et al. (2014), Skulmowski et al. (2014), Kang et al. (2019)] and the Footbridge variant [Francis et al. (2016), (2017), (2018)] have been virtually simulated for research purposes. Although these contributions open valuable doors for moving from studies of moral judgments to studies of moral actions in these dilemmas, some concerns have also arisen. In particular, Erick Ramirez & Scott LaBarge (2020) have recently argued that the virtual simulation of the Bridge variant of the trolley dilemma for research purposes is largely unethical.

In this article, I address the theoretical grounds and the normative implications of VR experiments with trolley dilemmas. My argument proceeds as follows. In the first section, I summarize the rationale and the main results of studies carried out on real behaviour in virtual trolley simulations. Then, in the second section, I present the argument of Ramirez & LaBarge (2020) about the ethical impermissibility of simulating the Footbridge dilemma. I counter-argue the reasons they presented, showing that their arguments are not convincing enough to support such a claim. Finally, I approach the dispute over the normative value of the empirical evidence raised by trolley-style sacrificial dilemmas, advocating the need for further research to make empirically informed public decisions on some contemporary moral controversies.

I. A Proactive Turn to Trolleyology?

In the study of human morality, there is a relevant distinction between what we say and what we do [FeldmanHall et al. (2012)]. Although judgement and behaviour are connected [Edmonds (2014), p. 141], they often respond to different psychological mechanisms [Tassy et al. (2013)]. Similarly, in the empirical research on trolley cases, hypothetical moral judgement and actual behaviour do not necessarily need to correspond and they might be influenced by different factors. Still, studying the actual behaviour in trolley (i.e., sacrificial) scenarios in the real world would be practically challenging — and arguably unethical. In this section, I approach various experiments about the choice of action of participants in trolley dilemmas using VR technologies. But before presenting them I shall briefly clarify what VR is.

VR is a set of technologies that enable immersive experiences in digitally-created environments [Slater & Sanchez-Vives (2016), Bailenson (2018)]. When the technological affordances of VR (generally through a head-mounted display) elicit a remarkable degree of immersiveness, they have the potential to enable the sense of presence — the psychological

experience of "being there" [Heeter (1992)]. This place illusion – or qualia of being located in the virtual environment [Slater (2009)] – is what facilitates the vividness of immersive experiences. The subjective feeling of presence is the psychological guarantee that makes VR work. When the sensation of presence is successfully achieved, our perceptual, physiological, and motor systems function similarly to how they would do in the "real world," that is, behaving akin to a non-virtual situation [Slater (2009), Slater & Sanchez-Vives (2016), Bailenson (2018)]. Consequently, it has been argued that VR has a remarkable ecological validity to research moral dilemmas in social neurosciences and experimental psychology [Parsons (2015)]. This potential has attracted the attention of researchers interested in studying the actual behaviour of people in trolleyesque scenarios.

In 2011, Xueni Pan and Mel Slater proposed going beyond the mere questionnaire-based study about moral judgement in trolley dilemmas [Pan & Slater (2011)]. They argued that VR offers the possibility of placing participants *in vivo* in trolley-like situations to test their effective behavioural responses in these moral crossroads. However, they found it practically challenging to mimic the classic trolley settings and created a different studio design. In their pilot study, participants encountered a sacrificial scenario in an art gallery. The only way of saving the lives of the five visitors from a shooter on a lift was by switching down the aggressor on the elevator to the ground floor, where another visitor who was standing there would die as a result. Pan and Slater showed that participants responded vividly to this dilemma: "we observed nervous and panicked responses of the participants, and the post-experiment discussions with the participants supported the notion that participants had found themselves responding as if it were real" (p. 51). They also showed that participants gave more utilitarian responses[1] in the more immersive CAVE [a cube-shaped VR room, see Cruz-Neira et al., (1993)] than in a non-immersive desktop (i.e., computer screen).

Shortly thereafter, in 2012, David Navarrete and colleagues published a seminal experiment in which they tested the actual behaviour of participants in the Switch version [Navarrete et al. (2012)]. To meet that purpose, they created a 3-D virtual simulation of the trolley scenario. The immersive environment included visual and audio representations, and the option to switch the rail was enabled via the manipulation of a force-feedback joystick. The results obtained were similar to a previous large-scale survey of Hauser et al. (2007): 89% of participants adopted at a practical level the typically utilitarian response of saving five at the cost of one life. Navarrete et al. also measured the autonomic arousal of par-

ticipants through electrodermal activity showing that emotional arousal reduced the likelihood of acting in an utilitarian way, consistently with the dual-process theory of Joshua Greene and colleagues (2001). This pioneering behavioural study on the morality of harm with virtual trolleys has been followed by subsequent experiments.

Indrajeet Patil and colleagues (2014) showed that participants in the non-immersive VR desktop (i.e., computer monitor) condition were more emotionally aroused and acted in a more utilitarian way than those in the text condition, who were less emotionally aroused and gave fewer utilitarian responses. Autonomic arousal was measured through skin conductance activity. Patil et al. claimed that the increase in emotional arousal in the VR condition could be due to the contextual saliency of the dilemma in the virtual environment in comparison to the text version. Aware that their results were inconsistent with the dual-process theory of Greene et al. (2001), which states that emotional arousal interferes with the cognitive-related process of utilitarian responses, they interpreted their finding as supporting the second dual-process framework of Fiery Cushman (2013). This dual-system theory ascribes different value representations based on the action-type or on the outcome. Both of these competing and complementing processes have affective content that motivates behavioural responses. Patil et al. hypothesized that the contextual saliency of the virtual experience could make the negative consequences of inaction more explicit, consequently making participants more sensitive to outcomes.

Moreover, Alexander Skulmowski and colleagues (2014) tried to overcome some of the limitations of Navarrete et al.'s study through a different experimental setup. Unlike Navarrete et al.'s studies in which participants were passive bystanders, Skulmowski et al. placed research subjects as the drivers of the train[2]. With this change to the first-person perspective, researchers tried to elicit higher involvement and spatial presence in participants. Moreover, instead of recording electrodermal activity, they used pupillometric measurements of emotional arousal. The increase in the pupil diameter of participants in all conditions showed an increased level of arousal, which, however, did not affect the tendency to give utilitarian responses. Skulmowski et al. also modified across conditions the features of the target people to be sacrificed, varying their gender, skin colour, and body orientation (avatar on the rails facing or not towards the driver), without achieving significant results — except that more men were sacrificed than women. Overall, the study shows that participants sacrificed one to save the group in 96% of the cases, repli-

cating results similar to pen-and-paper studies and reinforcing the validity of VR as a research tool for trolley dilemmas.

Kathryn Francis and colleagues also investigated the disparity between moral judgement and moral action in the trolley problem [Francis et al. (2016), (2017), (2018)]. They were the first to create the Footbridge variant in VR for research purposes. Since their objective was to understand the differences between how we judge the morality of this personal dilemma and how we would really act in it, they divided participants into two conditions. In the judgement condition, participants were asked not only how they judge (whether they consider it morally permissible or not) throwing the large person, but also what they would *do* in that situation. In the action condition, participants were exposed to the virtual dilemma being able to push the large person. The results of their studies are surprising. According to the study with the bigger sample, in the judgment condition, only 10% of participants considered pushing the person morally acceptable and also only 10% of participants said that they would perform that action [Francis et al. (2016)]. Astonishingly, in the action condition, 63.3% of participants endorsed a utilitarian action — 15.55 times higher than in the judgment condition. In other words, participants were more utilitarian in their actions than in their judgments regarding the morality of pushing a person to the rails to save five. They also showed that participants in the action condition were more aroused than in the judgment condition — physiological arousal being measured by recording the heart rate. Moreover, they stated that the increase in the emotional arousal was not only a consequence of the virtual experience per se, but it was also an outcome of the moral content of the action of throwing the person. In 2017, they published another article with similar results, but in which the action of pushing was even more realistic via haptic feedback devices — a robotic haptic interface that enabled a faithful physical action with personal force or an interactive human-like sculpture mechanism [Francis et al (2017)]. To explain these results, they followed Patil et al. (2014) and interpreted this tendency as congruent with the dual-system theory of Cushman (2013). Although the action of pushing a person to death has great disvalue, the contextual features of the virtual task make the consequences of inaction more salient and, as a result, the outcome of allowing five people to die achieves greater disvalue.

After having considered the previous experiments, the rationale of these trolley studies through VR might seem more manifest. VR opens promising avenues for moral cognition research. Simulating virtual actions permits testing the morality of diverse decisions in life-or-death sit-

uations. VR experiments also offer more realistic experiences than vignette-based text versions. In this way, we can bridge the gap between what we say and what we do, studying the differences and similarities that pervade between mere moral judgement and more realistic (even not real-world) virtual behaviour. Furthermore, VR research can integrate various physiological measures (e.g., heart rate, pupil diameter, skin conductance, electrodermal activity, and so on) and haptic input devices that record sensorimotor data with the aim of unravelling further factors that underlie our moral decisions.

Of course, some could object that very famous pen-and-paper studies of the trolley problem had also recurred to other technologies to shed light on the cognitive and affective factors that underpin moral evaluations. The conspicuous fMRI investigations of Joshua Greene and colleagues (2001) could be suitable examples of that. Still, David Edmonds claimed that those text-based neuro-studies had limited ecological validity:

> Brain-scanning is indeed still a crude tool with crude measurements. And gauging the response of subjects while they are lying prone in a long tube can hardly replicate any real-life dilemma. However deeply the patients immerse themselves in the dilemma, however successful they are in imagining themselves inside it, in suspending disbelief, they're unlikely to feel the thumping heart, the sweaty palms, the fear, panic, and anxiety of real life. The ordinary sounds, smells, and sights are absent. (…) But real life does contain multiple influencing factors, so we should be wary of extrapolating from the white tube to real life. [Edmonds (2014), p. 150].

Research with VR can circumvent that objection of the limited ecological validity to vignette-based studies. In fact, it offers various remarkable advantages in this respect. First and foremost, moral cognition is context-dependent. VR permits to include the situative factors that influence our embodied moral decision-making. Unlike test-based questionnaires, the virtual scenario can include the experimental stimuli of life-like visual information and sensory engagement to achieve more contextual saliency [Patil et al. (2014), Skulmowski et al. (2014)]. Second, to the extent that VR includes the situational features of the environment, it also overcomes the psychological limitation of relying on the mental representations of participants in perspectival thought experiments [Ramirez, (2017), p. 519]. Third, the ecological validity of the body of evidence gathered in these VR experiments can illuminate the genuine disparities between moral action and moral judgement. It seems that inactions get greater disvalue in VR than on pen-and-pencil studies — especially in the

Footbrige version [Francis et al. (2016), (2017)]. Although we must be cautious when extrapolating these results to the real-world, these studies suggest how our hypothetical moral judgments may differ from our moral behaviour in real (and exceptional) trolley-like situations in non-virtual situations. Fourth, it has been claimed that researching on the actual behaviour on real trolley dilemmas in the non-virtual world would be unsafe and unethical [Pan & Slater (2011), Skulmowski et al. (2014)]. Conversely, the experimental execution on VR would avoid putting anybody's life in danger — except for those non-sentient virtual avatars.

Summarizing, VR may pave the way to a more ecologically valid study of moral decisions on trolley-style dilemmas. In particular, the previous experiments shed light on the need to reinforce the research on the differences between what we say and what we actually do in sacrificial scenarios. However, the fact that the trolley goes virtual is not excluded from some (ethical) troubles. In the next section, I will address some research ethics concerns related to the study of the Footbridge variant of the trolley problem in VR.

II. The Argument of Ramirez & LaBarge and its Problems

Erick J. Ramirez and Scott LaBarge (2020) have argued that simulating the Footbridge version in VR is a misleading aspiration. They developed the following two arguments to support that claim: (*a*) this dilemma is *practically* impossible to simulate, and, moreover, (*b*) it is *unethical* to expose participants to such a situation.

On the one hand, they defended that the practical simulation of the Footbridge version is doomed to fail. Its simulation is practically challenging because the design could not meet two requirements that they considered fundamental. According to Ramirez and LaBarge (2020), pp. 3317-8, the virtual footbridge scenario cannot comply neither with *perspectival fidelity* ("the degree to which a simulation accurately reproduces the structural perspectival features of human experience") nor with *contextual realism* ("the degree to which the content of a simulation coheres with the rules of the actual world, as understood by the user"). In particular, they criticized the experiment of Francis et al. (2016). Some design features could make that research setting a context-unreal environment. For instance, they argued that the simulation of the pushing experience was poorly recreated. (However, Ramirez & LaBarge (2020) ignored that this limitation was already overcome by the experiment of Francis et al. (2017), in which a realistic pushing experience was achieved via robotic

haptic interface or via a human-like interactive sculpture.) Furthermore, they criticized the way Francis et al. conveyed crucial information to participants. Whereas a key element of the original Footbridge dilemma is that the protagonist *knows* (by "being expert on trolleys") that shoving the fat man *will* stop the trolley [Thomson (1985), p. 1409], Francis et al. (2016) used a non-diegetic voice-over to alert participants that the runaway trolley was coming and that pushing the person would stop its trajectory. In what follows, I will not address these practical issues further in order to focus more closely on their ethical concerns.

On the other hand, Ramirez & LaBarge noted that, even if the practical problems could be overcome, ethical issues should prevent the simulation from being performed. They presented two arguments to defend their position: (*1*) the generation of trauma in the research subjects, and (*2*) the long-term development of undesirable moral traits (such as seeing people as mere instruments) that can be condemned by virtue ethics, deontology, and consequentialism. After analysing each objection, I will argue that neither of the two arguments is satisfactory enough to claim that the simulation of the Footbridge variant is unethical.

Firstly, this simulation may cause undue stress to participants. This is an empirical and plausible concern. One reason to hypothesize that participants could suffer distress comes from a previous influential experiment. Mel Slater and colleagues (2006) simulated the famous study on obedience of Stanley Milgram (1963), replicating its results and showing that participants felt distressed by giving electric shocks to the virtual learner.[3] That is, research subjects can get stressed if they think that they are harming the virtual avatars — even if they know that the scene is not "real". Therefore, if the distress caused in the virtual Footbridge dilemma is so intense as to create trauma in the participants, it would be ethically impermissible to simulate it.

While this argument may have some appeal, it masks a fundamental omission. Ramirez and LaBarge do not mention that simulating the Switch version may also cause similar problems. In fact, in the seminal experiment of Navarrete et al. (2012), 15 participants out of 365 were removed from the study because they did not finish the task due to distress or discomfort. Moreover, in their pilot study, Pan and Slater (2011), p. 51, also reported that they "observed nervous and panicked responses". Patil et al. (2014) and Skulmowski et al. (2014) also mentioned a significant level of arousal in their participants. Could these experiments on the Switch version have caused any trauma in their research subjects? Although they certainly created some stress, this is not enough to affirm that

participation produced trauma. We would need specific post-experimental studies to be able to confirm the generation of trauma. Likewise, we do not have evidence to claim that the studies on the Footbridge version were traumatic for the experimental subjects. Francis et al. (2016), (2017), (2018), did not report the withdrawal of any participants due to stress.[4] These studies received, moreover, ethical approval from their university ethics committee – following its specific guidelines and regulations – and obtained informed consent from all participants.

Consequently, at least two things are clear. First, it is very striking that Ramirez and LaBarge ignored or made no reference to the psychological stress generated in experiments of the virtual Switch variant of the dilemma.[5] Second, for consistency with their argument, they should also reject these experiments in case they also had a traumatizing potential — a fact that I believe we do not have sufficient evidence to affirm at all. Furthermore, research in both classic dilemmas share the characteristic that, beyond remuneration and the moral value of contributing to scientific progress, neither benefits the participating subjects. This is also sometimes the case in biomedical research. Thus, drawing a dividing line to allow virtual experiments on the Switch variant and not allowing the Footbridge version is an ethically unjustifiable arbitrariness in the absence of further evidence on its *hypothetical* traumatizing effects.

Secondly, Ramirez and LaBarge argue that the Footbridge experiment can inculcate the bad trait of seeing persons as means to solve moral problems. The inculcation of an instrumental view of persons can lead to an undesirable character change. They argue that this subsequent change in moral traits is regrettable from virtue ethics, deontology, and consequentialism (especially from rule-bound versions and from accounts that take into account the long-term impact of particular moral traits in society's welfare). I think that there is some element of truth in this second concern. Primarily, it is correct that lived experiences in VR can affect our moral traits in the real world, as demonstrated by the case of empathy enhancement [Rueda & Lara (2020), Lara & Rueda (2021)]. Indeed, the promises and fears of the diverse therapeutic, recreational, and social awareness applications of VR are rooted in its impact on our behaviour in the non-virtual world.

However, I find this second argument controversial and, all things considered, too weak. There are various problems with this argument. On the one hand, this inculcation would only take place in the cases of people who actually pushed the person onto the tracks. In the cases that refrained from pushing – the not negligible percentage of around 40% of

participants according to Francis et al. (2016) – could we say that good moral traits such as not seeing others as means have been inculcated? Would they constitute cases of moral enhancement? Ramirez and LaBarge appear not to have considered this possibility. And, if one accepts the premise that experiencing this virtual dilemma can instil both morally good and bad traits, depending on the action-choice, we should also take into account the opportunity cost of not reinforcing good moral traits in the people that abstain from shoving the large man.

On the other hand, and more importantly, this argument takes too many things for granted that should be proved. First, they should demonstrate that such undesirable traits are indeed inculcated in the long run. This is an empirical question. One way to test this hypothesis could be, after some time has passed, to conduct experiments with the same subjects (who pushed the person onto the pathways) by placing them in other dilemmas in which there is the possibility of using other people as means. Thus, for the moment, the authors lack evidence to support that claim. Second, they should prove that pushing the person is produced by (or at least through the co-presence of) the moral trait that leads to instrumentalization. It may be that this action is produced by other moral traits, further situational factors, or different ethical commitments. Third, they should persuade us that pushing the person is a morally wrong choice in this dilemma. It seems to me that they are asserting in a veiled way that shoving the person is morally wrong — because the victim would be treated as a means, which shows an undesirable moral trait of the agent. But, to my knowledge, the authors have not developed their position regarding what is the best choice in this dilemma. I think that, therefore, they beg the question.

Last but not least, Ramirez and LaBarge's assertion of the ethical impermissibility of this simulation is too categorical. In fact, I think that some of their ethical concerns are largely mitigable. There are a variety of ways to underpin the ethical permissibility of this experimental simulation:

- Warning in the informed consent document about the potentially stressful task of the experiment.
- Guaranteeing the possibility of withdrawing at any time without any penalty concerning remuneration.
- Offering professional psychological consultations after completion of the study to detect possible cases of trauma and provide free treatment in such diagnosed cases.

• Increasing incentives (either monetary, academic credits for university students, or charity donations), and consequently offering a greater proportional reward to the research subjects.

Those are very simple strategies for reducing the research ethics concerns raised by these simulations. This way we would not have to forego the value provided by these virtual trolley experiments, which help us to cement scientific knowledge about human morality. I will discuss precisely the value of these empirical studies and other sacrificial dilemmas in the next section.

III. Rethinking the Means and Ends of Trolleyology

Although the history of experimental ethics is just beginning to be written [Dworazik & Rusch (2014)], this (inter)discipline has a long way to go. What role should the trolley problem play in future research? And what is its value for public controversies? In this last section, I will connect trolley experiments with other studies on sacrificial dilemmas. My purpose is to show that the relevance of these investigations is not so detached from the reality of some contemporary moral problems. Then, I will briefly address the controversy over what normative relevance derives from the empirical evidence gathered on these issues.

The trolley problem is not *a* problem. Rather, it may be conceived as a multifaceted framework in which a wide variety of investigations about how humans morally evaluate life and death decisions may proliferate. In other words, this framework is one of the sacrificial settings par excellence. A sacrificial dilemma is a situation in which any of the alternative courses of action (potentially) entails the death of one or multiple subjects. Although the individual peculiarities of each sacrificial dilemma vary from case to case, there are a number of moral aspects that tend to recur: the distinction between actions and omissions, the consequences, the distribution of harm, positive and negative duties, intention (including the differences between intended effects versus foreseen side effects), and so on. Those characteristics have not only been discussed concerning the trolley variants but are also present in other sacrificial dilemmas. Consider triage practices and autonomous vehicles.

On the one hand, triage is the "process of classifying patients according to their medical needs and severity" [Rueda (2021a)]. Triage becomes necessary when the demand for specific medical services or resources (e.g., mechanical ventilators or intensive care unit beds) significantly exceeds the supply. In mass casualty situations, rationing those

life-saving resources often implies that the non-selected patients may eventually die. In that sense, the COVID-19 pandemic has boosted the interest of experimental ethics surveys on these dilemmatic decisions [Rueda et al. (2020), Kneer & Hannikainen (2021)]. For example, Markus Kneer and Ivar Hannikainen (2021) have argued that studies on critical care dilemmas related to COVID-19 are more psychologically salient and have greater ecological validity than hypothetical trolley scenarios and that, therefore, they provide results that are more generalizable to the real-world. Moreover, VR has been used for paramedic training on triage and mass casualty incidents [Berndt et al. (2018)].

On the other hand, autonomous vehicles will encounter trolley-like unavoidable collisions [Keeling (2020)]. Recently, famous studies have addressed the public preferences on the diverse fatal accidents of driverless cars [Bonnefon et al. (2016), Awad et al. (2018)]. VR has also been used to test participants' responses to autonomous vehicles in traffic dilemmas [Sütfeld et al. (2017), Faulhaber et al. (2019)]. Overall, media attention and academic interest in this topic is growing steadily due to the pressing need for a solution from car manufacturers and legal systems on the liability in these trolley-type crashes.

So, experimental ethics studies on triage and autonomous vehicles are overcoming the alleged limits of the ecological validity of hypothetical trolley situations — sometimes also using VR. But why are we interested in knowing what people think and do in such sacrificial dilemmas? What are the benefits of researching those topics for public policy debates? Inferring normative conclusions (about what we should do) from these descriptive experiments (about what people do or think) is not unproblematic. In consequence, what is the normative relevance of the empirical moral research on trolley-like sacrificial dilemmas? I will develop three responses: the first two are unsatisfactory and the last one is, in my view, the most attractive one.

The first response states that these studies have no normative relevance at all. A traditional objection to the trolley dilemma pointed to the artificiality of the scenario and its normative uselessness in translating to real contemporary problems [see, for instance, Midgley, cited in Edmonds (2014), p. 100-101]. We have already seen that this is not true. Indeed, the existence of real dilemmas that share structural similarities with hypothetical trolley scenarios makes it practically useful to test our intuitions on them [Edmonds (2014)]. Besides that, a more sophisticated objection claims that intuitive responses to the trolley problem have no ethical value because intuitions are quite unreliable. Cognitive science has

frequently shown how fallible, illogical, biased, and irrational many of our intuitive preferences can be. In fact, moral intuitions in text-based trolley dilemmas are subject to morally irrelevant factors such as order [Liao et al. (2012)], frame [Cao et al. (2017)], or mood [Pastötter et al. (2013)]. However, the fact that there are wrong or biased intuitions does not mean that intuitions do not have *any* epistemic or moral value. Dismissing intuitions because they are subject to implicit psychological factors in favour of armchair ethical theorizing is inconsistent. Empirical evidence should play a role in normative theorizing on trolley dilemmas as long as ethical theorizing is also subject to implicit psychological factors — and which experimental research can help to make explicit [Kahane (2013)].

The second option states that what should be done as public policy on sacrificial dilemmas is what the majority of people say or do in those situations. In other words, the descriptive results of the experiments show us how we should act at the normative level. Consider the following example from the debate of self-driving vehicles: "We thus argue that any implementation of an ethical decision-making system for a specific context should be based on human decisions made in the same context" [Sütfeld et al. (2017)]. So, as most people act in a utilitarian way in VR simulations of traffic dilemmas, autonomous cars should act similarly in analogous situations [Sütfeld et al. (2017)].

Although this view might have some appeal, it has at least two problems. Firstly, it runs the risk of conflating the 'is' with the 'ought'. Recent debates in neuroscience of morality have warned again of the need to avoid the naturalistic fallacy — the logical invalidity of inferring moral properties from any natural set of facts [Greene (2003), (2014), Berker (2009), Bruni et al. (2014), Rueda (2021b)]. The kernel of the caveat lies in the notion that from a *mere* factual statement we cannot by itself (in the absence of other premises) infer a normative statement [Kahane (2016), Aguiar et al. (2020), appendix]. Thus, although people may act in a utilitarian way in traffic dilemmas, it does not logically follow that the moral algorithm of autonomous vehicles should act that way — even if one is theoretically sympathetic to utilitarianism. Second, and more importantly, this view leaves little room for expert judgements on the application of the most consolidated ethical theories. From this position, that is irrelevant because the only thing that matters is what ordinary folk do or think. However, as Savulescu and colleagues (2021), p. 4, said, "(t)o make ethical decisions a matter of referendum is to eschew ethical expertise and professional responsibility". In fact, "we might reasonably expect an autonomous vehicle to be programmed to make better moral

decisions in a collision than human drivers make in analogous collisions" [Keeling (2017), p. 2]. And to raise the ethical bar for policies on triage and autonomous cars, we not only need to collect public preferences, but we also need to test them against leading moral theories.

Third and last, there is a more promising way to devise empirically-informed public ethics on sacrificial dilemmas. From this view, public attitudes derived from experimental studies should not play an overriding contribution to normative guidelines. Public preferences can often be subject to misinformation, bias, prejudice, bad reasoning, or indifference to major ethical theories [Savulescu et al. (2019)]. However, collecting the responses and discovering their underlying psychological factors can contribute to responding to substantive ethical questions. Of course, experimental studies do not generate normative conclusions on their own, but rather provide the empirical data that underpin the reliability of the premises that are necessary for normative inferences [Earp et al. (2021)]. If it is discovered, for example, that predominant responses are caused by morally irrelevant factors, they could be called into question. Experimental ethics can therefore help to understand the cognitive processes that influence people's moral judgements and actual behaviours in ecologically valid contexts. And in the discussion of sacrificial dilemmas, it is important to know whether responses and decision processes change from abstract scenarios to life-like realistic contexts (such as VR simulations) in which one can actually react [Savulescu et al. (2021), p. 7].

I also believe that this third option has an attractive advantage. It allows us to seek coherence between folk intuitions and behaviours with ethical principles and theories. A prominent way in the search for ethical coherence is that of reflective equilibrium, repeatedly mentioned in experimental ethics [Kahane (2013), Aguiar et al. (2014), Savulescu et al. (2021)]. John Rawls (1971) famously proposed the method of the reflective equilibrium as a mutual adjustment between ethical principles and people's moral judgement on particular cases. This method constitutes a two-way street (with top-down and bottom-up routes) that permits confronting folk moral intuitions with major ethical commitments. In this sense, Savulescu et al. (2021) have suggested that the strongest public preferences (of laypeople and experts) on self-driving cars dilemmas should be confronted with leading ethical approaches. The idea is not to hide disagreements but to give greater legitimacy and justification in cases where public preferences cohere with general theories. In short, applying the reflective equilibrium to sacrificial dilemmas allows us to reconnect the

scientific-descriptive part of experimental methodologies with the normative function of ethics.

Last but not least, if experimental studies on these contemporary sacrificial dilemmas can contribute in some way to normative debates, as I have just argued, I believe this is an added incentive to continue their research. To the extent that these ethical debates are still ongoing, and their corresponding public policies are under constant construction, we should keep investigating how citizens realistically react in such tragic situations. And therefore, we have ethical reasons to continue researching them through VR.

CONCLUDING REMARKS

Recent debates in experimental ethics show an apparent paradox in which trolleology seems to move between two contradictory inclinations. On the one hand, moral judgments about thought experiments have clear limitations, so we need realistic simulations to study how people truly act in trolley-style situations. On the other hand, realistic VR simulations of such dilemmas may raise research ethics controversies, and beyond their descriptive scientific value, the moral value of those experiments might seem unclear. However, this apparent paradox is not insurmountable. In this article, I have first defended that research ethics concerns in virtual trolleology are mitigable. Secondly, I have also claimed that the experimental study of trolley-like scenarios may pave the way to (modest) normative inferences through empirically informing public guidelines about real-life sacrificial dilemmas like triage or autonomous vehicles. The paradox is thus dissolved. Virtual trolleys are here to stay.

Department of Philosophy 1
University of Granada
Faculty of Psychology, Campus Universitario de Cartuja
CP: 18071, Granada, Granada, Spain
E-mail: ruetxe@ugr.es

ACKNOWLEDGEMENTS AND FUNDING

This article has benefited from the previous discussion with Kathryn Francis, Francisco Lara and Belén Liedo, to whom I am deeply grateful. I would also like to acknowledge the valuable suggestions of María Jiménez Castro that

have helped me to improve my writing. This research is funded by an INPhINIT Retaining Fellowship of the La Caixa Foundation (grant number LCF/BQ/DR20/11790005), and the projects EthAI+3 (grant code PID2019-104943RB-I00) and INEDyTO II (grant code PID2020-118729RB-I00), both funded by the Spanish Research Agency.

NOTES

[1] Kahane (2015) challenged the very label of 'utilitarian' to refer to the choice of saving the most lives through instrumental harm in folk studies of the trolley dilemma because it does not generally refer to an impartial concern for the greatest good. For simplicity, here I will use the term 'utilitarian' to be consistent with the predominant literature.

[2] Kang et al. (2019) also tested the responses of participants in the driver position through a VR experiment. However, they were testing driver condition responses in contrast to the victim condition (i.e., where participants could decide to self-sacrifice themselves to save five).

[3] In Milgram's (1963) conspicuous study on obedience, research subjects believed that the study was a learning experiment about the effects of punishment in memory. However, research subjects were deceived by a confederate of the experimenter, who simulated the pain caused by a bogus shock generator when he failed to respond to the correct answers. Since participants believed that they were genuinely causing pain to the learner, they suffered from extreme levels of tension and stress during the experiment. In their virtual reprise of Milgram's obedience experiment, Slater's group also reported that many of the participants were very stressed and even some of them withdrew early from the study [Slater et al. (2006)].

[4] In their study with the helping profession and the Footbridge dilemma, Francis et al. (2018) found that trained individuals showed less arousal and regret than untrained individuals.

[5] It does not appear that Ramirez and LaBarge are interested in discrediting studies of the Switch version of the scenario. In fact, Ramirez et al. have developed a VR simulation model of the original version of Foot's (1967) dilemma. They have altruistically made it open access through PhilPapers, so that it can be used for research or teaching purposes: *https://philpapers.org/rec/RAMVRT-2* (last access on 24 June 2021).

REFERENCES

AGUIAR, F., GAITÁN, A. and RODRÍGUEZ-LÓPEZ, B. (2014), 'Robust Intuitions, Experimental Ethics, and Experimental Economics: Bringing Reflective Equilibrium into the Lab', in Luetge, C., Rusch, H. and Uhl, M. (eds.), *Experimental Ethics. Toward an Empirical Moral Philosophy*, Cham, Springer, pp. 191–208.

AGUIAR, F., GAITÁN, A. and VICIANA, H. (2020), *Una introducción a la ética experimental*, Madrid, Cátedra.

AWAD, E., DSOUZA, S., KIM, R., SCHULZ, J., HENRICH, J., SHARIFF, A., BONNEFON, J. F., and RAHWAN, I. (2018), 'The Moral Machine Experiment'. *Nature*, vol. 563, no. 7729, pp. 59–64. *https://doi.org/10.1038/s41586-018-0637-6*

BAILENSON, J. (2018), *Experience on Demand. What Virtual Reality Is, How It Works, and What It Can Do*, New York, W. W. Norton & Company.

BERKER, S. (2009), 'The Normative Insignificance of Neuroscience', *Philosophy and Public Affairs*, vol. 37, no. 4, pp. 293–329. *https://doi.org/10.1111/j.1088-4963.2009.01164.x*

BERNDT, H., WESSEL, D., WILLER, L., HERCZEG, M., and MENTLER, T. (2018), 'Immersion and Presence in Virtual Reality Training for Mass Casualty Incidents', in *Proceedings of the 15th ISCRAM Conference*.

BONNEFON, J. F., SHARIFF, A., and RAHWAN, I. (2016), 'The Social Dilemma of Autonomous Vehicles', *Science*, vol. 352*,* no. 6293, pp. 1573–1576. *https://doi.org/10.1126/science.aaf2654*

BRUNI, T., MAMELI, M., and RINI, R. A. (2014), 'The Science of Morality and its Normative Implications', *Neuroethics*, vol. 7, no. 2, pp. 159–172. *https://doi.org/10.1007/s12152-013-9191-y*

CAO, F., ZHANG, J., SONG, L., WANG, S., MIAO, D., and PENG, J. (2017), 'Framing Effect in the Trolley Problem and Footbridge Dilemma', *Psychological Reports*, vol. 120, no. 1, pp. 88–101.

CRUZ-NEIRA, C., SANDIN, D. J., and DEFANTI, T. A. (1993), 'Surround-Screen Projection-Based Virtual Reality: the Design and Implementation of the CAVE', in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques,* Anaheim, CA, pp. 135–142.

CUSHMAN, F. (2013), 'Action, Outcome, and Value: a Dual-System Framework for Morality', *Personality and social psychology review*, vol. 17, no. 3, pp. 273–92.

DWORAZIK, N. and RUSCH, H. (2014), 'A Brief History of Experimental Ethics', in Luetge, C., Rusch, H. and Uhl, M. (eds.), *Experimental Ethics. Toward an Empirical Moral Philosophy*, Cham, Springer, pp. 38–56.

EARP, B. D., LEWIS, J., DRANSEIKA, V., and HANNIKAINEN, I. R. (2021), 'Experimental Philosophical Bioethics and Normative Inference', *Theoretical Medicine and Bioethics*, vol. 42, no. 3-4, pp. 91–111. *https://doi.org/10.1007/s11017-021-09546-z*

EDMONDS, D. (2014), *Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tells Us About Right and Wrong*, Princeton, Princeton University Press.

FAULHABER, A. K., DITTMER, A., BLIND, F., WÄCHTER, M. A., TIMM, S., SÜTFELD, L. R., STEPHAN, A., PIPA, G., and KÖNING, 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*, vol. 25, no. 2, pp. 399–418. *https://doi.org/10.1007/s11948-018-0020-x*

FELDMANHALL, O., MOBBS, D., EVANS, D., HISCOX, L., NAVARDY, L., and DALGLEISH, T. (2012), 'What We Say and What We Do: The Relationship Between Real and Hypothetical Moral Choices', *Cognition*, vol. 123, pp. 434–441. *https://doi.org/10.1016/j.cognition.2012.02.001*

FRACIS, K. B., HOWARD, C., HOWARD, I. S., GUMMERUM, M., GANIS, G., ANDERSON, G., and TERBECK, S. (2016), 'Virtual Morality: Transitioning from Moral Judgment to Moral Action?', *PLoS ONE*. *https://doi.org/10.1371/journal.pone.0164374*.

FRACIS, K. B., TERBECK, S., BRIAZU, R. A., HAINES, A., GUMMERUM, M., GANIS, G., and HOWARD, I. S. (2017), 'Simulating Moral Actions: An Investigation of Personal Force in Virtual Moral Dilemmas', *Scientific Reports*, vol. 7, no. 1, pp. 1–11. *https://doi.org/10.1038/s41598-017-13909-9*

FRACIS, K. B.; GUMMERUM, M.; GANIS, G.; HOWARD, I. S. and TERBECK, S. (2018), 'Virtual Morality in the Helping Professions: Simulated Actions and Resilience', *British Journal of Psychology*, vol. 109, no. 3, pp. 442–465.

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

GREENE, J. D., SOMMERVILLE, R. B., NYSTROM, L. E., DARLEY, J. M., and COHEN, J. D. (2001), 'An fMRI Investigation of Emotional Engagement in Moral Judgment', *Science*, vol. 293, no. 5537, pp. 2105–2108. *https://doi.org/10.1126/science.1062872*

GREENE, J. (2003), 'From Neural "Is" to Moral "Ought": What Are the Moral Implications of Neuroscientific Moral Psychology?', *Nature Reviews Neuroscience*, vol. 4, no. 10, pp. 846–850. *https://doi.org/10.1038/nrn1224*

— (2014), 'Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics,' *Ethics*, vol. 124, no. 4, pp. 695–726. *https://doi.org/10.1515/lehr-2015-0011*

HAUSER, M., CUSHMAN, F., YOUNG, L., JIN, R. K.-X., J., and MIKHAIL, J. (2007), 'A Dissociation between Moral Judgments and Justifications', *Mind & Language*, vol. 22, pp. 1–21. *https://doi.org/10.1111/j.1468-0017.2006.00297.x*

HEETER, C. (1992), 'Being There: The Subjective Experience of Presence', *Presence: Teleoperators and Virtual Environments*, vol. 1, pp. 262–271. *https://doi.org/10.1162/pres.1992.1.2.262*

KAHANE, G. (2013), 'The Armchair and the Trolley: An Argument for Experimental Ethics', *Philosophical Studies*, vol. 162, no. 2, pp. 421–445. *https://doi.org/10.1007/s11098-011-9775-5*

— (2015), 'Sidetracked by Trolleys: Why Sacrificial Moral Dilemmas Tell Us Little (or Nothing) About Utilitarian Judgment', *Social Neuroscience*, vol. 10, no. 5, pp. 551–560. *https://doi.org/10.1080/17470919.2015.1023400*

— (2016), 'Is, Ought, and the Brain', in Liao, M. (ed.), *Moral Brains: The Neuroscience of Morality*, New York, Oxford University Press, pp. 281–311.

KAHANE, G., EVERETT, J., EARP, B. D., CAVIOLA, L., FABER, N. S., CROCKETT, M. J., and SAVULESCU, J. (2018), 'Beyond Sacrificial Harm: A Two-Dimensional

Model of Utilitarian Psychology', *Psychological review*, vol. 125, no. 2, pp. 131–164. *https://doi.org/10.1037/rev0000093*

KANG, S., CHANENSON, J., GHATE, P., COWAL, P., WEAVER, M. and KRUM, D. M. (2019), 'Advancing Ethical Decision Making in Virtual Reality', in *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1008-1009. *https://doi.org/10.1109/VR.2019.8798151*

KEELING, G. (2020), 'Why Trolley Problems Matter for the Ethics of Automated Vehicles', *Science and Engineering Ethics*, vol. 26, no. 1, pp. 293–307. *https://doi.org/10.1007/s11948-019-00096-1*

KNEER, M., and HANNIKAINEN, I. R. (2021), 'Trolleys, Triage and Covid-19: The Role of Psychological Realism in Sacrificial Dilemmas', *Cognition and Emotion*. *https://doi.org/10.1080/02699931.2021.1964940*

LARA, F., and RUEDA, J. (2021), 'Virtual Reality Not for "Being Someone" but for "Being in Someone Else's Shoes": Avoiding Misconceptions in Empathy Enhancement', *Frontiers in Psychology*, vol. 12(August), pp. 10–13. *https://doi.org/10.3389/fpsyg.2021.741516*

LIAO, M. S., WIEGMANN, A., JOSHUA, A., and VONG, G. (2012), 'Putting the Trolley in Order: Experimental Philosophy and the Loop Case', *Philosophical Psychology*, vol. 25, no. 5, pp. 661–671.

MIKHAIL, J. (2007), 'Universal Moral Grammar: Theory, Evidence and the Future', *Trends in Cognitive Sciences*, vol. 11, no. 4, pp. 143–52.

MILGRAM, S. (1963), 'Behavioral Study of Obedience', *Journal of Abnormal and Social Psychology*, vol. 67, pp. 371–378.

PAN, X., and SLATER, M. (2011), 'Confronting a Moral Dilemma in Virtual Reality: A Pilot Study', in *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, Newcastle upon Tyne, British Computer Society, pp. 46–51.

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. *https://doi.org/10.3389/fnhum.2015.00660*

PASTÖTTER, B., GLEIXNER, S., NEUHAUSER, T., and BÄUML, K.-H. T. (2013), 'To Push or Not to Push? Affective Influences on Moral Judgment Depend on Decision Frame', *Cognition*, vol. 126, no. 3, pp. 373–377.

PATIL, I., COGONI, C., ZANGRANDO, N., CHITTARO, L., and SILANI, G. (2014), 'Affective Basis of Judgment-Behavior Discrepancy in Virtual Experiences of Moral Dilemmas', *Social Neuroscience*, vol. 9, no. 1, pp. 94–107.

RAMIREZ, E. (2017), 'Empathy and the Limits of Thought Experiments', *Metaphilosophy*, vol. 48, no. 4, pp. 504–526.

RAMIREZ, E. J., and LABARGE, S. (2020), 'Ethical Issues with Simulating the Bridge Problem in VR', *Science and Engineering Ethics*, vol. 26, no. 6, pp. 3313–3331. *https://doi.org/10.1007/s11948-020-00267-5*

RAWLS, J. (1971), *A Theory of Justice*, Cambridge, MA, The Belknap Press of HUP.

RUEDA, J. (2021a), 'Ageism in the COVID-19 Pandemic: Age-Based Discrimination in Triage Decisions and Beyond', *History and Philosophy of the Life Sciences*, vol. 43, no. 3, pp. 1–7. *https://doi.org/10.1007/s40656-021-00441-3*

— (2021b), 'Socrates in the fMRI Scanner: The Neurofoundations of Morality and the Challenge to Ethics', *Cambridge Quarterly of Healthcare Ethics*, vol. 30, no. 4, pp. 604–612. *https://doi.org/10.1017/S0963180121000074*

RUEDA, J. and LARA, F. (2020), 'Virtual Reality and Empathy Enhancement: Ethical Aspects', *Frontiers in Robotics and AI*, 7: 506984. *https://doi.org/10.3389/frobt.2020.506984*

RUEDA, J., HANNIKAINEN, I. R., HORTAL-CARMONA, J., and RODRÍGUEZ-ARIAS, D. (2020), 'Examining Public Trust in Categorical versus Comprehensive Triage Criteria', *American Journal of Bioethics*, vol. 20, no. 7, pp. 106–109. *https://doi.org/10.1080/15265161.2020.1779867*

SAVULESCU, J., KAHANE, G., and GYNGELL, C. (2019), 'From Public Preferences to Ethical Policy', *Nature Human Behaviour*, vol. 3, no. 12, pp. 1241–1243. *https://doi.org/10.1038/s41562-019-0711-6*

SAVULESCU, J., GYNGELL, C., and KAHANE, G. (2021), 'Collective Reflective Equilibrium in Practice (CREP) and controversial novel technologies', *Bioethics*, 1–12. *https://doi.org/10.1111/bioe.12869*

SLATER, M., ANTLEY, A., DAVISON, D., SWAPP, D., GUGER, C., BARKER, C., PISTRANG, N. and SÁNCHEZ-VIVES, M. V. (2006), 'A Virtual Reprise of the Stanley Milgram Obedience Experiments', *PLoS ONE*, 1, e39.

SLATER, M. (2009), 'Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments', *Philos. Trans. R. Soc. B Biol. Sci.* vol. 364, pp. 3549–3557. *https://doi.org/10.1098/rstb.2009.0138*

SLATER, M., and SÁNCHEZ-VIVES, M. V. (2016), 'Enhancing our Lives with Immersive Virtual Reality', *Frontiers in Robotics and AI*, 3:74. *https://doi.org/10.3389/frobt.2016.00074*

SÜTFELD, L. R., GAST, R., KÖNING, P., and 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*(October 2015), pp. 1–13. *https://doi.org/10.3389/fnbeh.2017.00122*

TASSY, S., OULLIER, O., MANCINI, J., and WICKER, B. (2013), 'Discrepancies Between Judgment and Choice of Action in Moral Dilemmas', *Frontiers in Psychology*, vol. 4(MAY), pp. 1–8. *https://doi.org/10.3389/fpsyg.2013.00250*

THOMSON, J. J. (1985), 'The Trolley Problem', *Yale Law Journal*, vol. 94, no. 6, pp. 1395-1415.