

European Portuguese adaptation and validation of dilemmas used to assess moral decision-making

Adaptação e validação para português europeu de dilemas utilizados para avaliar a tomada de decisão moral

Carina Fernandes,^{1,2,3} Ana Ribeiro Gonçalves,¹ Rita Pasion,¹ Fernando Ferreira-Santos,¹ Tiago Oliveira Paiva,^{1,2} Joana Melo e Castro,¹ Fernando Barbosa,¹ Isabel Pavão Martins,³ João Marques-Teixeira¹

Abstract

Objective: To adapt and validate a widely used set of moral dilemmas to European Portuguese, which can be applied to assess decision-making. Moreover, the classical formulation of the dilemmas was compared with a more focused moral probe. Finally, a shorter version of the moral scenarios was tested.

Methods: The Portuguese version of the set of moral dilemmas was tested in 53 individuals from several regions of Portugal. In a second study, an alternative way of questioning on moral dilemmas was tested in 41 participants. Finally, the shorter version of the moral dilemmas was tested in 137 individuals.

Results: Results evidenced no significant differences between English and Portuguese versions. Also, asking whether actions are "morally acceptable" elicited less utilitarian responses than the original question, although without reaching statistical significance. Finally, all tested versions of moral dilemmas exhibited the same pattern of responses, suggesting that the fundamental elements to the moral decision-making were preserved.

Conclusions: We found evidence of cross-cultural validity for moral dilemmas. However, the moral focus might affect utilitarian/deontological judgments.

Keywords: Cognition, emotion, forensic psychiatry, validation studies, psychometric.

Resumo

Objetivos: Validar e adaptar para português europeu um conjunto de dilemas morais amplamente utilizado, que poderá ser aplicado para avaliar a tomada de decisão moral. Além disso, comparou-se a formulação clássica dos dilemas com uma versão alternativa, que incluía uma questão mais focada na moralidade. Finalmente, testou-se uma versão reduzida dos cenários morais.

Métodos: A versão portuguesa dos dilemas morais foi testada em 53 participantes de várias regiões de Portugal, e a versão alternativa foi testada em 41 sujeitos. Finalmente, a versão reduzida dos dilemas morais foi testada numa amostra de 137 participantes.

Resultados: Não houve diferenças significativas entre a versão portuguesa e a inglesa. A pergunta "É moralmente aceitável...?" evocou menos respostas utilitárias que a versão original, embora não tenham sido obtidas diferenças significativas. A versão portuguesa original, a alternativa e a reduzida evidenciaram o mesmo padrão de respostas, sugerindo que os elementos fundamentais para a tomada de decisão moral estão preservados.

Conclusões: Encontramos evidências para a validade intercultural dos dilemas morais. No entanto, o foco moral colocado na questão final pode afetar os juízos utilitários/deontológicos.

Descritores: Emoção, cognição, psiquiatria forense, estudos de validação, psicometria.

¹ Laboratório de Neuropsicofisiologia, Faculdade de Psicologia e de Ciências da Educação, Universidade do Porto, Porto, Portugal. ² Faculdade de Medicina, Universidade do Porto, Porto, Portugal. ³ Laboratório de Estudos de Linguagem, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal.

Submitted Feb 18 2017, accepted for publication Nov 01 2017.

Suggested citation: Fernandes C, Gonçalves AR, Pasion R, Ferreira-Santos F, Paiva TO, Melo e Castro J, et al. European Portuguese adaptation and validation of dilemmas used to assess moral decision-making. *Trends Psychiatry Psychother.* 2018;40(1):38-46. <http://dx.doi.org/10.1590/2237-6089-2017-0022>

Introduction

Moral dilemmas have become a standard methodology to assess and understand human moral psychology.^{1,2} Dilemmas are fictional stories describing two conflicting options. As participants are forced to choose one of these options, the dilemmas allow investigating to which reason precedence is given, and which features are crucial for moral decisions.^{2,3}

Studying moral decision-making through moral dilemmas also allows the inclusion of many variables in their formulation, enabling a more holistic approach and ecological validity under a higher level of experimental control.² Additionally, moral dilemmas offer a valuable tool to study which factors trigger the psychological foundations of human moral cognition, allowing conclusions about real-life moral decision-making.^{1,2,4}

Kohlberg used moral dilemmas in his studies about the moral reasoning development. Following his work, moral psychology was dominated by theories that emphasized the role of reasoning and “higher cognition” in moral judgment.^{5,6} Greene et al. proposed a theory that links utilitarian moral judgment with controlled cognitive processes, and non-utilitarian (deontological) moral judgment with automatic emotional responses.^{1,7,8} To test this theory, Greene et al. proposed a set of classical moral dilemmas that distinguish between “personal” and “impersonal” scenarios.^{1,7,9} In each scenario, the participant must decide whether it is appropriate or not to commit some harm or moral violation to achieve a specific favorable outcome. The personal scenarios involve direct physical contact, as in the case of the footbridge dilemma¹⁰:

A trolley threatens to kill five people. You are standing next to a large stranger on a footbridge that spans the tracks, in between the oncoming trolley and the five people. The only way to save the five people is pushing this stranger off the bridge. Ought you to push this stranger to his death, to stop the trolley from hitting five people and save them?

In turn, the impersonal scenarios comprise more indirect rule violations,¹¹ as in the case of the trolley dilemma¹⁰:

A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save them is pulling a switch that will turn the trolley onto an alternate set of tracks, where it will kill one person instead of five. Ought one to divert the trolley, preventing it from hitting five people, to save them at the expense of one person?

The outcome of these dilemmas is the same: five individuals saved by sacrificing one. However, participants typically consent to pull the switch but decline to push the stranger. The decision to commit the harm to maximize positive consequences is considered a “utilitarian” response. In contrast, a “deontological” response would be the omission of any harmful action, based on the belief that certain duties must be respected.⁸

Several studies demonstrated a noticeable pattern of utilitarian moral judgment in populations with deficits in socioemotional processing, which substantiate the utility of using formal instruments to assess moral judgment. For instance, patients with lesions in the ventromedial prefrontal cortex and patients with frontotemporal dementia demonstrated a more utilitarian judgment.¹¹⁻¹⁴ Adults with psychopathy, who show pronounced deficits in emotional processing and inhibitory control, also exhibited severe disruption of moral behavior, and generally failed to distinguish moral from conventional violations.¹⁵ They had increased utilitarian judgment and diminished empathic concern.^{16,17}

However, results obtained with this methodological approach have been heterogeneous, and the lack of consensus regarding their interpretation is mainly due to the lack of validation of moral dilemmas sets.² Research with this set of moral dilemmas has been lacking of a thorough psychometric evaluation and, to the best of our knowledge, only one previous study reported measures of internal consistency, but only for high-conflict personal dilemmas.¹⁸ In line with this scenario, this study presents results of three experiments conducted to validate the set of moral dilemmas proposed by Greene et al., as well as to test methodological improvements.¹ Specifically, the first experiment provides a cross-cultural validation of moral scenarios, by adapting them to European Portuguese and further assessing their reliability and internal consistency. In the second experiment, following Greene’s personal suggestion, the original question (“Is it appropriate...?”) was replaced with the question “Is it morally acceptable...?”, with the aim of increasing the focus on moral judgment, by eliciting more reflection about the right and wrong according to a moral standpoint. In the third experiment, a shorter version of each scenario was developed by reducing the number of words without changing the meaning of the text. This reduction aimed to decrease the cognitive load of each scenario, by reducing the details of the fictional stories. Importantly, the internal consistency of the adapted and shortened versions of moral dilemmas is reported in the three experiments.

Experiment 1 – Adaptation and validation of moral dilemmas to European Portuguese

Greene et al. introduced this set of moral dilemmas in cognitive neuroscience as an attempt to deepen our understanding of the role of emotion in moral judgment.^{1,3} However, the results obtained with this methodological approach have been heterogeneous, and there is a lack of consensus regarding how to interpret them. Moreover, there is also a lack of cross-validation, which precludes the understanding of cultural influences on the pattern of responses.

Although cultural differences in moral decisions are a major theme in social psychology, the effect of culture in moral psychology is still far from being understood, since research in this field is largely restricted to English-speaking populations.¹⁹ However, even though some moral features are universal (e.g., “it is wrong to produce harm without any kind of justification”), morality highly varies across cultures, depending on specific norms, practices or values.¹⁹ According to Kohlberg & Candee, sociocultural context guides people’s moral evaluation and decision-making.²⁰ Two previous studies reported differences in the acceptance of harm between Colombian and Spanish participants, as well as in the neural correlates underlying moral decision-making between Chinese and Western participants.^{21,22} These findings reinforce that morality cannot be completely understood without considering sociocultural factors, supporting the need for cross-cultural studies in moral dilemmas.

In line with these results, the first experiment of the present study aimed to cross-culturally validate the moral dilemmas proposed by Greene et al., adapting each scenario to European Portuguese. Moreover, we aim to provide Portuguese normative values and assess the reliability and internal consistency of this instrument.

Methods

Participants

This study was approved by the local research ethics committee, and all participants gave informed consent. Inclusion criteria were being older than 18 years and having Portuguese nationality and fluency.

The Portuguese version of the moral dilemmas was tested with 53 individuals (39 women), recruited via convenience sampling, following advertisement of the study on social networks. This form of dissemination resulted in a sample of participants from the north (81%), center (8%) and south (2%) of Portugal.

Moreover, one participant (2%) was from the Azores, and three participants (6%) did not report their location of residence. Participants were between 18 and 52 years old (mean [M] = 23.6, standard deviation [SD] = 7.27) and reported 12 to 19 years of education (M = 14.9, SD = 2.66).

Two participants were excluded from the initial sample (n = 58) for having other nationalities, and three were excluded for being detected as outliers (see Data analysis below for information on outlier detection).

Materials

The set of moral dilemmas is composed of 40 scenarios, divided into 12 high-conflict personal scenarios (that elicited lower rates of agreement), nine low-conflict personal scenarios (that elicited near-perfect agreement) and 19 impersonal scenarios.^{1,13} Each scenario asks whether it is appropriate to commit a harm or transgression to achieve a more favorable outcome.

Translation and adaptation

After securing permission with the original author, the translation of the dilemmas to Portuguese was conducted according to procedures intended to maximize semantic (i.e., content and phrasing of the items) and normative equivalence (i.e., conformity between the instrument and the cultural rules of the target culture) between both versions.²³ This process comprised the following stages: 1) translation – two independent translations were carried out and reviewed to reach a consensual version, which was reviewed by a third researcher to optimize the translation; 2) pilot-administration with thinking out loud – the consensual version was administered to six participants, who discussed their understanding of each item and gave suggestions to improve the text, their answers, and the reasoning (as a result, 11 items were clarified); 3) back-translation – the Portuguese version was back-translated to English by a fourth researcher; 4) comparison with the original version – the back-translated version was approved by the original author. All researchers involved in translation and back-translation were proficient in both languages and experienced in moral judgment and social cognition research fields.²⁴ The final version of the Portuguese set of the moral dilemmas is available as Online-Only Supplementary Material.

Procedures

The Portuguese moral dilemmas were presented in random order to all participants in an online questionnaire (Qualtrics, 2016, Provo, UT, USA). The participants read and responded at their own pace,

choosing the option “yes” or “no” to the question “Is it appropriate...?”. After answering, the next scenario was displayed. All participants answered to all the dilemmas, as responding was mandatory to move to the next question.

Data analysis

The internal consistency of each category of dilemmas (high- and low-conflict personal dilemmas and impersonal dilemmas) was assessed using the Kuder-Richardson Formula 20 (KR20),²⁵ which is a special case of Cronbach’s alpha used for binary items. The results obtained with our data collection were compared to the results obtained in the original study. Statistically significant differences between the percentage of utilitarian responses to the Portuguese and the original versions were independently assessed for each category of dilemmas using the Mann-Whitney U test, since the normality (Shapiro-Wilk’s test) and homogeneity of variances (Levene’s test) assumptions were not met. The percentage of utilitarian responses

for both versions was also compared by dilemma, using the z-test for proportions. The Bonferroni correction was applied to control for multiple comparisons, to retain a family-wise statistical significance level of $p < 0.05$. Participants who selected a category that occurred with less than 5% of relative frequency were excluded as categorical outliers ($n = 3$) before computing KR20 and performing comparisons. All statistical tests were performed using Statistica version 13.0.

Results

The internal consistency was good for high-conflict personal dilemmas ($\rho = 0.84$), poor for impersonal dilemmas ($\rho = 0.53$), and unacceptable for low-conflict personal dilemmas ($\rho = 0.47$).

Nevertheless, the comparison between Portuguese and original percentages of utilitarian responses did not show significant differences for any category of moral dilemmas (Table 1). Additionally, no significant differences were found when the two versions were compared by dilemma.

Table 1 - Percentage of utilitarian responses to the Portuguese (original and modified questions) and original versions¹ of the moral dilemmas

Dilemma	High-conflict personal dilemmas			Dilemma	Low-conflict personal dilemmas			Dilemma	Impersonal dilemmas		
	Original	Portuguese	Modified question		Original	Portuguese	Modified question		Original	Portuguese	Modified question
Crying baby	60	38	12	Architect	1	4	5	Donation	62	70	66
Euthanasia	63	77	61	Country road	3	2	0	Environmental policy A1	75	72	80
Footbridge	21	38	22	Plane crash	2	25	5	Environmental policy A2	84	89	80
Lawrence of Arabia	82	68	51	Grandson	1	0	2	Environmental policy B1	24	15	17
Modified bomb	90	75	78	Hard times	9	2	0	Environmental policy B2	25	11	24
Modified lifeboat	71	74	49	Hired rapist	4	0	0	Eyes	3	0	7
Modified safari	22	60	34	Infanticide	5	0	0	Five-for-seven trolley	5	2	24
Sacrifice	51	26	17	Smother for dollars	7	9	2	Guarded speedboat	91	85	93
Sophie’s choice	62	38	29	Transplant	12	4	0	Illegal lunch	9	4	10
Submarine	91	81	66					Lost wallet	16	17	17
Vaccine test	79	64	46					Resume	9	11	10
Vitamins	35	32	22					Sculpture	96	92	95
								Speedboat	99	87	90
								Standard trolley	82	68	49
								Standard fumes	76	64	44
								Stock tip	12	11	17
								Taxes	24	23	27
								Three-for-seven fumes	10	4	12
								Vaccine policy	85	85	85
Mean (SD)	60.6 (24.3)	56.0 (20.1)	40.6 (21.1)		4.89 (3.79)	5.11 (8.00)	1.56 (2.13)		46.7 (37.0)	42.6 (36.6)	44.6 (33.2)
Mann-Whitney U test		64.5	42.0		30.5	29.5			161.0	204.5	
p		0.67	0.089		0.39	0.34			0.58	0.49	

SD = standard deviation.

Experiment 2 – “Is it appropriate...?” vs. “Is it morally acceptable...?”

The formulation of the final question of moral dilemmas, designed to elicit moral judgment, has given rise to some controversy, since the type of question influences the participant’s moral decision.^{2,26,27} For instance, in a study that used four question formats (wrong, inappropriate, forbidden, and blameworthy), participants judged moral transgressions more severely when the words “wrong” or “inappropriate” were part of the formulation.²⁸ Similarly, distinct response patterns were found with the questions “Is it wrong to...?” and “Would you...?”. The later formulation resulted in faster reaction times in moral than in non-moral scenarios, while the question “Is it wrong to...?” did not result in differences between both conditions.²⁹

In this set of moral dilemmas, participants are typically questioned with an appropriate–inappropriate dichotomy.¹ However, in some studies, participants may decide instead if they would choose to carry out the depicted action in a yes–no dichotomy, a decision that may be preceded by another dichotomous question asking whether the action is right or wrong.^{11,29} Another alternative is to ask the participants to indicate their judgment in a permissible–forbidden scale.³⁰ However, it is not the same to ask whether an action is permissible or appropriate.³ While the first term relates to the legal permissibility of the action, “appropriate” suggests whether the participant finds the action obligatory to the situation. Similarly, the right–wrong dichotomy hints the legal permissibility of the action, while the “would you...?” question does not give information about the judgment of the action (participants could decide to do what they take to be wrong).³ Moreover, the typical formulation (“appropriate”) is ambiguous and may be understood as related to conventional instead moral rules.²⁹

In line with these findings, the formulation of the question should be a matter of research, and this experiment provides a further test of the influence of question formulation on moral decision-making. Different questions may entail differential moral judgments, which would interfere with the development of a unified theory of moral judgment.³

Considering that dilemmas are used to test moral judgment, with our modification (“Is it morally acceptable” instead of “Is it appropriate”) we intend to make the question less ambiguous and increase the focus on moral judgment, by eliciting more reflection about right and wrong according to a moral standpoint. To the best of our knowledge, this is the first study that tests this type of formulation, since previous studies

have only used this probe in a Likert scale (e.g., “to what extent is the action morally acceptable?”), but not in a dichotomy.³⁰ However, dichotomous questions are typically preferred, since they force participants to choose one of the options, allowing to investigate to which reason precedence is given in moral decisions. We hypothesized that the percentage of utilitarian responses would decrease with the increasing focus on moral decision-making, which may raise controlled cognitive processes.

Methods

Participants

We tested an alternative way of questioning on moral dilemmas in 41 participants (22 women), aged 19 to 65 years ($M = 32.0$ years, $SD = 10.0$) and with 9 to 22 years of education ($M = 16.4$ years, $SD = 3.50$). All participants gave informed consent and were recruited by e-mail, using a convenience sampling approach. This form of recruitment was selected to ensure that participants who responded to this version of moral dilemmas would not respond to the two alternative forms (Experiments 1 and 3) that were going on simultaneously. Participants were recruited from the north (39%), center (56%) and south (5%) of Portugal. One participant with another nationality was previously excluded from the sample.

Procedures

The procedure was replicated from Experiment 1, changing the way of questioning in the Portuguese version of the moral dilemmas to “Is it morally acceptable...?”, rather than the original “Is it appropriate...?”. This modification may alter how participants approach the task and makes it more appropriate for the aim of the moral dilemmas.³¹ Inter-method reliability was tested by assessing the differences in the percentage of utilitarian responses obtained in Experiment 1.

Results

Internal consistency, measured by the KR20 test, was good for high-conflict personal dilemmas ($\rho = 0.86$), poor for impersonal dilemmas ($\rho = 0.50$), and unacceptable for low-conflict personal dilemmas ($\rho = 0.47$).

The percentages of utilitarian responses to the Portuguese version of moral dilemmas, with the original and modified questions, are presented in Table 1. The comparison between the two ways of questioning did not evidence statistical significance for any category of moral dilemmas.

Experiment 3 – Validation of a shorter version of the moral dilemmas

This set of moral dilemmas has been repeatedly used in moral judgment research, both in healthy and in pathological populations, to understand the underpinnings of morality and how specific conditions affect moral judgment. As stated in a recent review, moral dilemmas imply a lot of reading, and keeping them as short as possible prevents the participants from getting tired. Moreover, the cognitive load of each scenario can be decreased by reducing the details of the fictional stories.³

According to this evidence, a shorter version of each scenario was developed by reducing the number of words. The major challenge of this experiment was to keep the meaning of the text intact, as word framing is known to affect decision-making.³² Under certain circumstances, participants violate the principle of invariance, which postulates that moral choices must depend on the situation and not on the description. For instance, participants have different preferences over equivalent situations because of the way they are described, preferring a situation in which half of the population is saved to one in which half the population dies in an epidemic.³

Thereby, the aim of the present experiment was to create and validate a shorter version of this set of moral dilemmas, without changing the meaning of each scenario. We hypothesized that both versions would reach the same pattern of responses, evidencing a preserved focus on moral judgment despite a reduction in cognitive demands of each dilemma. This shorter version may be particularly useful to investigate moral judgments in clinical populations or for use with research techniques that require fast events, such as event-related designs for studies with electroencephalography (EEG) or functional magnetic resonance imaging (fMRI).

Methods

Participants

This experiment included a sample of 137 individuals (114 women), aged 18 to 62 years ($M = 22.5$ years, $SD = 7.16$) and with 9 to 22 years of education ($M = 14.2$ years, $SD = 2.53$). All participants gave electronic informed consent and were recruited by an e-mail sent to all students of the local university. Participants were from the north (83%) and center (9%) of mainland Portugal, as well as from the Azores (1%) and Madeira (1%). Eight participants (6%) did not report their location of residence.

Shorter version of moral dilemmas

The dilemmas tested in the Experiment 2 (Portuguese version of moral dilemmas with the modified question) were abbreviated by shortening the story of each scenario. To the reduction process, two researchers abbreviated each dilemma (eliminating the information considered unessential to moral judgment), and a final consensual version was developed. Subsequently, both the long and the short versions of the dilemmas were administered, in a balanced order, to eight participants, who discussed their understanding of both versions. Participants consensually reported that the shorter version was easier to answer, due to fewer details to memorize. On average, each scenario was reduced in 29.7% of total words ($SD = 7.37$, range: 16.5-45.8%). The shorter version of the moral dilemmas is also available as Online-Only Supplementary Material.

Procedures

Since the aim of the current experiment was to test the convergent validity of the short version of moral dilemmas, in order to guarantee that the moral content of each scenario was preserved despite the length reduction, a within-subjects design was adopted. This design is recommended to test convergent validity, since it allows to test the correlation of responses between both measures while reducing individual variability.^{33,34} Thus, the long and short dilemmas were presented in random order in an online questionnaire (LimeSurvey 2.14, LimeSurvey Project Hamburg, Germany), to prevent carryover effects. Participants read and responded at their own pace, choosing the option yes or no to the question "Is it morally acceptable to...?". After answering, the next scenario was displayed. All participants answered to all the dilemmas, as responding was mandatory to move to the next question. Data analysis was replicated from the previous experiments, but the phi coefficient, a measure of correlation for two binary variables, was calculated between the two versions.³⁵

Results

Internal consistency was equivalent for both versions of dilemmas: good for high-conflict personal dilemmas (long and short versions: $\rho = 0.85$), and unacceptable for impersonal (long version: $\rho = 0.40$; short version: $\rho = 0.41$) and low-conflict personal dilemmas (long version: $\rho = 0.29$; short version: $\rho = 0.28$).

No significant differences were found between the utilitarian responses to both versions of moral dilemmas (Table 2), and the phi coefficient revealed robust associations in every case ($r_\phi = 0.97$ to high-conflict personal dilemmas; $r_\phi = 0.85$ to impersonal dilemmas;

$r_{\phi} = 0.85$ to low-conflict personal dilemmas). Moreover, significant differences were not found when the two versions were compared by dilemma (Table 2).

Discussion

Moral dilemmas have been extensively studied in the field of moral psychology, and their use has much to contribute to our understanding of moral decision-making.^{1,3} However, methodological critiques highlight the need for research on this instrument, including psychometric evaluation, to enable an accurate interpretation of the results.³ Moreover, it is important to test the dilemmas across different countries, languages, and cultures, to understand the effects of culture on moral judgment. In line with these issues, the aim of the first study was to validate moral dilemmas to European Portuguese and to assess its reliability in a sample of Portuguese adults.¹ Moreover, two experimental improvements were also tested.

The results of Experiment 1 evidenced that the Portuguese version of the dilemmas is quite satisfactory, and the comparison of our results with the ones from the original version did not show significant differences in the percentage of utilitarian responses. This lack of differences confirms the results of a previous cross-cultural validation of a Spanish version of the same instrument.³⁵ This evidence supports the cross-cultural validity of the dilemmas despite cultural differences between both countries (i.e., more individualistic vs. more communitarian) and may suggest the existence of universal moral judgments, independent of cultural or educational background.^{1,35} However, as these findings contrast with results of other cross-cultural validations, further studies assessing the effects of each culture on moral judgment are needed.^{17,19}

The results obtained with the internal consistency assessment of the Portuguese set was good for high-conflict personal dilemmas ($\rho = 0.84$), suggesting that this type of stimuli is reliable to assess moral judgment in Portuguese individuals. Despite a lack of

Table 2 - Percentage of utilitarian responses to the Portuguese long and short versions of moral dilemmas

Dilemma	High-conflict personal dilemmas		Dilemma	Low-conflict personal dilemmas		Dilemma	Impersonal dilemmas	
	Short	Long		Short	Long		Short	Long
Crying baby	28	25	Architect	1	1	Donation	64	64
Euthanasia	72	74	Country road	0	1	Environmental policy A1	82	82
Footbridge	20	22	Plane crash	20	18	Environmental policy A2	81	80
Lawrence of Arabia	71	75	Grandson	3	0	Environmental policy B1	15	14
Modified bomb	82	80	Hard times	1	1	Environmental policy B2	16	18
Modified lifeboat	65	68	Hired rapist	0	0	Eyes	2	3
Modified safari	65	65	Infanticide	1	1	Five-for-seven trolley	12	8
Sacrifice	26	29	Smother for dollars	1	1	Guarded speedboat	93	94
Sophie's choice	36	35	Transplant	5	5	Illegal lunch	8	8
Submarine	78	79				Lost wallet	13	12
Vaccine test	66	67				Resume	8	5
Vitamins	22	20				Sculpture	99	99
						Speedboat	96	96
						Standard trolley	67	72
						Standard fumes	69	69
						Stock tip	15	15
						Taxes	18	18
						Three-for-seven fumes	17	14
						Vaccine policy	80	83
Mean (SD)	52.5 (24.0)	53.3 (24.6)		3.49 (6.57)	3.00 (5.66)		44.9 (36.4)	44.9 (37.4)
Mann-Whitney <i>U</i> test		68.0			39.5			179.5
<i>p</i>		0.84			0.80			0.95

SD = standard deviation.

psychometric evaluation of this instrument, a previous study focused only on high-conflict personal dilemmas reached a similar reliability ($\rho = 0.82$), calculated through Tarkkonen's ρ .¹⁸

The poor and very low internal consistency found for impersonal and low-conflict personal dilemmas was consistent along our three experiments (conducted with different samples), which reveals their reduced reliability in assessing moral decision-making. To the best of our knowledge, no previous studies have reported the internal consistency of these subsets of moral dilemmas, but this result was expected. According to Greene et al., high-conflict dilemmas share a similar structure, in which one person can be harmed to achieve a greater benefit.¹ Conversely, impersonal and low-conflict moral dilemmas have different structures and involve no clear conflict between utilitarian and non-utilitarian moral principles. Thus, recent studies have focused only on high-conflict dilemmas.¹⁸ However, impersonal and low-conflict dilemmas can also be used as "fillers" to mix in with high-conflict dilemmas and to test whether participants are responding by chance. Alternatively, as dilemmas can be modified to involve different grades of conflict, these scenarios can be adapted to increase this low internal consistency.²

Concerning Experiment 2, the adapted version of the moral dilemmas with a new way of questioning did not reveal significant differences in the responses to each category of dilemmas. However, the results revealed a lower percentage of utilitarian responses to the modified questions. In the case of personal dilemmas, this reduction was consistent across all scenarios, except one, and the lack of significant statistical differences may be due to lack of statistical power, which is a limitation of the present study. A future high-powered replication study may help resolve this question.

As the question "Is it appropriate...?" may downplay the moral relevance of the scenarios²² and was reported as ambiguous, we adopted the new way of questioning in Experiment 3, in which a shorter version of the moral dilemmas was proposed, aiming to decrease the cognitive load by reducing the details of each moral scenario.²⁸ This methodological modification attained successful results: equivalent internal consistency for both versions of the dilemmas, no significant differences between the responses elicited, and robust associations between both personal and impersonal scenarios. Importantly, the aim of Experiment 3 was to test the convergent validity of the short version of moral dilemmas, to guarantee that the moral content of each scenario was preserved despite a reduction of length. According to this purpose, a within-subjects design was adopted, which does not allow to make substantive

claims about moral judgment based on the data. They are only informative of the convergent validity of the short version of moral dilemmas.

The present study has some limitations that may be addressed in further investigations. First, the questionnaire was applied online, limiting the control of experimental settings and confounding variables. Furthermore, we did not include a measure of socioeconomic status, which may influence moral judgments. Also, the study included a sample with a high educational level, limiting generalization to different educational levels. Finally, we argue that our short version may reduce the cognitive load of each scenario, but this hypothesis was not directly assessed.

Furthermore, other characteristics of moral scenarios should be tested, such as instructions, presentation format, similar word counting between dilemmas, and order of presentation, among others.³ Control of these aspects will contribute to increase the methodological homogeneity of this instrument and its application.

In conclusion, we adapted and validated a widely used set of moral dilemmas to European Portuguese, and we have made it available as Online-Only Supplementary Material. This instrument will allow research into moral decision-making in Portuguese-speaking cultures for the first time, in both healthy and pathological populations. Moreover, we tested a new way of questioning, which is a matter of research concern.³ This new question formulation was tested for the first time, and it intended to decrease the ambiguity of the traditional formulation. Finally, we tested a reduced version of the moral scenarios, also available as Online-Only Supplementary Material. This shorter version will be faster to apply, will prevent the participants' fatigue, and will decrease the cognitive load of each scenario. Importantly, this aim was reached without changing the meaning of each scenario.

This translated version will be particularly useful to investigate moral judgments in clinical Portuguese-speaking populations. Nevertheless, this study did not include clinical samples, and the normative values reached cannot be used as reference to clinical participants. However, as previous findings have evidenced that pathologies related to deficits in emotional processing and inhibitory control exhibited an abnormal pattern of utilitarian moral judgment, further studies must extend moral judgment research to pathologies with marked deficits in emotional processing and inhibitory control, such as personality and major psychiatric disorders.¹¹⁻¹³ The shorter version of moral dilemmas will be particularly useful for simultaneous application with other research techniques that require fast events, such as event-related designs for studies with EEG or

fMRI. These experiments will significantly contribute to investigate the neural mechanisms underlying moral decision-making.

Acknowledgements

This research was supported by a grant from Fundação Bial. Carina Fernandes and Tiago O. Paiva were supported by a doctoral fellowship from Fundação para a Ciência e a Tecnologia (SFRH/BD/112101/2015 and SFRH/BD/93831/2013, respectively).

The authors would like to thank Joshua Greene for his support and permission to adapt the moral dilemmas. The authors are also grateful to Inês Mares for the back-translation of the Portuguese version.

Disclosure

No conflicts of interest declared concerning the publication of this article.

References

- Greene JD, Morelli SA, Lowenberg K, Nystrom LE, Cohen JD. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*. 2008;107:1144-54.
- Christensen JF, Flexas A, Calabrese M, Gut NK, Gomila A. Moral judgment reloaded: a moral dilemma validation study. *Front Psychol*. 2014;5:607.
- Christensen JF, Gomila A. Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neurosci Biobehav Rev*. 2012;36:1249-64.
- Hauser M., Cushman F, Young L, Jin R, Mikhail J. A dissociation between moral judgments and justifications. *Mind Lang*. 2007;22:1-21.
- Kohlberg L. Development of moral character and moral ideology. In: Hoffman ML, Hoffman LW, editors. *Review of child development research*. New York: Russell Sage Foundation; 1964. p. 383-432.
- Kohlberg L. Stage and sequence: The cognitive-developmental approach to socialization. In: Goslin DA, editor. *Handbook of socialization theory and research*. Chicago: Rand McNally; 1969. p. 347-480.
- Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. *Science*. 2001;293:2105-8.
- Greene J, Haidt J. How (and where) does moral judgment work? *Trends Cogn Sci*. 2002;6:517-23.
- Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD. The neural basis of cognitive conflict and control in moral judgment. *Neuron*. 2004;44:389-400.
- Thomson JJ, Parent W. *Rights, restitution, and risk: essays in moral theory*. Harvard: University Press; 1986.
- Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, et al. Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*. 2007;446:908-11.
- Koenigs E, Muccioli M, Ladavas E, di Pellegrino G. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Soc Cogn Affect Neurosci*. 2007;2:84-92.
- Moretto G, Ladavas E, Mattioli F, di Pellegrino G. A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *J Cogn Neurosci*. 2010;22:1888-99.
- Mendez MF, Anderson E, Shapira JS. An investigation of moral judgment in frontotemporal dementia. *Cogn Behav Neurol*. 2005;18:193-7.
- Koenigs M, Kruepke M, Zeier J, Newman JP. Utilitarian moral judgment in psychopathy. *Soc Cogn Affect Neurosci*. 2012;7:708-14.
- Cima M, Tonnaer F, Hauser MD. Psychopaths know right from wrong but don't care. *Soc Cogn Affect Neurosci*. 2010;5:59-67.
- Kahane G, Everett JAC, Earp BD, Farias M, Savulescu J. 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*. 2015;134:193-209.
- Laakasuo M, Sundvall J. Are utilitarian/deontological preferences unidimensional? *Front Psychol*. 2016;7:1-18.
- Prinz J. Is morality innate? In: Sinnott-Armstrong W, editor. *Moral psychology*. Cambridge: MIT Press; 2008. Vol. 1: The evolution of morality: adaptations and innateness. p. 367-406.
- Kohlberg L, Candee D. The relationship of moral judgment to moral action. In: Kurtines WM, Gewirtz JL, editors. *Morality, moral behavior, and moral development*. New York: Wiley; 1984.
- Olivera-La Rosa A, Corradi G, Villacampa J, Martí-Vilar M, Arango OE, Rosselló J. Effects of suboptimally presented erotic pictures on moral judgments: a cross-cultural comparison. *PLoS ONE*. 2016;11:e0158690.
- Wang Y, Deng Y, Sui D, Tang YY. Neural correlates of cultural differences in moral decision making: a combined ERP and sLORETA study. *Neuroreport*. 2014;25:110-6.
- Behling O, Law KS. *Translating questionnaires and other research instruments: problems and solutions*. Thousand Oaks: Sage; 2000.
- Almeida PR, Seixas MJ, Ferreira-Santos F, Vieira JB, Paiva TO, Moreira PS, et al. Empathic, moral and antisocial outcomes associated with distinct components of psychopathy in healthy individuals: a Triarchic model approach. *Pers Individ Dif*. 2015;85:205-11.
- Kuder GF, Richardson MW. The theory of the estimation of test reliability. *Psychometrika*. 1937;2:151-60.
- Bouwmeester S, Verkoeijen PPJL, Aczel B, Barbosa F, Bègue L, Brañas-Garza P, et al. Registered replication report: Rand, Greene, and Nowak (2012). *Perspect Psychol Sci*. 2017;12:527-42.
- O'Hara RE, Sinnott-Armstrong W, Sinnott-Armstrong NA. Wording effects in moral judgments. *Judgm Decis Mak*. 2010;5:547-54.
- Borg JS, Hynes C, Van Horn J, Grafton S, Sinnott-Armstrong W. Consequences, action, and intention as factors in moral judgments: an fMRI investigation. *J Cogn Neurosci*. 2006;18:803-17.
- Cushman F. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*. 2008;108:353-80.
- Cikara M, Farnsworth RM, Harris LT, Fiske ST. On the wrong side of the trolley track: neural correlates of relative social valuation. *Soc Cogn Affect Neurosci*. 2010;5:404-13.
- Ferreira-Santos F, Sousa P, Mauro C, Paiva TO, Pereira MR. Extending the moral-conventional task: Conceptual and methodological issues in the study of the judgment of moral transgressions. 20th Meeting European Society for Philosophy and Psychology; August 28-31 2012; London. p. 146.
- Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science*. 1981;211:453-8.
- Guilford J. *Psychometric methods*. New York: McGraw-Hill; 1936.
- Groth-Marnat. *Handbook of psychological assessment*. New Jersey: John Wiley & Sons; 2009.
- Carmona-Perera M, Vilar-Lopez R, Perez-Garcia M, Verdejo-Garcia A. Using moral dilemmas to characterize social decision-making. *Clin Neuropsychiatry*. 2013;10:95-102.

Correspondence:

Carina Fernandes
Laboratório de Neuropsicofisiologia
Faculdade de Psicologia e de Ciências da Educação
Universidade do Porto
Rua Alfredo Allen
4200-135 - Porto - Portugal
Tel.: +351 226 079 700.
E-mail: carinafernandes@fpce.up.pt