

What makes things funny?  
Evidence for the benign-violation theory of humour

Sarah Westwood  
Department of Psychology  
University of Cape Town

Author's Note:

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Supervisor: Prof. Johann Louw

Word Count:

Abstract: 286

Main Body: 9532



### **Plagiarism Declaration**

1. I know that Plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the American Psychological Association formatting for citation and referencing. Each significant contribution to, and quotation in, this project from the work or works, of other people has been attributed, cited and referenced.
3. This project is my own work.
4. I have not allowed, and will not allow anyone to copy my work with the intention of passing it off as his or her own work.

Signature:

Date:

### **Acknowledgements**

My sincere thanks to my supervisor, Professor Johann Louw, for his wise guidance and consistent availability (and for laughing at my jokes!).

Thank you to Dr Pedro Wolf and Michelle Hoogenhout for their help in solving methodological and statistical conundrums.

*Soli Deo Gloria*

### Abstract

Research has demonstrated that experiencing amusement is commonly associated with physical, emotional, social and psychological benefits. As such it is valuable to understand what it is that makes certain things funny. Peter McGraw and Caleb Warren suggested that the answer to this question is that a stimulus must be perceived as a benign-violation. That is, the stimulus must be appraised: a) to be wrong, b) to be non-threatening, and c) these appraisals must occur simultaneously. This study had two aims. The first aim was to demonstrate that, in agreement with McGraw and Warren's (2010) research, that people find moral violations amusing when they are appraised to be both acceptable and unacceptable. The second aim of this study was to test whether the benign-violation hypothesis is still confirmed when more appropriate and robust methods of data analysis are employed. The sample in this study was made up of conveniently sampled university students ( $N = 204$ ) who were assigned to one of eight groups as defined by the  $2 \times 2 \times 2 \times 4$  mixed model design. Participants were required to complete an online survey in which they responded to a number of (potentially justifiable) moral violation and control scenarios. Participants' responses were then scored in three different ways and entered into a mixed ANOVA. The results of this study indicate that participants were more likely to be amused by moral violation versions of scenarios than by comparable control versions. This effect was significant in all three analyses ( $p < .001$ ). Additionally, the size of the effect of the version factor increased substantially as the methods of data analysis were appropriately altered (Range  $\eta^2_p$ : .37 - .52). This study suggests that humour is a socially adaptive response to benign-violations that people regularly encounter.

*Key words:* benign-violation; humour; amusement; moral violation; pleasure.

## Introduction

Humour is ubiquitous. It is present in all human societies, and has also been observed in animal populations (Gervais & Wilson, 2005). That is not to say that every culture has the same sense of humour or that all people will laugh at the same things, but rather that in every culture humour that elicits positive emotions is valued (Billig, 2005). Humour is defined as a form of play within a matrix of the social context, cognitive process, and an emotional response that is often expressed through laughter (Martin, 2007). The pervasiveness of humour in its multiple forms suggests that it must have a positive effect on both the source and receiver of the humorous exchange. Indeed, humour has been described as “a modern human addiction” (p. 1) suggesting that it is perceived as essential to individual survival (Hurley, Dennett & Adams, 2011).

Being humorous is a desirable personality trait that confers social, physical and psychological benefits to an individual (Martin, 2007). Humour is an essential component of social interaction (Wyer & Collins, 1992). It influences the friends, dates and mates we choose, softens criticism, and establishes and maintains social boundaries (Billig, 2005; Gervais & Wilson, 2005; Martin, 2007). Research also suggests that humorous experiences can improve various aspects of physical health including pain tolerance, immunity and blood pressure (Martin, 2002). Humour can also alleviate undesirable psychological states such as anxiety and grief (Gervais & Wilson, 2005). Many of these changes come about because the perception of humour can cause actual changes in various bodily systems, the induction of desirable affective states, the development of a cognitive buffer against stress, and an increase in the quality and quantity of social support (Martin, 2002).

There is no doubt that humour is pervasive and everyone experiences humour regularly. Even though humour is so familiar and it may often seem that we intrinsically understand what it is, there are nevertheless many interesting psychological questions about humour that require empirical investigation. Some of these questions include “What are the mental processes involved in ‘getting a joke’?”, “Why is humour so enjoyable?”, and “How does a sense of humour develop in children?” (Martin, 2007). This research endeavoured to answer the broad question “what makes things funny?” or in other words “what criteria must a stimulus meet for it to be considered amusing?”. In a previous attempt to answer this question Peter McGraw and colleagues (2010) published a theory of humour, which they call the benign-violation theory of humour. Their initial study was, however, methodologically flawed, and as such the validity and reliability of their results have been called into question.

The research presented in this paper is a methodological revision of the initial study conducted by McGraw and Warren (2010) to determine whether or not the benign-violation theory is still a valid explanation as to what makes something funny.

### **Theoretical developments in the study of humour**

There are many different types of humour including jokes, sarcasm, irony, puns, satire, slapstick, and teasing. There are also various mediums through which humour is conveyed including verbal (e.g. joke telling), visual (e.g. caricatures) and physical humour (e.g. tickling). As a result of the vast variability in the types of humour we express and experience, an overarching theoretical explanation of mechanisms required to produce humour is not easy to articulate. Many researchers have tried to articulate theories of humour production and elicitation. These theories fall into two main groups: Domain-specific theories and general humour theories. Domain-specific theories address only certain types of humour and are thus incapable of providing an answer to the more general question of ‘what makes things funny?’. General humour theories try to delineate the underlying cause of multiple forms of humour. Notable general theories of humour include tension release (Freud, 1908), incongruity (Suls, 1972 as cited in McGraw & Warren, 2010) and superiority (Gruner, 1997 as cited in Martin, 2007). The disadvantage of these theories is that the presence of the supposed antecedent of humour can also be an antecedent of tragedy. For example, accidentally shooting ones child may indicate the release of pent-up tension, incongruence and the assertion of superiority, but it is in no way humorous. In his comprehensive review of existing humour theories Martin (2007) concluded that no general theory of humour was sufficiently able to delineate the antecedents of amusement. A more recent review of the literature revealed that there are two conditions that seem to predict humour arousal. They are that a situation must be a violation, non-threatening, and these conditions must occur simultaneously (McGraw & Warren, 2010).

### **The benign-violation theory of humour**

Based on their analysis of the literature, particularly research conducted by Veatch (1998), McGraw and Warren (2010) proposed the benign-violation theory of humour. This theory states that in order for a situation to elicit a humorous response it “must be appraised as a violation, appraised as benign, and these two appraisals must occur simultaneously” (McGraw & Warren, 2010, p. 1142).

A situation is appraised as a violation when it threatens the way we think things ought to be (McGraw & Warren, 2010). Violations of this kind include apparent threats, breaches of norms (social, moral, linguistic etc.), or taboo content. A benign situation is one that is non-

serious, safe and playful; it is essentially the opposite of a violation. Neither the perception of a benign situation nor a violation of some form will induce a humorous response. For example, watching a friend walk down a flight of stairs is benign and not funny. Watching a friend fall down a flight of stairs and be badly hurt is a violation and not funny. If, however, a friend falls down a flight of stairs and is unhurt it is funny. This example demonstrates that a humorous response is achieved when a situation is both benign and a violation. The process of holding two contradictory ideas at the same time, as in the case of appraising a situation as benign and a violation, is called simultaneity and is necessary to elicit humour.

### **Preliminary evidence for the benign-violation theory of humour**

A series of experiments were undertaken in order to provide support for the benign-violation theory (McGraw & Warren, 2010; McGraw, Warren, Williams & Leonard, 2012). These studies tested three different ways of making a moral violation seem more benign, and therefore more humorous. In their first paper McGraw and Warren (2010) demonstrated that moral violations become more benign if a) there is another norm that suggests that the violation of the moral norm is acceptable, b) the violated norm is not something that the individual is strongly committed to, and c) the violation is spatially distant. The theory that spatial distance from a violation can increase or decrease the threat a violation poses is part of a broader hypothesis that ‘psychological distance’ can influence whether a violation is benign or malign by causing the violation to be construed in an abstract or concrete manner (Liberman & Trope, 2008). There are four types of psychological distance – spatial (e.g. a kilometre is more distant than a metre), temporal (a month is more distant than a week), hypothetical (an imagined situation is more distant than a real one) and social (a stranger is more distant than a romantic partner) (Liberman & Trope, 2008). The perception of humour in moral violations is significantly affected by variations in all four types of psychological distance (McGraw et al., 2012). Furthermore, the level of severity of a violation moderates the effect that psychological distance has on humour (See figure 1). For example, stubbing one's toe is a “mild mishap” that is more funny if it happened yesterday as opposed to five years ago. Conversely, being hit by a car is a “tragedy” and is likely to be perceived as more humorous if it happened five years ago as opposed to yesterday.



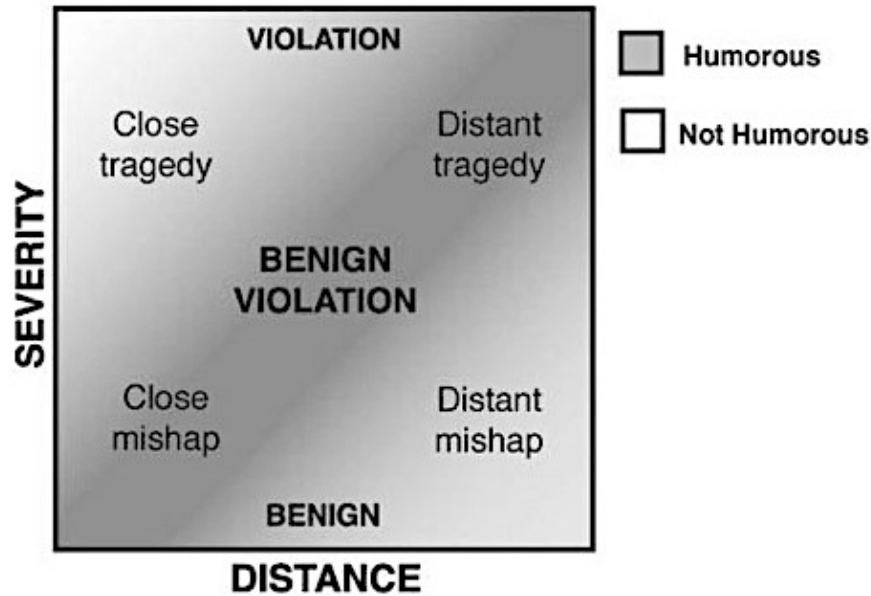


Figure 1. Benign-Violation account of the perception of humour. Violation severity by psychological distance. (McGraw, et al., 2012).

Owing to the comparatively recent development of the benign-violation theory, a literature search retrieves only one other study that sought to provide support for the theory. The study is an unpublished paper that sought to determine whether the benign-violation theory also explained humorous responses in an Indian sample, and if a particular humour style was associated with amusement at benign-violation scenarios (Eriksson, 2011). The same moral violation scenarios used by McGraw and Warren (2010) were used in this study. The results indicated that Indian participants generally found the scenarios more amusing than their contemporaries from the United States. The Indian participants were, however, unable to identify that the scenarios contained moral violations, whereas participants from the United States were able to understand the violations. The Indian participants' inability to perceive the violations suggests that the benign-violation theory may not be the best explanation for these results. For example, it is possible that due to different linguistic and cultural knowledge, Indian participants showed a response bias toward high levels of amusement because that is what they think their response should be, rather than because they were able to appraise the scenarios as violations and as benign. Future research should develop benign-violation scenarios that are valid to use in cross-cultural studies. The Humor Styles Questionnaire (HSQ; Martin, Puhlik-Doris, Larsen, Gray & Weir, 2003) was also distributed to all participants, and scores on the individual subscales (i.e. affiliative,

enhancing, aggressive, and self-defeating humour) were correlated with amusement levels. The only significant correlation occurred between amusement and an aggressive humour style in both the Indian and American samples. Because of the Indian sample's lack of understanding of the moral violations no theoretical inferences can be made. The significant correlation in the American sample indicates that personality factors may be a mechanism by which benign-violations are perceived as amusing.

### **Methodological shortcomings of McGraw and Warren's work**

McGraw and Warren (2010) reported five small studies in their first paper. The first of these studies was titled "*Benign moral violations elicit laughter*" and is the foundational study on which all their subsequent studies are based. This study appears to be a neat study that clearly demonstrates that participants laughed at behaviours that they considered immoral, however a more critical look at the methods used to collect and analyse the data reveals two significant methodological shortcomings. The first shortcoming is that asking participants if the scenario made them *laugh* is an unreliable way of determining whether or not a situation was humorous. The second shortcoming is that the statistical method that was used to analyse the data was inappropriate because categorical data should not be entered into an analysis of variance (ANOVA). Before discussing these two shortcomings further it is important that the reader is aware that the aim of the remainder of this paper is to improve upon the study conducted by McGraw and Warren by making appropriate changes so as to minimise methodological shortcomings and in so doing improve the reliability of the results.

**Laughter as a measure of the perception of humour.** Although laughter is a common response to humorous situations, laughter does not necessarily mean that something is funny nor does perceiving something to be funny necessarily lead to laughter (Billig, 2005). The biological psychology literature relating to humour corroborates this as studies of the neural underpinnings of laughter have revealed that there are two neural pathways that, when excited, lead to laughter. In the first pathway laughter occurs as a response to feelings of pleasure that an individual experiences in response to a stimuli, and in the second pathway laughter occurs in the absence of significant levels of pleasurable affect (Wild, Rodden, Grodd & Ruch, 2003). Based on this biological evidence there are two main reasons why asking participants to report if they laughed at a scenario is inadvisable. People will often laugh in the absence of perceiving something to be humorous because it is a socially beneficial response. This indicates that laughter, as a form of communication, is dependent on the social context in which individuals find themselves (Van Hooff & Preuschoft, 2003). When one considers that participants were not engaging in direct social

contact with anyone as they read the scenarios it is reasonable to assume that they did not always laugh when they found a scenario amusing because there was no need for them to communicate this inner state of pleasure with the outside world. A second reason for asking whether or not participants laughed at a scenario is unwise in that some of the scenarios used in the study contained moral violations and as such it is possible that participants' laughter was a release of awkward tension as opposed to pleasant affect (Billig, 2005).

Based on the evidence above, asking participants whether or not they laughed at a scenario is a poor way of evaluating if something is humorous. There are a number of possible alternative questions that could be asked to try and gauge if a scenario is humorous. One possibility is to ask participants if they thought the scenario was *funny*, but this is also unreliable as the word funny can have more than one meaning. That is, participants might say the scenario is funny because it is humorous or they might say it is funny because it is not what they expected or in more colloquial terms they thought the scenario was weird (Hurley et al., 2011). Martin (2007) prefers to use the affective term *mirth* to gauge whether or not a participant finds something humorous. The downfall of asking this question is that English speakers who have grown up during the late 20th and early 21st century do not commonly use the word mirth. Eventually it was decided that the best way to improve upon the measurement used by McGraw and Warren was to ask participants if they thought the behaviour described in the scenario was *amusing*. The word amusing is commonly used by English speakers and its meaning is not ambiguous, thus it is a better choice than *funny* or *mirth*. *Amusement* is also a term that is used to describe an affective state which is preferable to a behavioural act like laughter.

**Mismatches between data and statistical methods.** Following each scenario participants in McGraw and Warren's study were asked two questions: 1) *Is the behaviour described in this scenario wrong?* and 2) *Did this scenario make you laugh?* Participants could only answer 'yes' or 'no' to each question, and thus the type of data that the questionnaire elicited was binomial categorical data. McGraw and Warren then erroneously entered this categorical data into an ANOVA. Although entering categorical data into an ANOVA is fairly common in social science literature it is inadvisable for a number of reasons. The first is that ANOVAs are concerned with descriptive statistics such as means, variances and confidence intervals all of which are made difficult to interpret because the numbers do not mean specific things. For example, if 'yes' is coded as 1 and 'no' is coded as 0, it is not immediately obvious what an average of say 0.87 means (Agresti, 2002). Additionally, the confidence intervals attained when proportions or percentages derived from

categorical data are entered into an ANOVA are not interpretable because the interval ascribes probable mass to results that cannot realistically occur, thus underestimating the probability of results that can meaningfully occur (Jaeger, 2008).

There are two general ways that the data analysis used by McGraw and Warren can be improved upon while still using the original stimulus material and questions. The first is to change the way that the responses are scored so as to create a continuous dependent variable that can then be appropriately entered into an ANOVA. The second possible way is to use a multilevel regression method of data analysis to analyse the original categorical data. The current paper uses the first of these methods to determine if the data still supports the benign-violation hypothesis when they are treated more appropriately.

### Aims

The overarching aim of this study was to test the benign-violation theory of humour. That is, this study sought to determine whether the answer to the question “what makes things funny?” is that a benign-violation must be perceived. McGraw and Warren’s (2010) original study was modified and replicated in order to investigate this aim. It was expected that participants would find scenarios that were constructed to be benign-moral violations more amusing than comparable scenarios that were simply benign or non-threatening. The data that was gathered during this project was analysed in three different ways. Each of these three analyses primarily aimed to answer the question “what makes things funny?”. A second purpose for conducting three separate analyses of the same data was to demonstrate that the benign-violation theory of humour is still supported when more robust methods for analysing the data were used, and if indeed the effect for the version factor (i.e. benign-violation version vs. non-violation version) scenario is in fact strengthened by the use of these more appropriate methods.

### Methods

#### Design and setting

**Independent Variables.** This study is a 2 x 2 x 2 x 4 mixed design. The four independent variables in this study are version (violation, control), sex (male, female), version order (violation first, control first) and scenario order (1234, 2413, 3142, 4321). Version is varied within subject, and the remaining three factors are between subject variables. The order factors are included so as to ensure that their being exposed to prior scenarios does not influence participants’ responses to subsequent scenarios.

**Dependent Variable.** This study set out to measure if participants find the violation and control versions of the scenarios they read humorous. The dependent variable was

operationalized differently for each of the analyses in this study. This will be explained further in the ‘data analyses’ section of this paper.

**Setting.** Data were collected from undergraduate students at the University of Cape Town. The data was collected via an online survey that participants could access from their personal computers.

### **Participants**

204 undergraduate students from the University of Cape Town were recruited to participate in this study. The sample consisted of 168 females and 36 males. All participants were recruited using a convenience sampling method. Specifically, participants were recruited via the Student Research Participation Programme (SRPP) at the University of Cape Town. Consequently, the sample used in this research was made up of undergraduate students from the University of Cape Town. This is comparable to McGraw and Warren’s study as they also used university students as their participants.

No criteria were used to exclude participants from enrolling this study. Although research suggests that severe neuropsychological and psychiatric conditions affect the perception of humour it is highly unlikely that university students would be suffering from these conditions as they would be unable to function in a university setting (Marziali, McDonald & Donahue, 2008; Wild et al., 2003).

The 204 participants were divided into eight groups (Table 1). These eight groups were defined by the version order and scenario order factors. Effects for sex will be tested for in the analyses, but because McGraw and Warren did not find a significant effect for sex in their study equal numbers of males and females in each group are not required.

Table 1  
*Sample characteristics*

Scenario Order	Version Order	Sex	<i>N</i>	
1234	Control first	Female	22	
		Male	4	
		Total	26	
	Violation first	Female	19	
		Male	7	
		Total	26	
	Total	Female	41	
		Male	11	
		Total	52	
	2413	Control first	Female	21
			Male	2
			Total	23
Violation first		Female	21	
		Male	3	
		Total	24	
Total		Female	42	
		Male	5	
		Total	47	
3142		Control first	Female	22
			Male	4
			Total	26
	Violation first	Female	23	
		Male	4	
		Total	27	
	Total	Female	45	
		Male	8	
		Total	53	
	4321	Control first	Female	22
			Male	4
			Total	26
Violation first		Female	18	
		Male	8	
		Total	26	
Total		Female	40	
		Male	12	
		Total	52	
Total		Control first	Female	87
			Male	14
			Total	101
	Violation first	Female	81	
		Male	22	
		Total	103	
	Total	Female	168	
		Male	36	
		Total	204	

## Procedure

This study was granted ethical approval by the University of Cape Town's Research ethics committee before data collection was initiated (Appendix A). In line with good ethical practice, participants were informed as to the purpose of the study and what would be required of them should they choose to participate. The only foreseeable risk that participants may have encountered while participating in this study was that they might find some of the scenarios offensive. Participants were ensured that their responses to the questions following each scenario would not be used to endorse or condone the behaviour described in the scenario.

An electronic invitation to participate in this study was sent out to potential student participants via the SRPP website (Appendix B). The students were offered one SRPP point as compensation for their participation (Acquiring SRPP points is a requirement of many psychology courses at the University of Cape Town). Recipients of the email who chose to participate in the study were required to email the researcher in order to obtain the URL link to one of the eight versions of the questionnaire (Appendix C).

Upon accessing the questionnaire, participants were presented with a consent form (Appendix D). After reading the consent form participants were asked whether or not they give their consent to participate in the study by selecting either 'yes' or 'no'. Participants who chose to participate were then asked to identify their sex and to provide the course code that they would like their SRPP point to count towards. Participants then read and responded to each version of each scenario as set out according to their group characteristics. Once participants had completed the survey they were thanked for their participation, debriefed, and asked to submit their responses. The survey took participants 20 minutes to complete.

**Stimulus material.** A control and a violation version of four short scenarios were used as stimulus material (Appendix E). These scenarios are more-or-less the same scenarios that McGraw and Warren used. The scenarios were adapted slightly to make them more culturally relevant. For example, where the original version of scenario three speaks about money in terms of US dollars, the scenario in this study speaks about South African Rands. The violation versions of the scenarios describe behaviours that violate widely recognised moral norms. However, the violations are phrased in such a way that the situation is unusual enough to appear hypothetical, and the scenario can also be appraised as acceptable in light of a different norm. In other words the violation scenarios were designed to be benign-

violations because the hypothetical and potentially justifiable nature of the scenarios decrease the severity of the violation of the moral norm.

Following each version of each scenario the participants are asked to answer two questions to which they can respond “yes” or “no”. The first question is “Is the behaviour described in this scenario wrong”, which seeks to detect whether participants appraise the scenario to be a violation. The second question is “Is the behaviour in this scenario amusing?”.

These scenarios were set up on the online survey platform *Survey Monkey*. Eight separate versions of the survey were created – one for every group. The reason so many versions of the survey were developed is that the output that *Survey Monkey* produces does not indicate the order in which participants have seen the scenarios or the order in which they see each version of the scenarios. A hardcopy of a survey (Version order: Violation first; Scenario order: 1234) can be seen in Appendix F.

## Results

The data in this study was analysed using the statistics software package SPSS version 21. In order to improve on the methods used by McGraw and Warren (2010) participants’ responses were scored in three different ways and thus three separate sets of data analyses were undertaken. In line with the primary aim of this study, each analysis aimed to demonstrate that participants were more likely to be amused by benign-violation versions of a scenario than by non-violation versions, thus supporting the benign-violation theory of humour. Each of the three analyses and their respective results all support the benign-violation theory of humour. The second aim of this study was to demonstrate that using more sophisticated statistical techniques to analyse participants’ responses to the original stimulus material used by McGraw and Warren would also lead to results that support the benign-violation theory of humour and that the size of the effect of the within-subject version factor (benign-violation vs. non-violation) would be increasingly high as the type of analysis used becomes more sophisticated. The results do indeed indicate that this is the case as the results indicate that the assumptions of the mixed-ANOVA design that was used in this research were violated to an increasingly lower extent for each subsequent analysis thus demonstrating that the results become increasingly more reliable. Furthermore, the effect size for version was highest for the third analysis which is the most refined of the three.

### Analysis I: Direct replication



This data analysis was a replication of the analyses used by McGraw and Warren (personal communication, 5 May 2013). The results of this analysis were used to demonstrate that participants' responses to the stimulus material were similar to those that McGraw and Warren obtained in their study, and to show that if analysed in the same way (albeit a faulty way), the results will indicate that participants were more amused by benign-violation scenarios than by non-violation scenarios.

**Dependent variables.** Responses to the questions following each scenario were assigned a value of 1 ("yes") or 0 ("no"). Participants' responses to the questions "is the behaviour in this scenario wrong?" and "is the behaviour in this scenario amusing" were then averaged across the two versions of the scenario. This produced four sets of proportions that indicate the percentage of participants who found each version of the scenarios wrong and amusing. A third outcome variable was then computed to indicate what proportion of the participants found each version of the scenarios both wrong and amusing. This third set of variables was computed with the aim of demonstrating that it is possible to be amused by a scenario that violates a moral norm held by the individual – an observation consistent with the benign-violation hypothesis.

**Statistical analyses.** Although problematic for the reasons described in the introduction to this paper, three  $2 \times 2 \times 2 \times 4$  mixed model ANOVAs were conducted to analyse these data. The answers to the *wrong* question was entered into the first ANOVA, answers to the *amusement* questions in to the second, and the computed variable for *both* was entered into the third ANOVA. An alpha level of .05 was used to determine whether or not it was appropriate to reject the null hypotheses. Owing to the presence of a repeated measures factor Partial Eta-Square ( $\eta^2_p$ ) values were calculated to determine the unique variance that was explained by any significant effects.  $\eta^2_p$  was chosen as the measure of effect size for this study for two reasons. Firstly,  $\eta^2_p$  is more appropriate than  $\eta^2$  for a design that contains a repeated measures factor (Bakeman, 2005). Secondly, a  $\eta^2_p$  value is most useful for comparing the effect a single factor has across numerous studies that use different designs (Pierce, Block & Aguinis, 2004). This second reason was particularly relevant to this study as a whole because the aim of this study is to compare the effectiveness of a new method for determining whether or not a violation version of a scenario was more likely to predict amusement than a non-violation version.

**Statistical assumptions.** Q-Q plots, kurtosis scores and skewness scores were used to determine whether or not the three sets of outcome variables were normally distributed (Figure 2, Table 2). Statistical tests for normality were not used because, given that this study

had a large sample, they are likely to indicate that data is non-normal even if it is relatively close to being normally distributed (Field, 2009). The assumption of normality was upheld for the three outcome variables for the violation versions of the scenarios, but the scores for the non-violation scenarios were clearly non-normal. The data for the non-violation scenarios and the kurtosis scores indicate that the data were clustered around a value that was not the mean value of the data. These problems in the data exist because the vast majority of the participants did not rate any of the non-violation scenarios wrong (81%) or amusing (63%) and as a result the data was very negatively skewed.

The assumption of homogeneity of variance was tested using Levene's test (Table 3). The assumption was only upheld for the outcome variables that measured whether the violation version was seen to be wrong, and whether the violation version was both wrong and amusing ( $ps > .05$ ). The variances for all the other outcome variables are significantly different ( $ps < .05$ ) and therefore they violate the assumption of homogeneity of variance. Field (2009) states that when the sample size is large Levene's test is more likely to be significant even if the data are relatively homogenous.

Sphericity could not be calculated for this dataset as the repeated measures factor (i.e. version) only has two levels. The data was therefore treated as if the assumption of sphericity had been upheld.<sup>1</sup>

It was not surprising that the data used in this analysis regularly violate assumptions because, as was stated earlier in this paper, this was categorical data and thus should not technically be entered into an ANOVA. This was particularly true of this data because participants were forced to choose between only two categories and thus the mean statistics, on which the assumptions of normality and homogeneity of variance are based, were especially hard to interpret (Jaegar, 2008). Although the data in this analysis was problematic no transformations were made to the data as the purpose of this analysis was to replicate the method used by McGraw and Warren (2010) as closely as possible.

---

<sup>1</sup> This also applies to both analysis 2 and 3.

Table 2

*Analysis II: Descriptive statistics relating to the assumption of normality.*

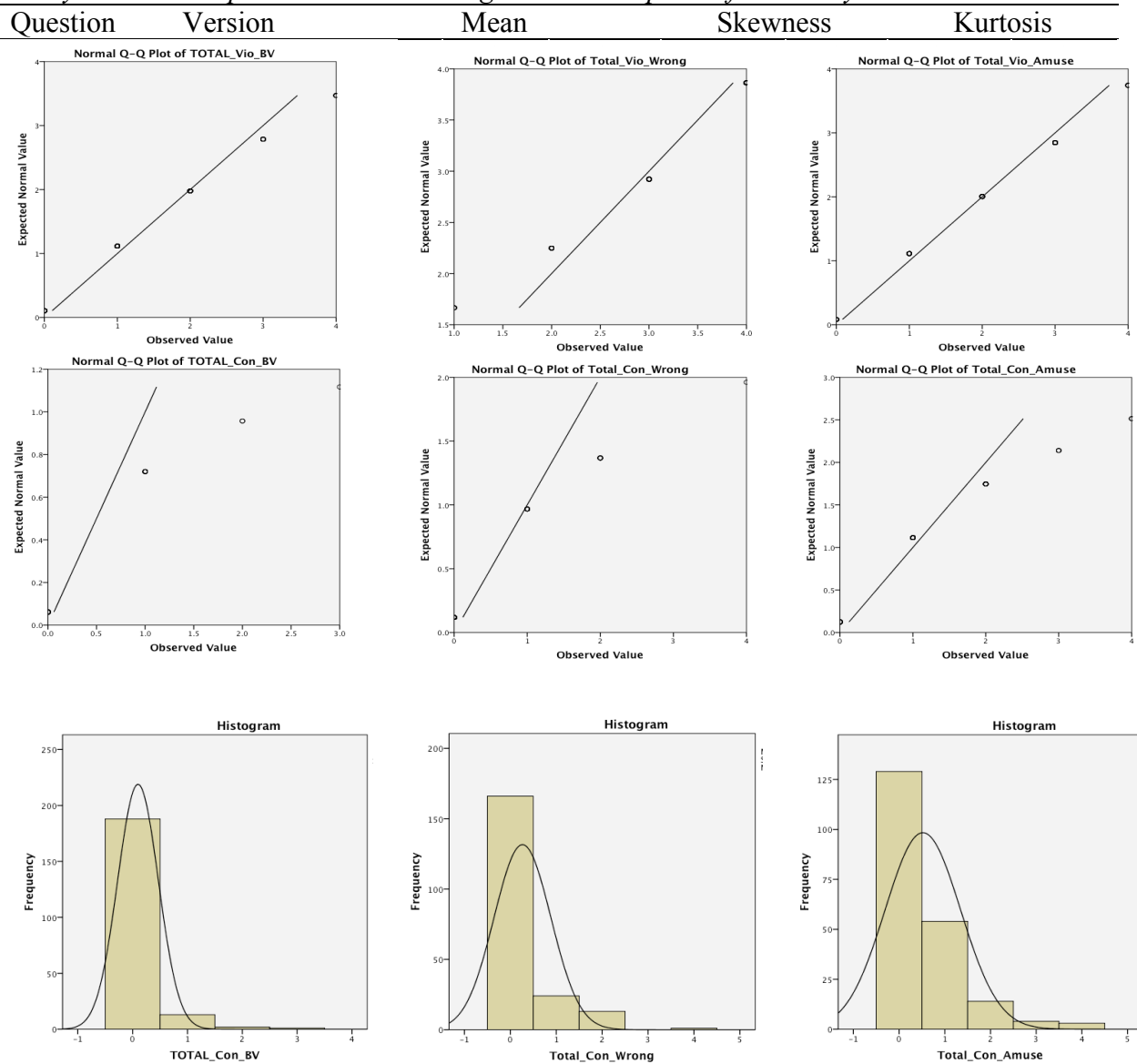


Table 3

*Analysis II: Levene's test for equality of variance*

Question	Version	$S^2$	$F$	$df1$	$df2$	$Sig.$
Wrong	Violation	.70	1.256	15	188	.234
	Non-Violation	.38	5.280	15	188	.000
Amusing	Violation	1.17	2.18	15	188	.008
	Non-Violation	.68	5.50	15	188	.000
Both	Violation	1.15	1.27	15	188	.224
	Non-Violation	.14	7.16	15	188	.000

**Results.** The analyses showed no significant effect for sex, version order or scenario order and no significant interaction effects were present in any of the analyses,  $ps > .05$ . The

Figure 2. Q-Q plots and histogram plots illustrating the distribution of the data for each

proportion of participants who answered “yes” to the questions following each scenario are reported below (Table 4). Overall the participants judged the behaviour in the violation version to be wrong more frequently than the behaviour in the non-violation version,  $F(1, 188) = 833.8, p < .001, \eta^2_p = .82$ . Participants were also more likely to be amused by the behaviour in the violation scenarios than the behaviour in the non-violation scenarios,  $F(1, 188) = 77.3, p < .001, \eta^2_p = .29$ . Finally, participants were more likely to judge the behaviour in the violation scenarios to be both wrong and amusing,  $F(1, 188) = 119.3, p < .001, \eta^2_p = .39$ . This result is consistent with the benign-violation hypothesis as it indicates that participants perceive a scenario to be a moral violation yet they are still able to laugh at it indicating that they also perceive the scenario to be acceptable or benign.

These results appear to be similar to the results reported by McGraw and Warren (2010) in that they too found that all three of their ANOVAs were significant. They also reported no main effects for the between-subject factors and no significant interactions. However, McGraw and Warren did not report effect sizes for any of their  $F$  tests, and thus it is very difficult for any future researcher to replicate their study because the strength of the predictive power of the version factor is unknown. The effect sizes reported for the significant  $F$ -tests in the present analysis range from  $\eta^2_p = .29$  to  $\eta^2_p = .82$ . These statistics are evidence that a significant ANOVA is not sufficient to ascribe real meaning to the relationship between a particular independent variable and the dependent variable. A comparison between the two studies is also difficult to make because effect sizes are contingent upon sample size, and this study had a larger sample size than McGraw and Warren’s study (204 vs. 66). On the surface the results from analysis one appeared to support the benign-violation hypothesis because the version factor strongly predicted ( $\eta^2_p = .39$ ) whether a scenario was perceived to be both amusing and wrong. However, due to the inappropriateness of using categorical data in this analysis and the lack of effect sizes reported by McGraw and Warren the results of this analysis are not reliable.

Table 4  
*The proportion of positive (“yes”) responses to the questions following each scenario.*

Scenario and version	Response (%)		
	Behaviour is wrong	Behaviour is amusing	Both
Snorting Remains			
Violation	87	43	33
Non-Violation	9	19	4
Selling Virginity			
Violation	89	10	8

Non-Violation	3	4	1
Stealing Tips			
Violation	90	42	36
Non-Violation	8	12	2
Endorsing Pork			
Violation	68	72	50
Non-Violation	7	17	3
Average			
Violation	84	42	32
Non-Violation	7	13	3

### Analysis II: Humour scale measuring amusement at violations

Given the numerous problems associated with the categorical data in the first analysis, a second analysis was undertaken in which steps were taken to improve the appropriateness of the match between the type of data that was elicited from the stimulus material and the statistical methods used to analyse the data. This second analysis spoke to both aims of this study in that it sought to demonstrate once again that benign-violations are more humorous than non-violation scenarios, as well as to undertake a new method of analysing the data that was expected to be more reliable.

**Dependent variable.** In this analysis the questionnaire was scored in a novel manner so as to elicit continuous data that could be entered into an ANOVA as the dependent variable. In order to do this each scenario was viewed as an item on a scale that measures people's level of amusement at both levels of the version factor (i.e. violation or non-violation version). This helped to solve one of the major problems in the first data analysis which was that three separate ANOVAs were used to analyse the data. This was a problem because the use of multiple ANOVAs increases the chance of incorrectly rejecting the null hypothesis (i.e. a Type 1 error). Consequently, computing a single dependent variable to measure amusement at both levels of the repeated measures version factor results in the need to only conduct one ANOVA.

In order to compute this new composite dependent variable, participants received a score of 0, 1 or 2 for each scenario. They received a 0 if a) they did not find the scenario wrong or amusing, or b) they found the scenario to be wrong, but not amusing. They received a 1 if they reported that a scenario was amusing, but not wrong. The reason these responses were allocated a value of 1 is that it is reasonable to assume that although the participant may not have deemed the scenario to be a particularly severe violation and thus they did not personally think it was wrong, they most probably still perceived that the scenario contravened a moral norm held by society in general and thus they found it amusing. Thirdly,

participants received a 2 if they perceived the scenario to be both wrong and amusing. These scores were then added up and each participant received two scores out of eight – one for the violation versions and one for the control versions – measuring the participants’ amusement at violation scenarios. This newly computed dependent variable was a continuous interval variable where a zero suggests that the participant showed low levels of amusement as a function of violation perception and an eight suggests that participants were highly amused by scenarios that they perceived to be a violation.

**Statistical analyses.** A single 2x2x2x4 mixed design ANOVA was performed. The participants humour scores for the control and the violation scenarios were entered into the levels of the repeated measures factor “version”. An alpha level of .05 was used to determine whether or not it was appropriate to reject the null hypotheses for each of the main and interaction effects. Effect sizes ( $\eta^2_p$ ) for each of the significant effects in the ANOVAs were also calculated and interpreted.

**Statistical assumptions.** The Q-Q plots and histograms (Figure 3) suggested that scores representing participants’ amusement at violations for the violation versions of the scenarios were fairly well normally distributed, but the scores for the control versions of the scenarios were not normally distributed. The skewness statistic indicated that scores on the control version of the scenarios were negatively skewed and the kurtosis statistic indicated that the data was clustered quite markedly around a value that is not the mean (Table 5)

Levene’s test for homogeneity of variance was used to check if the variances of the data were in fact homogenous. The variances of different groups’ humour scores for the violation scenarios were homogenous and thus the assumption was upheld for the violation level of the version factor. The homogeneity of variance assumption was violated for the scores on the non-violation version of the scenarios (Table 6).

Although both assumptions were violated for the non-violation level of the version factor no adjustments or transformations were made to the data because any changes made to the data based on non-violation scores will affect the normality and homogeneity of the violation scores. Furthermore, ANOVA is a robust statistical test (Wilcox, 2005). That is, its efficiency is not easily diminished by violations of assumptions. This robustness is particularly strong when the groups that the sample is divided into are equal. In this study group sizes were not significantly different from one another and therefore no changes to the data were made,  $\chi^2(7, N = 204) = .471, p > .001$ .

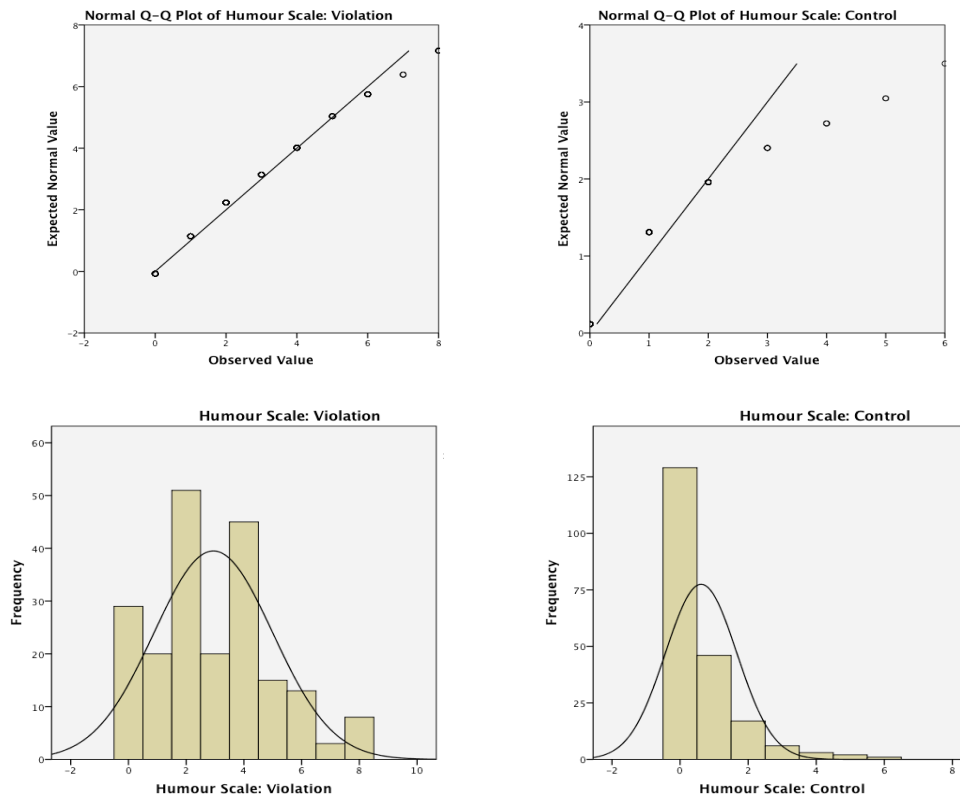


Figure 3. Q-Q plots and histogram plots illustrating the distribution of the data for the humour variable in analysis two.

Table 5

Analysis II: Descriptive statistics relating to the assumption of normality.

Version	Mean			Skewness		Kurtosis	
	Stat.	SE	SD	Stat.	SE	Stat.	SE
Violation	2.94	.14	2.06	.49	.17	-.22	.34
Control	.62	.07	1.05	2.31	.17	6.29	.339

Table 6

Analysis II: Levene's test for equality of variance

Version	$S^2$	F	df1	df2	Sig.
Violation	4.24	1.44	15	188	.13
Control	1.10	5.29	15	188	.00

**Results.** The mixed model ANOVA indicated that there were no significant main effects for sex or either of the order factors ( $ps > .05$ ). A significant main effect was observed for scenario version indicating that participants rated the violation scenarios to be more

amusing and wrong ( $M = 2.94$ ,  $SD = 2.06$ ) than the non-violation scenarios ( $M = 0.62$ ,  $SD = 1.05$ ),  $F(1, 188) = 111.4$ ,  $p < .001$ . This effect has a moderate effect size ( $\eta^2_p = .37$ ). This effect size is marginally smaller than the effect size observed in the first analysis ( $\eta^2_p = .39$ ) indicating that version explains slightly less of the variance when the data is viewed as a scale as opposed to separate scenarios. However, given the inappropriateness of the analyses used in the first analysis the results of this analysis are assumed to a better reflection of reality than the first analysis.

The results of this analysis also indicate that there were two significant interaction effects at play in this data set. The first significant interaction was VERSION X SCENARIO ORDER,  $F(3, 188) = 2.66$ ,  $p < .05$ . The effect size of this interaction was small ( $\eta^2_p = .04$ ) indicating that it exerts very little power over the dependent variable. Even though the effect size of this interaction was small three cautionary post-hoc pairwise comparisons were conducted to ensure that the data was understood as fully as possible. Conducting multiple comparisons increases the probability of making Type I error, therefore a Bonferroni corrected level of alpha will be used. This would be  $\alpha = .05/6 = 0.008$ . Based on the information depicted in Figure 4 the following three interactions will be investigated:

1. The difference between scores on the violation version and the control version for each scenario order.
2. The difference between control version scores when the scenario orders 2413 and 3142.
3. The difference between control version scores when the scenario orders 4321 and 3142.

The first comparison indicated that there is a significant difference between the control and violation scenarios at each of the four scenario orders ( $p < .001$ ). The second pairwise comparison indicated that the estimated marginal means for scenario order 2413 ( $M = .49$ ) and scenario order 3142 ( $M = .89$ ) are not significantly different,  $p = .06$ . Similarly the third pairwise comparison indicates that scenario order 3142 ( $M = .89$ ) and scenario order 4321 ( $M = .52$ ) are not statistically significant,  $p = .07$ .



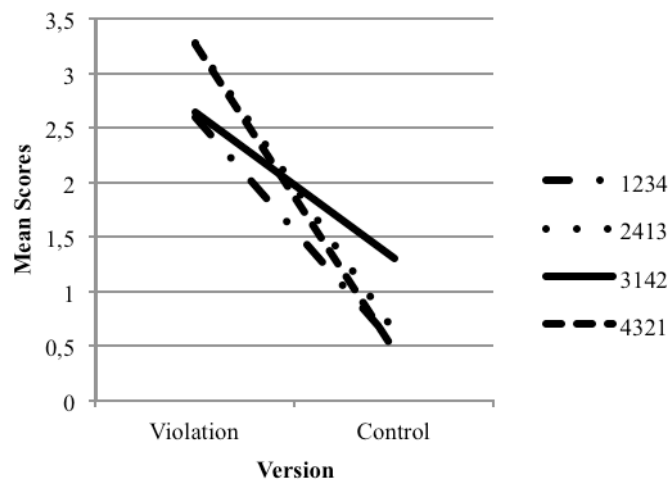


Figure 4. Visual representation of the version x scenario order interaction effect in analysis two

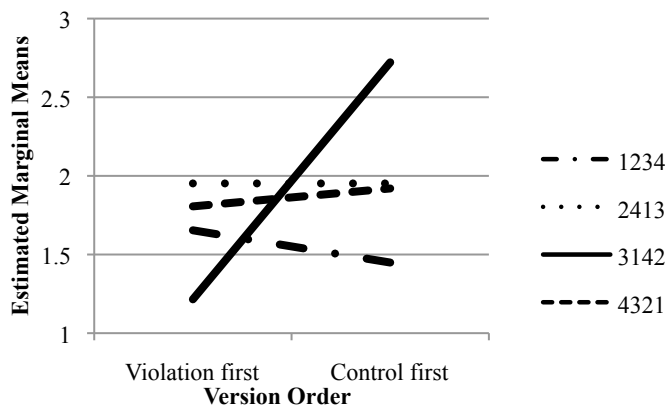


Figure 5. Visual representation of the version order x scenario order interaction effect in analysis two

The second significant interaction effect was between two between-subject factors VERSION ORDER X SCENARIO ORDER,  $F(3, 188) = 2.81, p < .05, \eta^2_p = .04$ . Figure 5 clearly indicates that scores for scenario order 3142 are affected by the order in which the version of each scenario was seen. The following pairwise comparisons were conducted to better understand this interaction:

1. The difference between non-violation and violation scores for scenario order 3142 when either the violation or control version was shown first.
2. The difference between both version scores for scenario order 3142 compared to all other scenario orders.

As a total of eight comparisons will be conducted therefore the significance level will be adjusted using Bonferroni's calculation. This would be  $\alpha = .05/8 = 0.006$ . The results of these pairwise comparisons are set out in Tables 7 and 8 below.

The results of the pairwise comparisons indicated that two of the eight comparisons were significant. The two significant comparisons were the scores for the non-violation version of the scenario and not the violation version. This difference is unsurprising given the non-normal distribution and equal variance of control version scores reported previously. The scenario order 3142 is clearly an anomaly in this dataset, given that participants got significantly higher scores when they saw the non-violation version of the scenario first ( $p = .002$ ). This may be because seeing the control version first may cause the violation in the violation version to be amplified. Why this should occur only in scenario order 3142 is not immediately clear. It is possible that this is simply a statistical artefact and given that the total variance that this interaction explains is small (4%) it is not necessary to tease out the intricacies of this interaction. Finally, scenario order 3142 the score on the control version of the scenarios when the control version is presented first is significantly different to the comparable ratings on scenario order 4321 ( $p = .005$ ).

Table 7  
*Scenario Order x Version Order: Pairwise comparison 1*

Dependent Variable	Scenario Order	Version Order	<i>MD</i> <sup>a</sup>	<i>SE</i> <sup>b</sup>	<i>Sig.</i>
Violation Version Scores	3142	Control first vs. Violation first	1.427	.565	.012
		Violation first vs. Control first	-1.427	.565	.012
Control Version Scores	3142	Control first vs. Violation first	.902*	.281	.002
		Violation first vs. Control first	-.902*	.281	.002

\* The mean difference is significant at  $\alpha = 0.006$  (Bonferroni adjustment)

a. *MD* = Mean Difference

b. *SE* = Standard Error

Table 8  
*Scenario Order x Version Order: Pairwise comparison 2*

Version Order	Scenario Order	<i>MD</i> <sup>a</sup>	<i>SE</i>	<i>Sig.</i>
Control First	3142 vs. 1234	.654	.283	.022
	3142 vs. 2413	.607	.292	.039
	3142 vs. 4321	.808*	.283	.005
Violation First	3142 vs. 1234	.021	.281	.939
	3142 vs. 2413	.194	.286	.498
	3142 vs. 4321	-.056	.281	.843

\* The mean difference is significant at  $\alpha = 0.006$  (Bonferroni adjustment)

a. Difference between the dependent variable scores for control scenarios.

### **Analysis III: Humour scale measuring amusement at violations as a function of level of immorality**

Like the second analysis, the third analysis also informs both aims of this study in that the effect of the version factor will be analysed to determine whether this data supports the benign-violation theory for a third time. In line with the second aim of this study, the method of data analysis is further adjusted in this analysis in order to determine if this model is more reliable than the previous two.

The second analysis undertaken in this paper assumed that all the violation versions of the scenarios were equally immoral and that all non-violation versions were equally moral. This was a faulty assumption because, for example, the data indicated that participants rated the violation version of the scenario about the mother taking the tips of the staff who had served at a wedding (Scenario 3,  $n = 184$ ) to be much more wrong than the Rabbi advertising pork products (Scenario 4;  $n = 139$ ). This indicates that participants saw stealing as more immoral than a religious leader acting in a manner incongruent with the beliefs of that religion. The number of participants who rated the violation version of each scenario to be wrong was different across the four scenarios,  $\chi^2(3, N = 204) = 7.92, p < .05$ . Consequently, it was deemed necessary that this inconsistency be taken into account when scoring and analysing the data.

This study has a large sample size ( $N=204$ ) and as such it is acceptable to assume that participants' ratings of the immorality of the scenarios in this study are representative of the moral norms that are held by the population as a whole.

**Dependent variable.** Participants' responses were first scored using the same method described in analysis II. The resulting scores out of two for every version of every scenario were then multiplied by their respective immorality index so as to account for the degree to which each scenario is morally wrong. The immorality index was computed by calculating what proportion of participants reported that they perceived each version of each scenario wrong (Table 9). Participants' adjusted scores for each version were then added up so that a composite measure of amusement, taking the degree of immorality into account, was created for both the violation and non-violation versions of the scenarios.

Table 9

*Analysis III: Immorality indices*

Scenario	Version	Immorality Index
1 - Snorting Ashes	Violation	.87
	Non-Violation	.09
2 – Selling Virginity	Violation	.89

3 – Stealing tips	Non-Violation	.02
	Violation	.90
4 – Rabbi promoting pork	Non-Violation	.08
	Violation	.68
	Non-Violation	.07

**Statistical analyses.** The statistical analyses used for analysis III is identical to that used in analysis II.

**Statistical assumptions.** The Q-Q plots and histograms (Figure 6) demonstrate that the humour scores that have been adjusted to account for degree of immorality were very nearly normally distributed for the violation level of the version factor, whereas the scores for the non-violation level were not normally distributed. The skewness and kurtosis statistics for the distribution of the control scenarios are still fairly concerning, but they are much less concerning than the scores for the previous analyses which suggests that the results of this analysis are more reliable than the previous analysis (Table 10).

The assumption of homogeneity of variance was also not upheld for the humour scores on the control versions, but as with the skewness and kurtosis scores, The  $F$  score obtained from Levene's test was lower than the  $F$  score in analysis II (Table 11). This indicates that the variance of the humour scores for each group in the control condition in analysis III were closer to being homogenous than the variances in the second analysis were, and as such the results of analysis III could be more reliably interpreted than previous analyses.

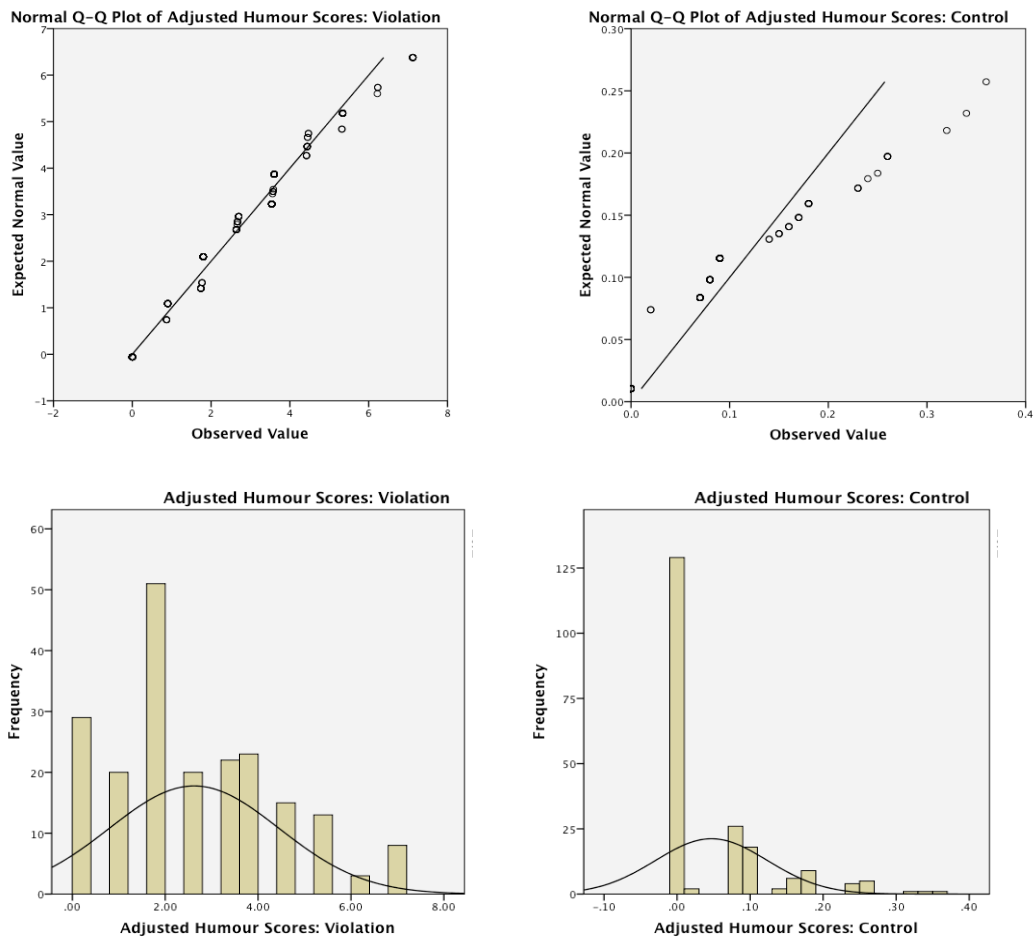


Figure 6. Q-Q plots and histogram plots illustrating the distribution of the data for the humour variable in analysis three.

Table 10

Comparison of the skewness and kurtosis scores for analysis II and III

Analysis	Version	Skewness		Kurtosis	
		Stat.	SE	Stat.	SE
II	Non-violation	2.31	.17	6.29	.339
III	Non-violation	1.83	.17	3.02	.339

Table 11

Comparison of Levene's test results for analysis II and III

Analysis	Version	Levene's test for homogeneity of variance				
		$S^2$	F	df1	df2	Sig.
II	Non-Violation	1.10	5.29	15	188	.00
III	Non-Violation	.01	4.41	15	188	.00

**Results.** The mixed model ANOVA indicated that there was a significant main effect for version,  $F(1, 188) = 203.22, p < .001$ . More specifically, participants rated violation scenarios ( $M = 2.62, SD = 1.83$ ) to be significantly more amusing than non-violation scenarios ( $M = .05, SD = .80$ ). The effect size for the effect of version is large ( $\eta^2_p = .52$ ). This suggests that 52% of the variance in amusement scores is explained by the within-subject version factor. This effect size is considerably larger than the effect sizes observed in analysis one ( $\eta^2_p = .39$ ) and analysis two ( $\eta^2_p = .37$ ).

No other main effects or interaction effects were observed in this dataset ( $ps > .10$ ).

### Discussion

The first aim of this study was to determine if benign-violations were more likely to be humorous than non-violation scenarios. The results of this study demonstrated that this is the case, and thus this research provides evidence in support of McGraw and colleagues' benign-violation theory of humour. That is, this data suggests that in order for something to be perceived as amusing it must violate an expectation that the recipient of the humorous exchange holds, yet the violation must also be justifiable in some way. In other words, the violation must be perceived to be both 'okay' and 'not okay' at the same time. Each of the moral violation scenarios used in this research could be appraised to be acceptable because they are hypothetical and thus psychologically distant from the participant (Liberian & Trope, 2008). This supports later research conducted by McGraw and Warren (2012) that demonstrated that when participants know that the stimulus material they are being shown is not real they are more likely to be amused than when they are told that the scenarios are true. It is also feasible to assume that these scenarios could also have been appraised to be benign because an alternate norm exists that justifies the immoral behaviour in the scenarios. For example, the scenario in which Keith snorted his dead father's ashes clearly violates a moral norm held by society – it borders on cannibalism! However, because Keith's father told Keith he could do whatever he wanted with his ashes, Keith's chosen method is not contrary to the wishes of the deceased, and thus could be seen as fairly acceptable. Incidentally, it may interest the reader to know that, although slightly exaggerated, this scenario is based on Rolling Stones singer-songwriter Keith Richards' treatment of his father's ashes (Richards, 2010).

As indicated in the introduction to this paper, the benign-violation theory of humour is the most recent of many other theories that have tried to explain what it is that makes something funny. Future studies will have to be done where two or more theories are pitted against one another to determine which theory best explains humour, however this study does

suggest that the benign-violation theory may be superior to the incongruity theory of humour (Suls, 1972 as cited in McGraw & Warren, 2010). The incongruity theory of humour agrees with the benign-violation theory of humour by suggesting that in order for something to be funny an expectation or a norm must be violated. Where they differ is that according to the benign-violation theory, it is essential that the violation not be perceived as so severe that it is threatening to a person's physical, cognitive, social or emotional well-being. Many participants in this study viewed the violation scenarios as wrong, yet they did not simultaneously report that they thought the scenario was amusing. This indicates that a perceived incongruity is not sufficient to predict amusement. A practical example that illustrates this is the common use of humour to soften criticism. Criticism involves informing someone that what he or she has done is incorrect. This is a type of violation because it indicates to the individual that something that although they may have considered their actions to be acceptable, they are in fact incorrect or unacceptable – this sort of criticism is not amusing. However, when someone wants to offer criticism in a more sensitive way they attempt to reframe the criticism in a more benign or light-hearted manner. This alteration in the language of communication often results in a humorous jibe that still communicates the criticism. For example, if a woman arrives at work with inappropriately messy hair, a supervisor could either choose to address this issue by saying “You look terrible.” or “You look like Einstein today!”. The first of these responses is not likely to be amusing because it is insulting and may threaten the woman's social and emotional wellbeing. However, by saying she looks like Einstein the supervisor is offering the woman the chance to laugh at her error, while still ensuring that she understands that she has done something wrong.

The particular assertion that something is humorous only when it is benign adds a subjective aspect to humour theory that has been lacking in previous attempts to understand what makes something funny (Martin, 2007; McGraw & Warren, 2010). This is indicative of the reality that no statement or behaviour is funny in and of itself; it is only humorous if *my* conscious experience appraises it as such. Thus, different things amuse different people because the degree to which something is perceived to be a violation is dependent on individual identity. This clarifies why, for example, prejudicial jokes about sensitive topics like race, gender or national identity are funny to some and not others (Billig, 2005). A joke that is disparaging about women may not negatively disturb the identity of a man, indeed it may strengthen their identity by placing them in a position of superiority and power, and thus they may find the joke humorous. Conversely, such a joke is much more threatening to a woman's subjective (and collective) identity as it directly undermines an integral and

consistent part of who they construct themselves to be. A woman is therefore far less likely to laugh at such a demeaning statement even if the intention of the teller was benign. When participants' responses to the scenarios in this research are compared to the responses of the American students in McGraw and Warren's study it appears cultural subjective identity may play a mediating role in determining whether something is funny. For example, in the second scenario in which Jenny decides to sell her virginity in order to contribute money towards getting her family out of debt, 45% of the American students sampled reported that this scenario made them laugh, whereas only 10% of the South African students in this study found the scenario amusing. This suggests that the American students sampled by McGraw and Warren were more liberal than the South Africans sampled in this study, and consequently the results of the studies are not automatically comparable.

Knowing that moral norms differ from one culture to another is an important reason why the methodological changes to the original measure were instituted. The inclusion of the immorality co-efficient in the third analysis is advantageous as it not only leads to a stronger design, but it increases the validity of the measure. Most importantly, the inclusion of the immorality co-efficient increases the ecological validity of the measure as it ensures that the outcome of interest (i.e. humour) is less likely to be confounded by different levels of morality in different cultures. Future research that seeks to validate the benign-violation theory of humour should therefore include an immorality co-efficient so as to improve comparability across studies. This speaks to the achievement of the second aim of this study that was to develop a more valid way to measure humour within the benign-violation theoretical framework.

It is inappropriate to compare the effect sizes from the direct replication conducted in analysis one to the subsequent analyses as the method used in study one is unreliable. Not only was frequency data entered into an ANOVA, but also the data was entered into three different ANOVAs which increases the chance of finding an effect even when one does not exist. The effect size for the version factor (violation vs. non-violation) in the second analysis wherein the scenarios were scored so as to create continuous quantitative data is fairly strong ( $\eta^2_p = .37$ ). The effect size for the third analysis in which adjusted scores based on population mean ratings of immorality was very large ( $\eta^2_p = .52$ ). Additionally, the two small interaction effects observed in analysis two were no longer significant in the third model.

The fact that both analyses two and three showed significant results, and that the third analysis' effect size is so strong is indicative of two important things. Firstly, it demonstrates that McGraw and Colleagues (2010; 2012) assertion that something is funny when it is a



benign-violation is not only true when weak methods are used to test the theory. Secondly, it indicates that all research must always be evaluated in a very critical manner, as three different ways of analysing the same data set were used in this paper, yet the results of the analyses were not wholly consistent with one another. This careful and critical approach to interpreting statistical results has been previously highlighted by writers such as Ian Hacking (1990) and Theodore Porter (1995) who caution against the temptation to accept numerical outcomes as truly factual information.

When the questionnaire was scored in the novel manner used in analyses two and three, each of the eight scenarios were viewed as items on a scale. Eight items is a relatively small number of items to have on a scale and thus may be considered as a limitation of this study. It was however, important for the purposes of this particular study that only the original scenarios used by McGraw and Warren (2010) be used so as to replicate their study as closely as possible. Future research should look to extend this scale to cover a wider range of moral violations and possibly to include scenarios in which the degree to which a norm is violated is varied. Future research could also make a further methodological adjustment to enhance the reliability of the results by using a multi-level statistical analysis to better understand the predictors of amusement.

In agreement with the work conducted by McGraw and colleagues (2010; 2012), this study has suggested that when a moral violation is appraised to be benign participants are more likely to be amused than when a scenario is not a moral violation. It is reasonable to surmise that other types of violations can be humorous when they can also be interpreted in a benign manner. A pun, for example, violates an anticipated linguistic expectation, if however, the violation can be quickly resolved by invoking an alternate linguistic rule the incidence is perceived to be amusing. It is therefore necessary for future researchers to carefully design studies to test whether the benign-violation theory of humour is able to explain myriad types of humour.

### **Conclusion**

The first aim of present study was to determine whether the answer to the question “what makes *immoral* things funny?” is that the immoral behaviour presented in scenarios must be perceived to be a benign-violation. That is, the situation must be appraised to be simultaneously acceptable and unacceptable. The results of this study indicated that this is indeed the case as participants were significantly more likely to be amused by scenarios that were designed to be benign-moral violations than by scenarios that were simply acceptable. Furthermore, the results indicate that simply perceiving a scenario to be wrong is not enough

to predict amusement, thus opposing a previous theory of humour that suggested that something was simply humorous when it was incongruous. In line with the second aim of this study, the results demonstrated that the significant effect for the version of the scenario found by McGraw and Warren (2010) was still valid when two new and more appropriate methods for scoring and analysing participants' responses were used. Whilst the results of this study cannot be used to assert that the benign-violation theory of humour is a reliable and valid general theory of humour, they do indicate that moral violations are humorous when they are perceived to be both acceptable and unacceptable. It therefore seems reasonable to hypothesise that other sources of humour are amusing because they are appraised to be both acceptable and unacceptable. In conclusion, the results of this study demonstrate that humour is a socially adaptive function wherein individuals are driven to reframe a violation that they encounter in a non-threatening light. This knowledge is important as individuals experience myriad stressors in everyday life that they need to face and overcome. In the words of Stephen Hawking (2004): "Life would be tragic if it weren't funny."

## References

- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3), 379-384.
- Billig, M. (2005). *Laughter and Ridicule. Towards a social critique of humour*. London: SAGE Publications Ltd.
- Eriksson, K. (2011). *Who finds benign moral violations amusing?* Unpublished manuscript, Center for the Study of Cultural Evolution, University of Stockholm, Sweden.
- Freud, S. (1908). *Wit and its relation to the unconscious*. London: Kegan Paul.
- Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *Quarterly Review of Biology*, 80, 395-430.
- Hacking, I (1990). *The taming of chance*. Cambridge, UK: Cambridge University Press.
- Hawking, S. W. (2004). The science of second-guessing. Retrieved from <http://www.nytimes.com/2004/12/12/magazine/12QUESTIONS.html>.
- Hurley, M. M., Dennett, D. C., & Adams, R. B. (2011). *Inside jokes: Using humor to reverse-engineer the mind*. Cambridge, Massachusetts: The MIT Press.
- Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.

- Liberman, N., & Trope, Y. (2008). The psychology of transcending the here and now. *Science*, *322*, 1201-1205.
- Martin, R. A. (2002). Is laughter the best medicine? Humor, laughter, and physical health. *Current Directions in Psychological Science*, *11*(6), 216-220.
- Martin, R. A. (2007). *The psychology of humor: An integrative approach*. Burlington, MA: Elsevier Academic Press.
- Martin, R.A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, *37*(1), 48-75.
- Marziali, E., McDonald, L., & Donahue, P. (2008). The role of coping humor in the physical and mental health of older adults. *Aging & Mental Health*, *12*(6), 713-718.
- McGraw, A. P., & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological Science*, *21*(8), 1141-1149.
- McGraw, A. P., Warren, C., Williams, L. E., & Leonard, B. (2012). Too close for comfort, or too far to care? Finding humor in distant tragedies and close mishaps. *Psychological Science*, *23*(10), 1215-1223.
- Pierce, C.A., Block, R.A., & Aguinis, H. (2004). Cautionary note on reporting Eta-Square values for multifactor ANOVA designs. *Educational and Psychological measurement*, *64*, 916-924.
- Porter, T.M. (1995). *Trust in numbers. The pursuit of objectivity in science and public life*. Princeton, New Jersey: Princeton University Press.
- Richards, K. (2010). *Life*. London, United Kingdom: Weidenfeld & Nicolson.
- Van Hooff, J.A., & Preischoft, S. (2003). Laughter and smiling: The intertwining of nature and culture. In F.B.M. de Waal & P.L. Tyack (Eds.), *Animal social complexity: Intelligence, culture, and individualised societies* (pp. 260-287). Cambridge, MA: Harvard University Press.
- Veatch, T. C. (1998). A theory of humor. *Humor: International Journal of Humor Research*, *11*, 161-215.
- Wilcox, R.R. (2005). *Robust estimation and hypothesis testing*. Boston, MA: Elsevier Academic Press.
- Wild, B., Rodden, F.A., Grodd, W., & Ruch, W. (2003). Neural correlates of laughter and humour. *Brain*, *126*, 2121-2138.

Wyer, R. S., & Collins, J. E. (1992). A theory of humor elicitation. *Psychological Review*, *99*, 663-688.

Appendix A

Evidence of ethical approval

1

What makes things funny?  
Evidence for the benign-violation theory of humour

Sarah Anne Westwood

Department of Psychology

University of Cape Town

*Ellis approval given by  
panel - not superior  
Docu 3/6/13  
Ref no. PSY2013-038*

Supervisor: Professor Johann Louw

Word Count: 4870

## Appendix B

### Invitation to participant in the study

We all laugh almost everyday. Humour is an important part of our social lives. In the words of Stephan Hawking: “Life would be tragic if it weren’t funny”.

You are invited to participate in a study that aims to find out what makes things funny. This study is open to all undergraduate psychology students.

If you would like to participate in this study you will need to fill in an online survey. It will take you no more than 15 minutes to fill in the survey. You will be awarded 1 SRPP point for your participation.

If you would like to participate in this study please send an email to the following address and the link to the survey will be sent to you. **psytutor.sarah@gmail.com**

Regards

Sarah

## Appendix C

### **Email sent to participants who expressed interest in participating**

Thank you for being willing to take part in this research.

Please fill in this survey carefully. The survey contains pairs of scenarios that are very similar. Please read them carefully so that you pick up the differences and can respond appropriately.

You will need to fill in your participant number at the start of the survey.

Participant Number:

The survey can be accessed at the following link:

Regards

Sarah Westwood

## **Appendix D**

### **University of Cape Town Consent to participate in a research study: What makes things funny?**

Dear participant,

#### **Study purpose**

You are invited to participate in a research study being conducted by a researcher from the Department of Psychology at the University of Cape Town. The purpose of this study is to determine what makes some things funnier than others.

#### **Study procedure**

If you do decide to participate in this study, you will be required to complete an survey via the online platform SurveyMonkey. The survey will include a short demographic questionnaire and three sets of written stimuli to which you will be asked to respond. The questionnaire will take approximately 30 minutes.

#### **Possible risks and benefits**

There are no real risks involved in this study. You may find some of the questions offensive. Please know that your answers will not be used to endorse any of the behaviours in the questions. Upon completion of the survey you will receive 1 SRPP point as compensation for your time.

#### **Voluntary participation**

Participation in this study is entirely voluntary. You are free to withdraw from the study at any time.



## **Confidentiality**

Your responses to the survey will be kept confidential. The information obtained will not be disclosed to anybody other than the researchers involved. Any reports or publications about this study will not identify you or any other study participant. The online account used to create the survey will be password protected to protect your data.

## **Questions**

Any questions or problems related to this study can be addressed to the following researchers:

Sarah Westwood  
wstsar002@myuct.ac.za  
084 568 5723

Prof. Johann Louw  
Johann.Louw@uct.ac.za  
021 650 3414

Should you still wish to participate in this study please complete the following:

I have read the consent form and am satisfied with my understanding of this study, its possible risks, benefits and alternatives. I hereby voluntarily consent to participate in this study.

Name: \_\_\_\_\_

Date: \_\_\_\_\_

## Appendix E

### Stimulus material

#### Scenario 1:

##### **Violation Version:**

Before he passed away, Keith's father told his son to cremate his body. Then he told Keith to do whatever he wished with the remains. Keith decided to snort his dead father's ashes.

##### **Control Version:**

Before he passed away, Keith's father told his son to cremate his body. Then he told Keith to do whatever he wished with the remains. Keith decided to bury his dead father's ashes.

#### Scenario 2:

##### **Violation Version:**

Jenny's family made some poor investments. Then her father lost his job. She wanted to help out, and so she decided to sell her virginity to help pay off family debt.

##### **Control Version:**

Jenny's family made some poor investments. Then her father lost his job. She wanted to help out, and so she decided to sell her jewelry to help pay off family debt.

#### Scenario 3:

##### **Violation Version:**

The servers and bartenders at a wedding are denied tips when the mother of the bride walks up to the bar and casually pockets the money in the tip jar.

##### **Control Version:**

The servers and bartenders at a wedding earn extra tips when the mother of the bride walks up to the bar and casually drops a ten-dollar bill in the tip jar.

**Scenario 4:**

**Violation Version:**

Woolworths decides to hire a rabbi as their new spokesperson for the company's line of pork products.

**Control Version:**

Woolworths decides to hire a farmer as their new spokesperson for the company's line of pork products.

**Appendix F**

**Questionnaire**

What sex are you? Male/Female

**1)** Before he passed away, Keith’s father told his son to cremate his body. Then he told Keith to do whatever he wished with the remains. Keith decided to snort his dead father’s ashes.

Is the behaviour described in this scenario wrong?	Yes	No
Is the behaviour described in this scenario amusing?	Yes	No

**2)** Before he passed away, Keith’s father told his son to cremate his body. Then he told Keith to do whatever he wished with the remains. Keith decided to bury his dead father’s ashes.

Is the behaviour described in this scenario wrong?	Yes	No
Is the behaviour described in this scenario amusing?	Yes	No

**3)** Jenny’s family made some poor investments. Then her father lost his job. She wanted to help out, and so she decided to sell her virginity to help pay off family debt.

Is the behaviour described in this scenario wrong?	Yes	No
Is the behaviour described in this scenario amusing?	Yes	No

**4)** Jenny’s family made some poor investments. Then her father lost his job. She wanted to help out, and so she decided to sell her jewelry to help pay off family debt.

Is the behaviour described in this scenario wrong?	Yes	No
Is the behaviour described in this scenario amusing?	Yes	No

**5)** The servers and bartenders at a wedding are denied tips when the mother of the bride walks up to the bar and casually pockets the money in the tip jar.

Is the behaviour described in this scenario wrong?	Yes	No
Is the behaviour described in this scenario amusing?	Yes	No

**6)** The servers and bartenders at a wedding earn extra tips when the mother of the bride walks up to the bar and casually drops a ten-dollar bill in the tip jar.

Is the behaviour described in this scenario wrong?	Yes	No
Is the behaviour described in this scenario amusing?	Yes	No

7) Woolworths decides to hire a rabbi as their new spokesperson for the company's line of pork products.

Is the behaviour described in this scenario wrong?	Yes	No
Is the behaviour described in this scenario amusing?	Yes	No

8) Woolworths decides to hire a farmer as their new spokesperson for the company's line of pork products.

Is the behaviour described in this scenario wrong?	Yes	No
Is the behaviour described in this scenario amusing?	Yes	No