Applied Analysis 1

Rohan Hore

August 2024

1 Relativism of moral judgements

Certain behaviours such as plagiarism or copying another student's homework, texting while driving or driving while intoxicated, corruption or abuse of public office for private gain, encouraging or offering support for hate-groups, and so on, are assessed by the majority in today's society as improper or immoral to varying degrees. Morality is, in part, a code of conduct for acceptable behaviour of individuals: see `https://plato.stanford.edu/entries/morality-definition` for further discussion and definitions of concepts. The study described below is concerned with morality judgements or assessments made by individuals about the activities of other individuals in certain settings.

Five skeleton scenarios were constructed that could be tweaked in various ways to highlight certain philosophical concepts. Each scenario is a short piece of text labelled by the protagonist, Driver, Construction worker, Footballer, Firefighter and Train rider, whose activity is to be judged by the assessor. The philosophical concepts to be studied are moral luck (ML), culpable causation (CC) identification (Id), hedonic state (HS), and impulsive versus deliberative action (IvD). For each concept, two versions of each scenario were constructed to highlight the distinction associated with that concept.

For example, the Driver scenario involves an individual who is drunk, falls asleep at the wheel and has a serious accident. Moral luck arises in the contrast between outcomes: in version ML-A, the driver hits a tree and there are no serious injuries; in version ML-B , the driver strikes a pedestrian who dies at the scene. The hedonic state concept arises in the contrast between attitudes following the event: in version HS-A, the driver is remorseful following the accident; in version HS-B, the driver shows no remorse for his actions. Culpable causation arises in the contrast between causes or reasons for the activity: in one version the driver is on his way home to conceal a stash of marijuana from his parents; in the other to conceal a present that he had bought for their wedding anniversary. Moral luck is a contrast solely between outcomes; hedonic state is a contrast between mental attitudes following the event, and culpable causation is a contrast between reasons or purposes leading to the event. Each version is a short piece of text called a vignette, so there are five scenarios, five philosophical concepts, and 25 combinations illustrated by 25 pairs of vignettes.

One of the issues that arises here is the extent to which moral luck, culpable causation and so on, play a role in our assessments of morality. Is driver ML-B judged by society to be morally more reprehensible because his or her actions led to the death of a pedestrian, while the same actions by ML-A having relatively less serious consequences are judged to be less reprehensible? Similar questions arise for culpable causation, and the other philosophical constructs.

For technical discussion of philosophical terms, see the Stanford Encyclopedia of Philosophy, which is available online. The entry for moral luck commences as follows: Moral luck occurs when an agent can be correctly treated as an object of moral judgment despite the fact that a significant aspect of what she is assessed for depends on factors beyond her control. Bernard Williams writes, "when I first introduced the expression moral luck, I expected to suggest an oxymoron (Williams 1993, 251). Indeed, immunity from luck has been thought by many to be part of the very essence of morality. And yet, as Williams (1981) and Thomas Nagel (1979) showed in their now classic pair of articles, it appears that our everyday judgments and practices commit us to the existence of moral luck." The problem of moral luck arises because we seem to be committed to the general principle that we are morally assessable only to the extent that what we are assessed for depends on factors under our control (call this the Control Principle). At the same time, when it comes to countless particular cases, we morally assess agents for things that depend on factors that are not in their control. And making the situation still more problematic is the fact that a very natural line of

reasoning suggests that it is impossible to morally assess anyone for anything if we adhere to the Control Principle. (Stanford Encyclopedia of Philosophy `https://plato.stanford.edu/entries/moral-luck/`).

For this study, 1068 participants were recruited for an online survey at the website Amazon Mechanical Turk. Each participant was asked for an assessment of the morality of the protagonist's activity in five pairs of vignettes, i.e., ten morality assessments by each participant. The menu of pairs was selected at random and independently for each participant subject to the constraint that each philosophical concept and each scenario be included exactly once. Morality assessments were made on a 1–7 scale, with 1 labelled 'Not at all immoral' and 7 labelled 'Extremely immoral'. For examination purposes, the scale is to be treated quantitatively.

For a randomly selected subset of 525 participants, labelled joint evaluation, the two versions of each vignette were presented side-by-side on the screen, and the response was entered simultaneously as an ordered pair. The five pairs were presented in random order, which was recorded. For the remaining 543 participants labelled single evaluation, or sequential evaluation, the ten vignettes were presented one at a time sequentially in random order. Participants were not given the opportunity to revise an earlier response. For a discussion of the difference between joint and separate evaluation, see the 1996 paper *The Evaluability Hypothesis: An Explanation for Preference Reversals between Joint and Separate Evaluations of Alternatives* by C.K. Hsee in *Organizational Behavior and Human Decision Processes 67, 247–257*

Data Reading

```
morality_data=read.table("morality.dat.txt",header=TRUE)
head(morality_data)
```

```
##   subj mode order   scenario concept first Yp Yn age gender  race
## 1    1   je     3 linebacker      hs   neg  5  6  36 Female white
## 2    1   je     1      train     ivd   pos  7  7  36 Female white
## 3    1   je     2       fire      cc   pos  2  6  36 Female white
## 4    1   je     4     driver   ident   pos  7  7  36 Female white
## 5    1   je     5  construct      ml   pos  5  7  36 Female white
## 6    2   se     5       fire   ident   neg  7  6  23   Male white
```

2 Possible Questions

Problem 1

"Is driver ML-B judged by society to be morally more reprehensible because his or her actions led to the death of a pedestrian, while the same actions by ML-A having relatively less serious consequences are judged to be less reprehensible?"– How do you statistically answer this question?

Problem 2

For the moral luck concept only, tabulate the mean differences $Y_p - Y_n$ by assessment mode and scenario. Compute an approximate standard error, and summarize what this table tells you about the effect of assessment mode on moral-luck judgements.

Problem 3

For the moral luck concept only, tabulate the mean differences $Y_p - Y_n$ by assessment mode and first-in-pair. Compute an approximate standard error, and summarize what this table tells you about the effect of assessment mode on moral-luck judgements.

Problem 4

For the subjects that were assigned to morality assessment in joint mode, tabulate the two-way tables of mean differences $Y_p - Y_n$ by philosophical concept, by scenario, and by left-right on-screen presentation order.

Problem 5

For the subjects that were assigned to morality assessment in single mode, tabulate the two-way tables of mean difference $Y_p - Y_n$ by philosophical concept, by scenario and by within-pair order of presentation.

Problem 6

Comment briefly on the similarities and differences in the summary tables for joint-mode versus single-mode assessment. Are there any differences that appear surprising?

Problem 7

Each respondent i provided one pair of assessments for each of the matched pairs of vignettes presented. That means five differences $Y_p - Y_n$ for each respondent. Under what sort of statistical assumptions is it reasonable to treat the differences as independent random variables?

Problem 8

For the joint-evaluation design, let $n = 2625$ be the number of vignette-pair assessments, let D be the n-component vector of differences, and let $\mathcal{X} \subset \mathbb{R}^n$ be the subspace.

$$\mathcal{X} = \text{scenario} * \text{concept} + \text{order} + \text{left-panel+age}$$

where scenario, concept and order are five-level qualitative factors, left-panel is a binary factor, and age is a quantitative variable. Fit the additive iid random-effects model in which the mean lies in $\mathcal{X}$, and

$$\text{cov}(D) = \sigma_0^2 I_n + \sigma_1^2 B$$

is a linear combination of two matrices in which B is the block factor, or indicator matrix, for assessors. Report the REML variance-component estimates (not necessarily positive). Report also the fitted regression coefficients for age and left-panel with standard errors.

Repeat this analysis for the vector of sums $Y_p + Y_n$. Comment on any major differences between these analyses.

Problem 9

If we consider the additive random-effects model directly to the $2n$ undifferenced assessments, what do you expect the correlation to be for two assessments made by the same subject? Give a numerical value and explain your reasoning.

Problem 10

The pairs of vignettes in this design were coded one pos the other neg, implying that the set of vignettes is a triple Cartesian product.

$$A = \{5 \text{ scenarios}\} \times \{5 \text{ concepts}\} \times \{\text{pos, neg}\}$$

Some of the analyses suggested in previous questions are based on the assumption that the correspondence between the actual levels and the Cartesian product set is natural. In fact, the actual vignette labels for the concept ML are lucky/unlucky, while those for IvD are impulsive/deliberative, and so on, with contrasting pairs of adjectives describing the other concepts. One might guess that lucky has been coded for administrative purposes as pos, and unlucky as neg, but the coding for IvD and other concepts is less obvious, and perhaps even entirely arbitrary. If that is the case, the set of vignettes is a two-fold Cartesian product

$$B = \{5 \text{ scenarios}\} \times \{\{\text{lucky, unlucky}\}, \cdots, \{\text{implusive, deliberative}\}\}$$

where the second factor is a partition of ten adjectives into five pairs, one pair for each concept. In that case, no 1–1 correspondence $A \iff B$ is natural in the sense of preserving Cartesian or other relationships among elements. Show that some of the models that were fitted and tested are not independent of the assignment of labels pos and neg to vignette pairs. Which ones?

Explain the statistical implications of dependence on administrative coding. For the models that are not independent of the label assignment, suggest a modification to remedy the problem.

Problem 11

For the difference vector, are the effects of order, age, sex and race of the assessor appreciable? Start with any reasonable baseline model of your choice for comparison. What are the implications of this for generalizability of the conclusions?