PATRICK SUPPES

SOME FORMAL MODELS OF GRADING PRINCIPLES*

I. INTRODUCTION

The present paper offers an analysis of grading principles from the viewpoint of statistical decision theory and game theory. The mistaken notion is widely held that the plain man is really clear about practical ethical and moral issues and that philosophers need only tidy up certain wayward corners of the subject. Personally I find difficult the problem of devising any general ethical rules of behavior for simple two-person games; the ethical complexities of progressive taxation, tariff barriers, or treatment of sexual psychopaths are beyond any exact conceptual analysis. That decisions are and must be made about these issues no more proves that their ethical aspects are completely understood than does the fact that the Romans built bridges prove that they had any quantitative grasp of the mechanical theory of stress.

It is pertinent to remark that the first model used in this paper is at the basis of much recent foundational work in statistics (see Blackwell and Girshick (1954) and Savage (1954)). The considerations in the last two sections are within the more general framework of the theory of games as developed by von Neumann and others. My particular concern is the embedding in this framework of a theory of two-person justice.

II. INDIVIDUAL DECISION MODEL

The structure of the first model to be considered is simple. We shall call an

* I am indebted to Richard Brandt, Donald Davidson and F. Studnicki for a number of useful and penetrating criticisms of a much earlier draft of this paper written in 1957 and circulated as a technical report in that year under the title, `Two formal models for moral principles'.

† Kant's views are typical: "... in matters which concern all men without distinction, nature cannot be accused of any partial distribution of her gifts; and that with regard to essential interests of human nature, the highest philosophy can achieve no more than that guidance which nature has vouchsafed even to the meanest understanding" (1949a, p. 666).
ordered triple $S = <S, C, D>$ an individual decision situation when $S$ and $C$ are sets and $D$ is a set of functions mapping $S$ into $C$. The intended interpretation is:

- $S =$ set of states of nature,
- $C =$ set of consequences,
- $D =$ set of decisions or actions.

Since the terms 'states of nature', 'consequences', 'decisions' and 'actions' are used here in a somewhat special manner, an example may help to make clearer their intended meaning.

**Example 1:** Suppose I come home and find a bottle of ink spilt on the rug, and also suppose I know immediately that it could have been spilt either by my four-year old daughter or by my cat. These two possibilities correspond to the two states of nature. I can take one of two actions, let us say: spank the child or do not spank the child. And the possible consequences are four in number, as illustrated in Table I. The rows correspond to the two states of nature, the columns to the two actions, and the entries in the table to possible consequences.

<table>
<thead>
<tr>
<th>states of nature</th>
<th>actions</th>
<th>$a_1$ - spank the child</th>
<th>$a_2$ - do not spank the child</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$ - child spilt the ink</td>
<td>$c_1$ - ink spilt by child and child spanked</td>
<td>$c_2$ - ink spilt by child and child spanked</td>
<td></td>
</tr>
<tr>
<td>$s_2$ - cat spilt the ink</td>
<td>$c_3$ - ink spilt by cat and child spanked</td>
<td>$c_4$ - ink spilt by cat and child spanked</td>
<td></td>
</tr>
</tbody>
</table>

Since the term 'states of nature' is not much used in philosophy there should be little objection to its special use here; the term 'action' is used in a way that is consonant with at least one of its major uses in ordinary contexts. But my use of 'consequence' is probably at variance with its primary use in the writings of moral philosophers. The consequence $c_1$ above, for instance, ink spilt by child and child spanked, would be regarded by many as the bare beginning of consequences. It is to avoid