TOWARD A THEORY OF MORAL REASONING

PHILLIP T. MONTAGUE

It is remarkable, considering the importance of answering the question how moral judgments are justified, that philosophers have devoted so little attention to the problem of formulating an adequate theory of moral reasoning. Though some writers have attempted to hang moral justification on a deductive framework, others have pointed out how difficult it is for deductive theories to account adequately either for the existence of moral conflict or for the relation between judgements and actions. Yet these non-deductivists give little hint of an alternative to the deductive model.[1]

In this paper I will attempt to furnish the foundation for a non-deductive theory of moral reasoning. I hope in the process to clarify certain related concepts which have become standard, though not completely digestible fare for the moral philosopher.

I

Few would contest the claim that making a justified moral judgment about an action requires a systematic consideration of the action's morally relevant features. Indeed, the latter statement has been overworked to the point where it seldom represents more than a sonorous phrasal of the truism that justified judgments are not merely guesses. This situation has tended to obscure the fact that not only the structure of moral reasoning, but also the concept of moral relevance is more than a little mysterious. What does it mean to say that an action has certain morally relevant features; and how, from a

[1] The deductivist position (i.e., the identification of cogent moral inferences with deductively valid inferences) is adopted categorically by R.M. Hare. (See especially *Freedom and Reason* [Oxford: The Clarendon Press, 1963], p. 88). Opposing stands have been taken by W.D. Ross *(The Right and the Good* [Oxford: The Clarendon Press, 1930]), G.E. M. Anscombe *(Intention* [Ithaca: Cornell University Press, 1957]), and M.G. Singer (*Generalization in Ethics* [New York: Alfred A. Knopf, 1961]).

Diálogos, 32 (1978) pp. 19-29.

logical standpoint, does the presence of a morally relevant feature in an action pertain to the question whether the action is right, wrong, obligatory, etc.? We might agree, for example, that being an instance of promise-keeping is relevant to whether an action is obligatory. We might also want to say that the relevance of being the keeping of a promise, together with the fact that some action A is an instance of promise-keeping, counts as a reason —though not a conclusive one— for thinking that A is obligatory. What we lack, however, is any sort of systematic account of these "non-conclusive" moral reasons— reasons which in some sense support but do not entail moral conclusions. Both their logical form and their logical ties to the conclusions they support are in need of specification.

One might claim, of course, that all moral reasons are conclusive (i.e., deductive), that the "premises" of cogent moral arguments entail the "conclusions" of those arguments. But this claim must be backed by a theory of moral reasoning which not only is purely deductive, but also exhibits the way moral reasoning involves considerations of moral relevance. Indeed it seems to me to be a primary task of a theory of moral reasoning to give substance to the notion articulated above that making a justified moral judgment about an action requires a systematic consideration of the action's morally relevant features. Whether purely deductive theories are equal to this task can only be determined by evaluating individual theories as they are proposed. But no deductive theory with which I am familiar deals adequately —if at all— with the concept of moral relevance.

Although moral philosophers (including those inclined to view moral reasons as non-deductive) have in general failed to provide any adequate *theoretical* account of what it is to be a morally relevant feature of actions, many writers have emphasized the latter concept on an informal level. Often motivating these writers is a devotion to "individual moral autonomy," an eschewing of "antecedently accepted moral principles," and an abhorrence of "legalism" in moral practice. The point seems to be that morally sensitive and mature persons make their moral decisions not by mechanically applying highly general principles which they have in hand on approaching the decision situation, but by assessing the situation in terms of its own peculiar features and in some sense creating principles tailored to fit each situation. Only in this way (so the argument runs) can one be said to make genuinely autonomous moral decisions.[2]

[2] I am assuming that the sort of position just outlined is a fairly familiar one. Parts of it have been supported by proponents of "situation ethics" and by

This line of thinking, though without question containing a core of truth, seems to me to present a picture of moral reasoning which is, to say the least, misleading.

Clearly, the morally sensitive and mature person does not confront decision situations empty-handed. In order for him to recognize morally relevant features of actions he must believe that certain features are, while others are not, morally relevant. Moreover, if these features are morally relevant in one situation they are morally relevant in all situations —or in any case there are other features which *are* applicable in all situations, and which tell us when features are morally relevant. But this implies that antecedently accepted principles —i.e., principles of moral relevance— do have a role in arriving at reasoned moral decisions. One must be careful, then, to specify the sort of principle the consideration of which he wants to exclude from the moral deliberations of the mature individual.

Proponents of the view we are examining would no doubt counter the above remarks with the claim that their concern *is* with principles of a certain type. These principles are exemplified by "Promises ought to be kept" and "It is wrong to hurt people." It is principles like these which allegedly have no real place in arriving at reasoned moral decisions.

But I think we must ask what precisely could be meant by this last claim. For example, does it imply that the principles in question are false? Would one really want to deny that it is wrong to hurt people? Probably not, though one might well hesitate to answer this question until a way of *interpreting* the disputed principles is provided.

It seems to me that attempting to construct a theory of moral reasoning on the unqualified assumption that principles like "It is wrong to hurt people" have no legitimate place in moral reasoning, involves an intolerable amount of philosophical legislation. Moreover, moral autonomy can be safeguarded and legalistic views of moral reasoning avoided without making so strong an assumption. What we must avoid are *certain interpretations* of the principles in question, which interpretations are dictated by adherence to a deductive model for moral reasoning. Consider, for example, the principle "It is wrong

some existentialists, as well as by philosophers who might be loathe to identify with either of these two points of view. Prominent among the latter is R.M. Hare *(Freedom and Reason)*, though we can also cite W.J. Rees ("Moral Rules and the Analysis of 'ought' ", *The Philosophical Review*, LXII [1953], p. 27) and Thomas D. Perry ("Moral Autonomy and Reasonableness", *The Journal of Philosophy*, LXV [June 27, 1968], pp. 383-401).

to hurt people". One who accepts the principle and assumes that it is a reason for believing "A, which is an instance of hurting someone, is wrong" only if the former entails the latter, might very well interpret the principle as "If *x* is an instance of hurting someone, then *x* is wrong." And this sort of position would promote the rigid, legalistic view of moral reasoning which devotees of moral autonomy reject. The situation appears in an entirely different light, however, if we give up the notion that all cogent moral inferences are deductive. The problems in question can be avoided by formulating a model for moral reasoning within which reasons can *support* conclusions while others can simultaneously support contrary conclusions. This would allow for the possibility of, in some sense, "weighing" *conflicting* reasons. As was indicated above, such a model is necesary if principles of relevance are to be counted as moral reasons. What I am now claiming is that a non-deductive model is also necessary to do justice to the role in moral reasoning played by principles like "It is wrong to hurt people."

Some of the foregoing points can more clearly be made if we state more precisely what is involved in formulating a theory of moral reasoning. The latter will be construed here as an analysis of the relation "*t* is a reason for believing *s*," where *s* and *t* are sentences, and *s* expresses a moral proposition. In rejecting a deductive model for moral reasoning I am denying that for *t* to be a reason for believing *s*, *t* must entail *s*. Reasons for rejecting this implication are scattered throughout the preceding discussion. They range from the claim that deductive theories are unable to account either for moral conflict or the relation between reasoned moral judgment and actions,[3] to the claim that such theories leave no room for principles of relevance in moral reasoning. It was also asserted that in order to show how principles like "It is wrong to hurt people" and "One ought to keep his promises" function as moral reasons a non-deductive theory of moral reasoning is required. None of these considerations, however, is being put forward as a knock-down argument against deductive theories of moral reasoning. All of them depend on judgments concerning the ability of such theories to account *adequately* for certain aspects of moral reasoning. I am simply claiming that one can better handle these tasks with a non-deductive theory.

To say that a theory of moral reasoning must specify how the

[3] The relation between these two problems and that of specifying the nature of moral reasoning is set forth clearly in an unpublished paper by Professor Donald Davidson entitled, "How is Weakness of the Will Possible."

two types of principles discussed above can count as moral reasons is to require that "*t* is a reason for believing *s*" be defined so that, e.g., (a) a sentence expressing the relevance of being an instance of hurting someone to being wrong conjoined with "A is an instance of hurting someone" is a reason for believing "A is wrong"; and (b) "It is wrong to hurt others and A is an instance of hurting someone" is a reason for believing "A is wrong." It might be worth emphasizing here that our concern is with a relation between sentences, and not with the truth values of possible arguments of the relation. We cannot expect a theory of moral reasoning either to provide us with true premises for moral arguments or to guarantee the truth of conclusions of these arguments.

II

Pervading the preceding remarks is the view that both beliefs about moral relevance and beliefs about the rightness, wrongness, obligatoriness, etc., of certain sorts of actions play a legitimate and significant role in moral justification. A task facing anyone wishing to give a theory of moral reasoning is to define the logical status of such beliefs, both with respect to each other and with respect to the conclusions they support.

When we speak of morally relevant features of actions, we are concerned with features the presence of which in an action is relevant to whether the action is, e.g., right, wrong, obligatory or forbidden. Recognition by an individual that an action he is thinking of performing possesses one of these features places him in the realm of *moral* decision. If his deliberation over whether to perform the action is affected appropriately by this recognition, then he is engaging in *moral* reasoning. But employing considerations of moral relevance in moral reasoning involves more than mere awareness that certain features are relevant while others are not. This is obvious from the fact that, e.g., we consider both an action's being an instance of promise-keeping and its being an instance of promise-breaking as relevant to whether the action is obligatory, though in different ways. This difference can be characterized by saying that the presence of one feature counts towards the action's being obligatory, while the presence of the other counts against this, and in fact towards the obligatoriness of refraining from performing the action.

It is possible, then, to distinguish three types of relevance arising from the notion of morally relevant features of actions. One is relevance *simpliciter;* one we may refer to as "positive relevance";

and the other can be labeled "negative relevance." All three types of relevance will be construed here as relations between sentences or sentential functions, and will be symbolized respectively as *rel(s, t)*, *pr (s, t)*, and *nr(s, t)*. Thus, for example, we can express the relevance *(simpliciter)* of being an instance of promise-keeping to being obligatory as *rel*("*x* is obligatory," "*x* is an instance of promise-keeping").

According to the way positive and negative relevance are being interpreted here, to say that the presence of a feature in an action is *negatively* relevant to its being e.g. obligatory, is to say that the presence of the feature is *positively* relevant to the action's nonperformance being obligatory. This allows us to define negative relevance in terms of positive relevance as follows:

(1) *nr*("*x* is obligatory," "*x* is K") if and only if *pr*("$\bar{x}$ is obligatory," "*x* is K).

where "obligatory" stands for any appropiate moral predicate, K refers to a morally relevant feature of actions, and "$\bar{x}$ is obligatory" means "Refraining from performing *x* is obligatory."

We will also stipulate that

(2) *rel(s, t)* if and only if *pr(s, t)* or *nr(s,t)*

and take the relation *pr(s, t)* as an undefined concept for the theory presented here. A second undefined notion required for our theory can be expressed by the relation "The consequences of performing action A are better than the consequences of performing action B."

A theory of moral reasoning was defined in the preceding section as an analysis of the relation "*t* is a reason for believing *s*," where *s* and *t* are sentences and *s* expresses a moral proposition. If we abbreviate this relation as *r(s, t)* we might want to say, for example, that *r*("A is obligatory," "Promise-keeping is obligatory and A is an instance of promise-keeping"). Substitution instances of *r(s,t)* represent arguments and, I will claim, express analytic propositions the truth values of which are independent of the truth values of any sentences filling the argument places of the relation. But substitution instances of *pr(s, t)*, express not analytic propositions but synthetic moral principles. They do not represent moral arguments though they may appear among the premises of such arguments.[4]

[4] It is not uncommon to see in the literature statements like "An action's being an instance of promise-keeping is a *reason* for believing that the action is

The above interpretation of *r(s, t)* and *pr(s, t)* and their relation to each other bears strong similarities to a plausible way of construing confirmation ("logical probability"), statistical probability and one way they are related. We can, for example, express the claim that the statistical probability that an occurrence of an A is an occurrence of a B is greater than one-half as *p*("*x* is B," "*x* is A") > 1/2 —a statement which is synthetically true or false. And we can use this fact as part of an inductive argument expressed as the confirmation of one sentence by another: *c*("*b* is B," "*p*('*x* is B,' '*x* is A') > 1/2 and *b* is A")— which can be viewed as expressing an analytic proposition the truth value of which is independent of the truth values of the terms of the relation.[5]

One motive for eschewing attempts to define inductive reasoning in terms of deductive reasoning is the apparent impossibility of reflecting, within a deductive framework, that aspect of inductive reasoning which involves accounting for and "weighing" relevant and often conflicting evidence. This last feature is shared by moral reasoning and was stressed in my earlier remarks opposing attempts to define moral reasoning on a deductive model. It should be noted, however, that in emphasizing certain similarities between inductive and moral reasoning I am not affirming that moral arguments are in any sense a species of inductive argument. Nothing said here hangs on a decision regarding this issue.

III

According to the definition given above, an analysis of *r(s, t)* —of the relation between the premises and conclusion of cogent moral arguments— would constitute a theory of moral reasoning. Rather than confront here the task of constructing such a theory in its

obligatory (or a reason for performing the action)". One also finds reference to the relevance of the premises of an argument to its conclusion. My insistence on the uses of "relevant" and "reason" described in the text is based not on a belief that alternative uses are inappropriate or incorrect, but merely on a desire to emphasize a distinction which I take to be a conceptual one. For example, in "An action's being an instance of promise-keeping is a reason for believing the action is obligatory" and in "The premises of a cogent argument are reasons for accepting its conclusion," different concepts of being a reason are employed. A similar remark applies if the two statements are expressed in terms of relevance rather than reasons.

[5] Cf. Carl G. Hempel, "Deductive-Nomological vs. Statistical Explanation," *Minnesota Studies in the Philosophy of Science*, ed. Herbert Feigl and Grover Maxwell (Minneapolis: University of Minnesota Press, 1962), III, pp. 128-149; Rudolph Carnap, *The Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950), pp. 19-36.

entirety, I will restrict my remarks to arguments with conclusions of the form "A is obligatory," "A is right," "A is forbidden," etc., where A is an action. These arguments include common forms of the standard "practical syllogism," and have traditionally constitutted the main focus of interest and controversy among philosophers concerned with moral reasoning.

It will be convenient at this point to introduce a technical term, and also some abbreviations for certain expressions to be used below:

(a) If *s* is a sentence of the form "A is obligatory," then a "syllogism sentence for *s* concerning A" (abbreviated SA) is a sentence of the form "*pr(*'*x* is obligatory,' '*x* is K') and A is K." A syllogism sentence for "Refraining from A is obligatory" (abbreviated SA) is "*pr*('*x* is obligatory,' '*x* is K') and refraining from the A is K". Syllogism sentences are such that the conjunction of two or more syllogism sentences for *s* and A is also a syllogism sentence for *s* and A.

(b) The expression "The consequences of performing A are better than the consequences of performing B" will be abbreviated as C (A, B).

We can now define *r(s, t)*, under the restrictions on its domain given above, as follows:

r(s, t) if and only if any one of the following conditions is satisfied:

1) *t* entails *s*;
2) *t* is equivalent to SA;
3) *t* is equivalent to SA & S$\bar{\text{A}}$ & C(A, $\bar{\text{A}}$)
4) *t* is equivalent to *u*, and *r(s, u)*
5) *s* is equivalent to *u*, and *r(u, t)*

As examples of each of the above conditions we can cite

1') If "A and B are obligatory" entails "A is obligatory," then the former is a reason for believing the latter; i.e., *r*("A is obligatory," "A and B are obligatory".

2') "*pr*('*x* is obligatory,' '*x* is an instance of promise-keeping') and A is an instance of promise-keeping" is a reason for believing "A is obligatory."

3') "*pr*('*x* is obligatory,' '*x* is an instance of promise-keeping') and A is an instance of promise-keeping and refraining from A is an instance of promise-keeping and the consequences of A are better than the consequences of refraining from A" is a reason for believing "A is obligatory."

4') If "*pr*('*x* is forbidden,' '*x* is an instance of promise-breaking') and refraining from A is an instance of promise-breaking" is equivalent to "*pr*('*x* is obligatory,' '*x* is an instance of promise-keeping') and A is an instance of promise-keeping," then, since the latter is a reason for believing "A is obligatory," so is the former.

5') If "A ought to be done" is equivalent to "A is obligatory," then since "*pr*('*x* is obligatory,' '*x* is an instance of promise-keeping') and A is an instance of promise-keeping" is a reason for believing the latter, it is also a reason for believing the former.

Because of the rule that the conjunction of several syllogism sentences for a sentence *s* and action A is also a syllogims sentence for *s* and A, it follows from conditions 2) and the definition of a syllogism sentence that "*pr*('*x* is obligatory,' '*x* is an instance of promise-keeping.) and A is an instance of promise-keeping and *pr*('*x* is obligatory,' '*x* is refraining from stealing') and A is refraining from stealing" is a reason for believing "A is obligatory."

A complication arises, however, when we are dealing with "competing" syllogism sentences as is the case in situations of moral conflict. Given a syllogism sentence for "A is obligatory" and one for "Refraining from A is obligatory", it is not possible in general to determine that the conjunction of the two syllogism sentences is a reason for believing either "A is obligatory" or "Refraining from A is obligatory."[6] But according to condition 3) of the definition of *r(s, t)*, the conjunction of two such syllogism sentences with "The consequences of A are better than the consequences of refraining from A" *is* a reason for believing "A is obligatory." This result seems to me to reflect the thinking of many moral philosophers concerning the question how to resolve moral conflict.[7]

Conditions 4) and 5) provide us with a way of accounting, within our theory, for the more familiar type of moral principle referred to in Section I. This type of principle is exemplified by "Promise-

[6] If an action is an instance of killing as well as an instance of promise-keeping, we might want to say that knowing this is sufficient to decide that refraining from the action is obligatory. But in referring to an action as an instance of killing, we are apparently describing it in terms of its consequences and, if so, are implicitly invoking condition three. The definition of *r(s, t)* is neutral with regard to this issue, however.

[7] See, for example, Stephen Toulmin, *Reason in Ethics* (Cambridge: Cambridge University Press, 1961), p. 147; William K. Frankena, *Ethics* (Englewood Cliffs, N.J.: Prentice-Hall, 1963), p. 42.

keeping is obligatory" and "It is wrong to hurt people." I will claim here that such principles are equivalent to principles of positive and negative relevance —that, e.g., "Stealing is forbidden" is equivalent to "*pr(x* is forbidden,' '*x* is an instance of stealing')." Thus "Stealing is forbidden and A is an instance of stealing" is a reason for believing "A is forbidden".

The above claim concerning the equivalence of the two types of moral principles is, however, far from unproblematic. On an informal level, we must face the fact that e.g. "Stealing is forbidden," seems in some sense to say more than does "An action's being an instance of stealing is positively relevant to the action's being forbidden." And if we agree —as seems reasonable— to formalize "Stealing is forbidden" as a universal conditional, then this principle, conjoined with "A is an instance of stealing," evidently entails "A is forbidden." This result would imply that our theory of moral reasoning is really a deductive theory in disguise, and we would be faced with all the problems which plague such theories.

What we must do, I think, is accept the well-worn distinction between "*prima facie*" and "strict" (or "actual") moral concepts, and state that our two types of principles are concerned with different kinds of moral concepts. For example, in the principle "Promise-keeping is obligatory," "obligatory" expresses the concept of *prima facie* obligation. But in "*pr*('*x* is obligatory,' '*x* is an instance of promise-keeping.)," "obligatory" expresses the concept of strict obligation. Given the equivalence of the two sorts of principles, we can define "*x* is *prima facie* obligatory (right, wrong, etc.,)" as "*x* possesses a feature positively relevant to its being strictly obligatory (right, wrong, etc.)."

Singular moral statements such as "A is obligatory" can be concerned either with *prima facie* or strict moral concepts. If in, say, "A is obligatory" "obligatory" is *prima facie*, then this sentence is entailed both by "Promise-keeping is obligatory and A is the keeping of a promise" and by "*pr*('*x* is obligatory,' '*x* is an instance of promise-keeping') and A is an instance of promise-keeping." Thus both the latter are reasons for believing the former. But neither of the second two sentences entails, though each is a reason for believing, "A is obligatory" if "obligatory" in this sentence is strict.

Given these last claims, however, one might feel compelled to comment as follows: it is all very well to know what goes on *within r(s, t)*, to know what to count as reasons for accepting singular moral sentences; but we must know more than this —we must know whether such sentences are true. Thus (so this argument might run), we must at some point be able to detach singular moral sentences for the reasons we have for believing them, and in so doing employ deductive reasoning to arrive at our moral conclusion. This fact implies that the non-deductive theory of moral reasoning presented above is at best incomplete and may even be totally misconceived.

In confronting this argument we must of course agree that knowing what to count as reasons for accepting singular moral sentences falls far short of knowing whether such sentences are true. To achieve the latter one must be able not only to recognize moral reasons but also to accept them, to "weight" them appropriately, and to have support for them. But this does not at all imply —as the argument presented above suggests— that moral knowledge requires the presence of *conclusive* reasons. If someone knows that a singular moral sentence *s* is true then (a) *s* is true, and (b) the person accepts some sentence *t* such that *r(s, t)*, as part of his being justified in believing *s*. Even if we also require *t* to be true, however, we cannot insist that the truth of *s follow* from the truth of *t* and condition (b) if we are to preserve the distinction between knowledge and justified belief. Indeed it seems to me that it is in virtue of blurring this distinction that the argument being considered here gains whatever plausibility it might possess.

IV

As was emphasized above, it is one thing to ask what to count as a reason for believing that some action A is (say) obligatory, quite another to ask when someone is justified in believing that A is obligatory. To believe with justification that A is obligatory, not only must one hold a belief which is a reason for believing that A is obligatory, but he must also have a reason for accepting that reason. This leads, in virtue of the role played by moral principles as reasons for believing singular moral sentences, to the question what we should regard as reasons for accepting principles. To answer this question would be to extend our definition of *r(s, t)*, allowing moral principles as values for *s*. And, if we are concerned with providing some sort of "ultimate justification" for moral principles, with avoiding an infinite regress of moral reasons, the obvious candidates for reasons for accepting moral principles will not suffice. We could not rest content, for example, with allowing as reasons for believing "Actions which are K are obligatory" either "Actions which are H are obligatory and K actions are H" or "Action A is K and A is obligatory." What sorts of reasons will serve our purposes, however, is an issue with which I am presently unprepared to deal.

Western Washington State College