RULES AS NESTS OF BELIEFS AND INTENTIONS

By Mark Fisher

Von Wright ([1]) has pointed out the relevance to legal realism of Anderson's proposal ([2]) that norm-statements can be reduced to alethic-necessity statements with a constant to be interpreted as a sanction. He argues that, although it is unsatisfactory to account for the existence of a norm in a society as the probable visiting of sanctions upon detected breaches of the norm, a satisfactory account can nevertheless be given in terms of necessary and sufficient conditions and the concept of liability, or of its contradictory, immunity. And he maintains that «the notion of liability can be drained of all 'deontic' content and used in a purely factual or naturalistic sense to mean that, under given circumstances, a punitive may or may not be consequent upon a certain type of action.» (93) He does not show how this is to be done, however, and I think it is fair to say that his argument showing that a distinction must be made between the existence of a norm in a society and the probability (which may be very low) of a detected offender's paying the legal penalty, also weighs heavily against his own proposal to define norm-existence in terms of liability to punishment if detected and tried — at least if liability is, as he suggests, a naturalistic concept.

If legal realism is to be defended, then, a different account of norm-existence must be given. It seems to me that the error of the usual legal-realist doctrine lies in its attempt to define an intentional concept in terms of non-intentional ones. A norm's existence cannot consist merely in the fact that persons *do* certain things; the existence of a norm that everyone shall ø, even, cannot consist in the fact that failure to ø is in reality commonly visited with a sanction. For it to *be* a sanction it must be visited upon the offender, or upon someone,

([1]) Von Wright (1969), 89-91.
([2]) Anderson (1958).

because the norm was, or was thought to have been, breached. And to describe *this* connexion between an act and a sanction we have to give the sanction-inflicter's reasons. This too is why an Anderson-type norm whose entailed sanction is a mere natural occurrence, such as an earthquake, cannot be a legal norm or any other sort of rule that persons make and try to follow.

It might now be concluded that legal realism is false. However, it does seem worthwhile to try to reformulate the doctrine in a more easily defensible way — at least, if the only alternative to it is the view that, as Von Wright puts it, there is a «realm of oughtness,» and that «the reality of norms is *sui generis*». Von Wright's suggestion is that «the reality of norms resides in some empirical facts about human and societal behaviour.» (89) Accepting this, but finding the behaviour of either legal or enforcing authorities unacceptable candidates for the role, I propose a different one and defend it in section III. The broadly naturalistic thesis of Von Wright, which I am going to defend, has been rejected by many modern logicians and moral philosophers who have insisted that «no 'ought' from 'is' can be derived.» A proposal to analyse norm-existence in terms of certain facts about persons must be able to meet their objections. For instance, a version (or echo) of Moore's refutation of the naturalistic fallacy can be put as a counter-example to my analysis. I defend it against this and other objections in section V.

The proposal I shall make is not intended to resolve any outstanding problems in deontic logic. Rather, I take it to be a desideratum of any analysis of the concept of rule-existence that it be possible to show how the main principles of deontic logic be at least compatible with it, and preferably deducible from it. In particular, I shall require that this be possible for the principles *COpPp, EOKpqKOpOq and CPApqAPpPq*. And it must always be possible for *Op* to be true while *p* is false.

In section II I discuss the concepts of belief and intention to be employed in the analysis, and introduce some notation.

II

The concept of belief or expectation, though obscure in many ways, is at any rate intentional. That a person believes something is a fact about his mind — though admittedly neither of these ways of putting the matter really clarifies the other. The same can be said of the concept of intention itself. An analysis of norm-existence in terms of the intentions and beliefs of persons should, therefore, escape at least from the objection previously raised to legal realism, that it analyses an intentional concept into non-intentional ones.

There are certain formal analogies between «*a* believes that *p*» and «it ought to be the case that *p*», which have been elaborated by Hintikka ([3]). For present purposes only a few features of the logic of such statements need be sketched — most of them, I believe, non-controversial. First, it is true of belief as it is of norms that the truth of what is believed (the occurence of what is required) does not follow from its being believed (required): neither *CBapp* nor *COpp* are principles of logic. But we do have *CBapNBaNp* just as we do have *COpNONp*. This principle about belief has, however, been questioned. For it apparently asserts (when the tacit universal quantifier is read in) that no-one holds to both members of any contradictory pair of beliefs, and people are notoriously fond of doing that. However, Hintikka's defence is that he is working out the principles of *rational* belief: his logical theses describe not what does happen but what a person must believe insofar as he is rational. We shall be employing this concept of rational belief in our analysis.

It has not been so widely appreciated that these formal analogies between belief and norm also hold for intention. If we write *Iap* to mean «*a* intends that it be the case that *p*,» then the parallel principle *CIapNIaNp* seems to hold good of *a*'s rational intentions, while evidently *CIapp* does not. We shall employ this parallel concept of rational intention in our analysis.

([3]) HINTIKKA (1962).

It is evidently part of the concept of belief and intention that rational belief and intention are in a certain way ideal: the concepts could not function as they do if the principles *CBapNBaNp* and *CIapNIaNp* had no regulative force at all. In a similar way, we shall try to spell out what it is for a rule to exist, in such a way that the fact that there can be rules even when the factual conditions we propose as our analysis do not *all* exist, need not prove decisive against our analysis. That is, our analysis will be an idealization, in the way that Hintikka's doxastic logic is an analysis of an idealised concept of belief.

Some notation will now be useful as abbreviating devices. (No interesting theorems will be looked for.) To say that there exists a rule in a community **C** that one ought to (must, is expected to) ø, we write ΠxOøx. (Note that no interpretation is offered of OΠxøx, which will be taken to be interchangeable with ΠxOøx.) Following Hintikka we write *BaBbøc* to mean that *a* believes that *b* believes that *c* is ø (øs). For our purposes *a, b, c* are the names of persons and ø, ψ, are predicates of action such as 'kills a person' or 'marries'. Similarly, *IaIbøc* will mean that *a* intends that *b* intend that *c* ø. In this first part of our analysis the universe of quantification is in effect the community **C**. Finally, we write *Dap* as a sentence-form that becomes a sentence when *D* is replaced by either *B* or *I*.

III

If ΠxOøx is true in **C**, it must at least be true that Πxy*B*xøy and Πxy*I*xøy, i.e. Πxy*D*xøy. For suppose people in **C** do not generally expect anyone to ø; suppose a person, if asked whether someone (perhaps he himself) will ø or not, usually replies that perhaps he will, perhaps he won't. Now, although there could be exceptional circumstances in which there did exist the rule in question although people did not have the expectation, this could not be the normal case. For if it were, then the information that a rule existed in a community would lose its utility. Again, suppose it is generally false that people intend, or intend others, to ø. This might exceptionally be true

in a community in which there was the rule ΠxOøx, but it could not be generally true of such communities. First, it could not be generally true of them that the common and accepted answer to the question «Do you intend to ø ?» was «I may or I may not — I have no intention in the matter.» For if it were, then the information that there existed such a rule would be empty. In this case it would also be found that the utility of having the rule would be lost. Second, less obviously, the same holds good if no-one intends anyone else to ø. For, suppose that in a certain community Πxy*B*xøy and Πx*I*xøx are true but it is false that Πxy*I*xøy: people expect everyone to ø and intend to ø themselves but generally don't intend anyone else to ø. In this community each person could, perhaps, truly say «It is *my* rule to ø,» and would expect others to ø. But, because no-one intended anyone *else* to ø, no-one could say that *his* rule was binding on others. Thus each might have his own rule; the community's behaviour might be regular; but the community would have no rule.

I am giving something like a contract theory of what it is for a community to have a rule. I am suggesting that for the rule to exist there must be, besides individual intention to follow it, and expectation (even well-grounded expectation) that others will follow it, some further «social glue» that binds the individuals together. But the glue is no mystery-substance: it consists (in part) of the individuals' intending that others shall follow the rule.

Perhaps this notion of one person's intending that another shall act is not much less puzzling than that of rule-existence. But it may be easier to clarify, for it does at least depend on the simpler notion of intending to act. And it may throw some light on the attractive yet unsatisfactory suggestion that a rule is a command. For the idea that everyone commands everyone to ø, while literally absurd, does seem meant to imply that everyone intends everyone to ø. And our account has the additional merit of explaining what is meant by «binding» when it is demanded of an account of the nature of norms, rules, or moral principles that it show how they are binding.

Is our account of what $\Pi x O \emptyset x$ means now sufficient ? So far we have analysed it as $\Pi xy D x \emptyset y$. But more seems to be required: the glue binding the community to ø is a more complicated substance. The sort of complexity I have in mind has recently been noticed by philosophers interested in such apparently disparate concepts as role ([4]), meaning ([5]), and sexuality ([6]). For a community to have a rule, its members must have beliefs and intentions not just about the *behaviour* but also about the *beliefs and intentions* of other members (and about their own). That is, for $\Pi x O \emptyset x$ to be true it must be true not only that $\Pi xy D x \emptyset y$ but also that $\Pi xyz D x D y \emptyset z$. And even this is not enough. For, even in a simple two-person situation, when there is a rule $\Pi x O \emptyset x$ it will generally be true not only that *a* believes *b* believes *a* will ø (and intends that too), but also that *b* believes *a* believes *b* believes *a* will ø. For, suppose this latter belief is generally lacking; informally, that in general *b* doesn't know whether *a* thinks he (*b*) trusts him (*a*) to ø. This seed of doubt (or merely non-belief), if it is general in the community, will constitute a subtle kind of group suspiciousness, whose existence would give *some* reason to doubt whether the community straightforwardly had the rule. The glue would then seem not to be holding properly. Similarly with intention: a variety of crude and subtle forms of hypocrisy can be distinguished. For example, if $Ix I y \emptyset x$ were generally false, we should have a group in which people didn't generally intend (want) others to intend (expect) them to ø.

In discussing roles Mayo suggests ([7]) that the number of layers, as it were, of belief that belief that ... involved in the existence of a role is infinite. (He calls the phenomenon «infinite etceteration.») But, attractive though it may seem as an explanation of the common philosophical notion that norms are different from facts, I do not see how, on that analysis, any norms could ever exist at all. The question then is: how many

([4]) Mayo (1968).
([5]) Grice (1957).
([6]) Nagel (1969).
([7]) Mayo (1968), p. 53.

layers must we allow for? In an ideally rational society perhaps there is no limit. Hintikka suggests ([8]) that a rational man must believe the logical consequences of his beliefs, if he takes them to be such:

CBaKpCpqBaq. Thus anyone who comes to know that a proposition can have an infinite number of consequences «must,» in this sense, believe an infinite number of things. Hintikka also maintains ([9]) that the S4 principle holds for belief, i.e. that *CBapBaBap:* if a man believes that *p* be must believe that he believes it. From this it would follow immediately that anyone who holds any belief at all «must» hold an infinite number of «nested» or «layered» beliefs that he believes that ... (Moreover, since *CBapp* does not hold there is no obvious way of proving the usual S4 reductions, so we cannot rely, as in S4, on all but 14 of these nested forms being redundant because equivalent to one of the basic 14.)

There seem to be three possibilities. First, we can search for a modal system that will allow us to distinguish exactly as many non-equivalent modalities as are necessary for our definition. Or, second, we can assume that there is some finite upper limit on the number of layers of belief or intention that is «possible.» Or, third, we can allow the possibility of infinitely many layers, possibly non-equivalent. The first is objectionable because there seems to be no non-arbitrary way of deciding just how many layers are needed for our analysis; I have argued that we shall need at least four (*BaBbBaøb*), but how many is the most we could ever need? The third, though the most elegant, strains credulity. I compromise between elegance and applicability to reality and adopt the second. It will be assumed that there is a maximum number of layers. (But we shall not go into the question of what determines the limit, or of what happens when a sentence is uttered or thought which exceeds the limit.) For simplicity we can further assume that the maximum is the same for beliefs and intentions and any mixture. We shall say that *Bap* has two layers, *BaBap* and

([8]) HINTIKKA (1962), pp. 31-38.
([9]) ibid., pp. 109-110.

BaBbp have three layers, and so on. A sentence with the maximum number of layers will then have the form $Da_1 \ldots Da_{n-1} \emptyset a_n$, where n is the maximum for that community. If $n = 7$ an example of such a maximally nested sentence might be *BaIaBbBcIcBcøc*, which could mean roughly: *a* thinks he intends *b* to think *c* thinks he (*c*) intends *b* to think he (*c*) øs. The advantage of making this assumption, that there is, for any community, a maximum possible number of layers, is that, with it, our analysis has some chance of being roughly true of some actual community; if the number of layers could be infinite it would have no chance.

The analysis we propose is then:

(R) «The rule $\Pi x O \emptyset x$ exists in community **C**» means «$\Pi x_1 \ldots x_n K D x_1 \emptyset x_2 K D x_1 D x_2 \emptyset x_3 K \ldots D x_1 \ldots D x_{n-1} \emptyset x_n$»

As I have emphasised, this analysis is an idealization. (R) will probably be false of all or most existing communities with respect to all or most existing rules. It may be possible, though, to weaken (R) by adding to it a set of probabilistic hypotheses, so as to make it fruitful in empirical inquiries.

IV

We required that *COpPp* should hold. *Pp* may be interpreted in our notation in two ways: as $\Pi x N O N \emptyset x$, or as $N \Pi x O N \emptyset x$, i.e. $\Sigma x N O N \emptyset x$. We call the former full, the latter partial, permission. The latter is obviously a consequence of the former since communities cannot have no members; it is easy to see that full permission is a consequence of obligation on our analysis, i.e. that $\Pi x C O \emptyset x N O N \emptyset x$. For, suppose this false. Then, applying (R),

(1) $\Pi x_1 \ldots x_n \quad K D x_1 \emptyset x_2 K \ldots D x_1 \ldots D x_{n-1} \emptyset x_n$,

(2) $\Sigma x_1 \ldots x_n K D x_1 N \emptyset x_2 K \ldots D x_1 \ldots D x_{n-1} N \emptyset x_n$.

Applying EI to (2) and detaching the first conjunct, we have

(3) $Da_1N\emptyset a_2$,

while applying UI to (1) and detaching the first conjunct, we have

(4) $Da_1\emptyset a_2$.

Now, we have assumed that $\Pi xCDxpNDxNp$ holds for both *B* and *I*. But this, together with (3) and (4), yields a contradiction. The supposition must therefore be false.

EOKpqKOpOq can be proved similarly, as can *EPApqAPpPq* for full permission; but for partial permission only *CPApqAPpPq* holds.

V

It is time to consider two major objections to our analysis of rule-existence. The first is that, since no «ought» from «is» can be derived, (R) may be true of some community for some rule and yet the question may still be raised: but *ought* the members of that community to follow that rule ?

One reply might be: it is quite true that this question can still be raised, but it is irrelevant, since our analysis did not attempt to give the meaning of «ought» but only of «a rule exists in **C**.» The reply is correct as far as it goes, but it has to be admitted that the aim of our analysis, as of the legal realists, was to show how at least some questions about what ought to be can be settled solely by appeal to what is. And while the objection merely insists that the adequacy of our analysis can be questioned, it has enough plausibility to deserve a more elaborate reply.

The objector's question, to make his point, must have a sense. He must attach some sense to his «ought.» It cannot be the sense (R) would assign to it. What other sense could it have ? Von Wright suggests there are broadly three types of

«ought» (at least): the «ought» of rules and laws; the «ought» of practical necessity; and a conceptual «ought,» which I suspect can be assimilated to the first type via the notion of meaning-rules. On this view «It ought to be the case that *p*» must mean either «There is a rule that requires that it be the case that *p*» or else «It is in someone's interest to make it the case that *p*.» (Both these analyses are of course only approximations.) I concur with Von Wright's conclusion: «I doubt whether it [the search for a supposedly sui generis moral «ought»] will give us an «ought» of a radically different conceptual nature from the three cases which I have here distinguished. But I do not wish to deny that there may exist important uses of the word 'ought' which may obey quite different conceptual patterns.» (102) It seems that our objector's «ought» will have to have a sense quite different from any of Von Wright's three. Perhaps there is such a sense, but it is now for him to say what it is.

One version of the objection, associated with the writings of Hare ([10]), can be formulated in this way: the question «Ought they to believe and intend as they do ?» remains open, and its *sense* (in the narrow, Fregean meaning of that word) need not be different from those you have mentioned, provided it be admitted that its *force* (in the Austinian meaning of that word) is. And its force is indeed very different, for while (R) is an elaborate report upon the thoughts of persons in **C**, my question asks not for a report at all but for an utterance with quite different, speaker-committing or action-guiding, force. You have a dilemma: if your «ought» is to be genuinely evaluative, not just a report of the evaluations of others, it must possess this dynamic, speaker-committing force. But no such judgement can be deduced from any set of judgements which do not, some of them, already possess it. Your analysandum, if it is genuinely evaluative, cannot therefore be deduced from your analysans, unless it has as part of its meaning a dynamic force which will destroy the claim of your analysis to be naturalistic.

([10]) HARE (1952).

One merit of our analysis is that it offers an explanation of the facts adduced by this objection. Consider a community **C** and a person *a*. Let us ask two questions. First, what determines whether *a* is a member of **C**? In particular, with respect to some rule existing in **C** that $\Pi x O \emptyset x$, second, from the fact that *a* believes there is this rule, does it follow that he «must», insofar as he is rational, be one of those persons to whom what is said of x, in (R) truly applies? That is, must he then (be prepared to) believe and intend all the things (R) says members of **C** believe and intend when the rule exists in **C**? I suggest that these two questions are connected at least in the following way. It is necessary, if *a* is to be a member of **C**, that he be prepared to infer from the existence of the rule in **C** to his own involvement in the beliefs and intentions that constitute the rule. This will not be sufficient, unless **C** is a peculiar community bound together by only one rule, or, perhaps, the rule is in some way solely constitutive of the community. Also, there can be degrees of acceptance — just as a rule may have a shadowy existence (it may be withering away), so a person may have only a shadowy membership in a community. Our objector is drawing attention to the fact that a person *a* who asserts that there is a rule $\Pi x O \emptyset x$ in **C**, but rejects such statements as $Ia\emptyset a$, $Ia\emptyset b$, and $\Pi x Ia\emptyset x$ is in a strong sense not a member, because he refuses to be a member, of **C**. The objector's point is that from the statement that there is the rule one cannot deduce anything about whether the person who made the statement is or isn't a member of **C**. This contention is quite true but it is quite different from, and does not support, the thesis that a moral obligation to follow some rule can never be deduced from any set of facts, even the fact that the rule exists in some community.

But with all this said, our answer to the objection remains rather inconclusive. For it is merely a challenge to the objector to provide us with an account of his «ought» which will be non-factual in the way he requires, i.e. distinct from our two «ought»s. We may allow, in a more constructive spirit, that the objection raises the following problem: if it is true that a community has the rule $\Pi x O \emptyset x$, what connexion must exist

between øing and the wants and aims of persons in that community ? Can it really be true, as our analysis implies, that the rule exists if and only if (R) is true, regardless of what sort of community it is and what kinds of lives the people in it lead, and in particular, regardless of what their aims and wants are, of what they *value* ?

I think it *could* be true: there could be such a community, with such rules. But perhaps it would be very difficult for human beings to understand very much of what happened in it. It seems very unlikely that any human community ever is of this kind. So (R) is, I am suggesting, an analysis of rule-existence in general, but very likely the richer concept of rule-existence in a possible human community, if its analysis is to take into account such things as earthly scarcities, human biology, and so forth, will involve a good deal more.

A second objection is this. ([11]) (R) fails to meet a basic requirement for any adequate analysis of an empirical concept, namely that it be possible for a user of the concept to be mistaken in thinking that the concept does not apply to some possible instance: here, it must be possible for a user of the concept of rule-existence to be mistaken in thinking that there does not exist in **C** the rule ΠxOøx. (R) does not meet this requirement, for if we suppose that it is true, and that *a* is in **C**, and that he mistakenly thinks the rule ΠxOøx does not exist in **C**, we can deduce a contradiction. If *a* thinks that the rule does not exist in **C**, then he thinks that in **C** *N*ΠxOøx. So, according to (R), he mistakenly supposes that some fact or other about the beliefs or intentions of members of **C** is not a fact. Let us imagine, for example, that he mistakenly supposes that *NIbøz*. Now according to (R) *a*, being in **C**, thinks that *Ibøc*, amongst other things. But according to Hintikka's principle of rational belief, adopted in this paper, if *BaIbøc* then *NBaNIbøc*. This contradicts our supposition that *a* does mistakenly think that *NIbøc*. Hence, either (R) is false, or it is logically impossible to be mistaken in believing that a rule

([11]) This objection was pressed by Leo Simons in discussion.

does not exist in one's community, and this latter alternative is absurd.

Two complementary things can be argued in reply. First, perhaps the right conclusion of the argument is the proposition just dismissed as absurd, namely in the supposed circumstances *a* cannot after all be a member of **C**. To take this line is to assert that **C** is a community in which all members are perfectly informed about all members' beliefs and intentions with regard to all members' actions. This reply is, I suggest, in the same spirit as Hintikka's reply to the already discussed objection to his principle of rational belief (on which the objection depends), namely, that his principle is intended to apply only to «perfectly rational» belief — indeed, is a partial account of what perfectly rational belief is. For it ought to be possible to describe the form that perfectly rational belief would have, in spite of the undoubted fact that no-one ever is a perfectly rational person; and similarly, the reply runs, it ought to be possible, as part of an account of the nature of a rational community, to describe the form that the beliefs and intentions of such a community's members would have.

A second, more positive, line of reply is this. The objection proves too much, for it seems, on the face of it, to show that norm-existence must be entirely independent of any belief of any member of **C** about members' beliefs. For the argument which deduced a contradiction from (R) and the assumption that *a* was mistaken in thinking that $N\Pi xO\o x$ could be simply adapted to deduce a contradiction from the weaker assumption that there is some proposition, belief in which by all members of **C** is a necessary condition of the existence of some norm $\Pi xO\o x$. Call it proposition *p*. We then get a contradiction from the assumptions that *a* is a member of **C** and mistakenly thinks that *Np*. It ought then to follow, in the spirit of the objection, that a norm could exist in **C** even if no proposition of the form $\Pi xBxp$ were true of the members of **C**. But surely there could not be such a thing as a community with rules existing in it yet no shared beliefs at all. Consideration of beliefs about the meanings of words is enough to show this.

I conclude that the objection shows, not that (R) is a defective analysis of the ideal case, but only that the ideal case cannot in practice be realised. However, it must be admitted that our idealised model is ideal in several different ways, and work remains to be done on the separation and clarification of these different sorts of ideal. ([12])

Middlesex Polytechnic, London Mark Fisher.

BIBLIOGRAPHY

ANDERSON (1958) A. R. Anderson, «A reduction of deontic logic to alethic modal logic,» *Mind* vol. 67 (1958), 100-103.

GOFFMAN (1967) E. Goffman, *Interaction ritual.* New York: Doubleday, 1967.

GRICE (1957) H. P. Grice, «Meaning,» *Philosophical review* vol. 66 (1957), 377-388.

HARE (1952) R. M. Hare, *The language of morals.* Oxford: Oxford University Press, 1952.

HINTIKKA (1962) J. Hintikka, *Knowledge and belief.* Ithaca, New York: Cornell University Press, 1962.

MAYO (1968) B. Mayo, «The moral agent,» in *The human agent,* Royal Institute of Philosophy Lectures, vol. 1, 1966-67. London: Macmillan, 1968. 47-63.

NAGEL (1969) T. Nagel, «Sexual perversion,» *Journal of philosophy* vol. 66 (1969), 5-17.

VON WRIGHT (1969) G. H. Von Wright, «On the logic and ontology of norms,» in J. W. Davis, D. J. Hockney and W. K. Wilson ed., *Philosophical logic.* Dordrecht: D. Reidel, 1969. Pp. 89-107. All page references are to this chapter.

([12]) The idea that nested intentional concepts may be involved in the way I have sketched in a number of important human relationships has been put forward in Mayo (1968), in Grice (1957), and in Nagel (1969). Nagel acknowledges his debt to Grice and to J.-P. Sartre. The general idea is perhaps rather a commonplace in sociological theory; I first came across it in Goffman (1967). I owe to correspondence with Von Wright the stimulus that led to my formulation of (R), and in its working out I was helped by discussion with colleagues and students at Cincinnati, especially Bill Todd, Bob Faaborg and Jenefer Robinson.