

MODAL LOGIC AND AGENCY

Douglas WALTON

The University of Winnipeg
515 Portage Avenue
Winnipeg, Manitoba, Canada.

Despite the recent flourishing of modal logic, and the obviousness of the need in many areas to understand the elementary modal properties of agency, not much work has been done on the study of agency as a sentence operator. A recent welcome exception is the work of Ingmar Pörn [6], who bases a rigorous study of the logical structure of influence relations, normative relations, rights, norms, and power, on a modal language of agency. This paper is a critical discussion mainly of one key assumption in Pörn's base-language.

1. *The Language L_1*

L_1 is a two-sorted language based on the language of first-order predicate logic with identity. Notably, L_1 contains praxiological sentence-operators relativized to individual symbols interpreted as denoting agents. Thus $\lceil D_i p \rceil$ is read as '*i* brings it about that *p*'. The semantics of L_1 , based on Hintikka-style model sets and model systems (see Hintikka [4]), yields a system in which the D-operator is isomorphic with the L-operator of the Feys-Gödel System T of standard alethic modal logic ($\lceil \cdot \rceil$). We will be primarily concerned here with some crucial syntactical properties of Pörn's D-operator. Hereafter, to conform with the notation of Walton [8], we write $\lceil \delta_a p \rceil$ where a is an agent symbol and δ stands for the praxiological operator D in Pörn's notation. The proposal of Pörn is equivalent to the suggestion that the following two axioms and rule obtain for δ , where \emptyset is a theorem.

- (A δ 1) $\delta_a p \supset p$
 (A δ 2) $\delta_a (p \supset q) \supset (\delta_a p \supset \delta_a q)$
 (R δ 1) \emptyset , to infer $\delta \emptyset$

As (R δ 1) makes explicit, it is a feature of L_1 that all theorems of L_1 are brought about by any agent. Thus $\delta_a (p \vee \sim p)$, $\delta_a \sim (p \& \sim p)$, and the like, will be theorems. According to Pörn, «... the appearance of oddity is counteracted as soon as one sees that what is asserted here concerns, not an action of an agent in regard to a state of affairs p or its contradictory opposite, but a feature inherent in 'bringing it about that'». (p. 7). We might add here that the reasonableness of this suggestion could be reinforced by regarding the δ -operator as vacuously applicable in such instances, (2) much as a quantifier is vacuous over a schema containing no free variables matching the variable of the quantifier.

The language based on (δ A 1), (δ A 2), and (δ R 1) is extremely interesting and potentially very fruitful, not only because of its applications but because isomorphism with T yields many theorems immediately, including significantly the following (see Pörn, p. 14).

- (T 1) $\delta_a p \supset \sim \delta_a \sim p$
 (T 2) $\delta_a (p \supset q) \supset (\sim \delta_a \sim p \supset \sim \delta_a \sim q)$
 (T 3) $\delta_a p \supset \delta_a (q \supset p)$
 (T 4) $\delta_a \sim p \supset \delta_a (p \supset q)$
 (T 5) $\delta_a p \supset \delta_a (p \vee q)$
 (T 6) $\delta_a q \supset \delta_a (p \vee q)$
 (T 7) $(\delta_a p \vee \delta_a q) \supset \delta_a (p \vee q)$
 (T 8) $(\delta_a p \& \delta_a q) \equiv \delta_a (p \& q)$
 (T 9) $\delta_a (p \equiv q) \equiv (\delta_a p \equiv \delta_a q)$

To confine the discussion to reasonable proportions, we here exclude consideration of other theorems developed by Pörn including theorems containing iterated δ -operators, and other praxiological operators. We will primarily be concerned to discuss some aspects related to (T 3) and (T 4), the praxiolo-

gical analogues to the paradoxes of strict implication. ⁽³⁾

(T 3) has as an instance: if Socrates drops the cup then Socrates brings it about that if the ass is running then the cup falls. (T 4) has as an instance: if Socrates brings it about that the cup does not drop then Socrates brings it about that if the cup drops the earth collides with the sun. Reflection on these puzzling theorems seemed to me a good reason to take a much harder look at (δ A 2) to see if his axiom might be too strong to yield a calculus of agency that would conform to the basic conception of agency that Pörn wants to utilize as a foundation for a logic of power. In the sequel, we see that these hesitations are justified.

To facilitate discussion, we consider a truth-functional equivalent of (δ A 2), namely

N 1.1.

$$(\delta A 2') \quad (\delta_a p \ \& \ \delta_a (p \supset q)) \supset \delta_a q$$

We now examine two clear cases of actions wherein, in both instances, the antecedent of (δ A 2') strongly appears to be true and the consequent false.

2. Two Counter-Examples

First Counter-Example: Consider the scenario wherein Smith brings it about that if Jones falls from the roof of a thirty-storey building he will die, i.e., Smith removes an intervening flagpole and so on. At the appropriate opportunity, Smith pushes Jones over the edge. As Jones plunges past the twentieth floor, Schmidt (a distant relative of Smith) fires from an open window, instantly killing Jones. This action-scenario satisfies the antecedent because (1) Smith brought it about that if Jones fell, he would die, and (2) Smith brought it about that Jones fell. The consequent, however, is not satisfied, for it is not clear that Smith brought it about that Jones died. It is true that Smith ensured Jones' imminent demise, but not clearly true that Smith was the person who *actually* brought about the death of Jones.

Discussion: The introduction of temporal indices, it can be countered, defeats the computer-example, for Smith brought it about that if Jones fell at t , he would die at t' , whereas Jones actually died at some time previous to t' . In other words, the proper statement of the principle in question should read:

$$(\delta A 2^*) \quad (\delta_a p_t \ \& \ \delta_a (p_t \supset q_{t'})) \supset \delta_a q_{t'}$$

Now the scenario above fails to satisfy the right conjunct of the antecedent of this schema because the fact that Jones was already dead at t' rules out the contention that Smith brought it about that if Jones fell at t , he would die at precisely t' .

The above rebuttal raises the question of whether the consequent of the shorter conditional needs to be satisfied in order for the entire antecedent of the schema to be satisfiable. The defender of $(\delta A 2^*)$ appears to argue as follows.

In the antecedent, we have it that $\delta_a p_t$ obtains, and hence by the axiom $\delta_a p \supset p$ that p_t obtains. By similar reasoning, we have it that $p_t \supset q_{t'}$ obtains. Now by *modus ponens* it follows that $q_{t'}$ obtains. Otherwise the entire antecedent

$$\delta_a p_t \ \& \ \delta_a (p_t \supset q_{t'})$$

cannot be satisfied by any scenario that is first-order consistent. But the satisfaction of the consequent itself is not really the crux of the rebuttal, as can be seen by observing that in the scenario, $q_{t'}$ really does obtain. That is, according to the example, Jones *is* dead at t' . The real issue is whether the entire schema $\delta_a (p_t \supset q_{t'})$ is realized.

And it seems that so far we have been offered no conclusive reason for believing that $\delta_a (p_t \supset q_{t'})$ is not realized in the scenario. True, Smith did not die at precisely t' , but he was dead at t' . Thus it becomes apparent that we need to distinguish

clearly between two interpretations of $\lceil \delta_a q_{t'} \rceil$:

- (1) Smith brought it about that Jones died at t' .
- (2) Smith brought it about that Jones was dead at t' .

Hoping that it does not beg too many questions, we will hereafter proceed on the assumption that (2) is the preferred interpretation and that (1) is, for our purposes, an elliptical rendering of (2). Expression (1) is, however, ambiguous, in that it appears to attribute a spurious agency to Jones. That is, it appears to assert 'Smith brought it about that Jones brought it about that Jones was dead at t' .' which would be translated

as $\lceil \delta_a \delta_b q_{t'} \rceil$. In any case, whether Jones was alive or dead before t' is a matter of indifference to the conception of agency we are here concerned with.⁽⁴⁾ Thus hereafter we will treat (1) for our purposes as generally translatable as equivalent to the more felicitous (2).⁽⁵⁾ This sharpens the

issue considerably. Now, even if we read $\lceil \delta_a (p_t \supset q_{t'}) \rceil$ very strongly as 'Smith ensured that Jones would die if he fell', the schema appears satisfied by the scenario since $\lceil \lceil \rceil$ actually obtains. The problem is that $\lceil \delta_a q_{t'} \rceil$ does not obtain.

It may still seem to remain an open question whether $\lceil \delta_a (p_t \supset q_{t'}) \rceil$ obtains, however, because Jones being dead at t' was not a consequence of his falling off the building-Jones was more significantly, a shooting victim. Accordingly, this new form of rebuttal continues, it is not true that Smith brought it about that if Jones fell he died. This rebuttal can be dismissed more summarily than the previous one. All that is required is that Jones' falling be sufficient for his death. No additional requirement of agent-causality can be imposed on expressions $\lceil \delta_a (p \supset q) \rceil$ without radically changing the entire intention and interpretation of the calculus. Clearly the material conditional is in no way adequate to the kind of agent-causality presup-

posed by this version of the rebuttal. (6)

To sum up, the counter-example emerges as quite strong, but since there may still be further arguments found to bear on the doubtful aspects of the contention that $\lceil q_t \rceil$ obtains in the scenario according to the best interpretation of the calculus, we consider a second putative counter-example where $\lceil q_t \rceil$ decisively obtains.

Second Counter-Example: Consider a scenario where Smith places an explosive charge in a mine shaft and ignites the fuse. Jones then extinguishes the fuse and then sets off the charge electrically at exactly the time it would have exploded if the fuse had been allowed to burn down normally. Here $\lceil q_t \rceil$ clearly obtains, i.e., the mine shaft collapsed just at t' where Smith brought it about that if the fuse was ignited at t , mine would collapse at t' . Since, in addition, Smith ignited the fuse at t , the antecedent appears to be satisfied whereas clearly the consequent is not, since Smith did not bring it about that the mine collapsed. Jones did.

Discussion: As before, the most questionable aspect of the counterexample is whether $\lceil \delta_a p_t \supset q_t \rceil$ obtained. We are inclined to respond that Jones' intervention overturns that allegation. Smith thought that he had brought it about that if the fuse was lit the mine would collapse. Jones proved it false. Yet our inclination to equivocate here draws out a need for additional temporal indices. Certainly at $t + \Delta$, it became false that $\lceil \delta_a p_t \supset q_t \rceil$ obtained, but previously no falsehood was evident, i.e., $\lceil \delta_a^t (p_t \supset q_t) \rceil$ is true. What is required here is that we specify the time of the bringing-about as well as the time of what was brought about. What still seems to have a true antecedent is the more specific principle,

$$(\delta A 2^o) \quad (\delta_a^t p_t \ \& \ \delta_a^t (p_t \supset q_t)) \supset \delta_a^{t'} q_t$$

Insofar as it was possible at t at all for Smith to bring it about

that $\overset{[}{p_t \supset q_t}$] (since the future is always uncertain) it is true that Smith created such a state of affairs at t . Of course the scenario does not constitute a counter-example to the principle,

$$(\delta_a^t p_t \ \& \ \delta_a^{t'} p_t \supset q_t)) \supset \delta_a^{t'} q_t,$$

The time t' is later than Jones' extinguishing the fuse, and thus at t' it is false that Smith brought it about that if the fuse was lit the mine would collapse. Nor is the counter-example affective if we construe the principle as

$$(\delta_a^t p_t \ \& \ \delta_a^{t'} p_t \supset q_t)) \supset \delta_a^{t'} q_t,$$

At t we have no strong reason to assert the falsity of the consequent $\overset{[}{\delta_a^{t'} q_t}$]. At this time it still might be true that Smith is the author of q_t . We might be somewhat puzzled here though at the notion of someone's having brought about $p_{t+\Delta}$, for some fairly lengthy Δ , at t .

This raises a general problem about the times of actions. Suppose in 1350 B.C. an Egyptian puts a poisonous device with a spring in a box in a tomb, and an archaeologist opens the box and consequently dies in 1974. When did the killing take place? Uncontestably the archaeologist became dead in 1974, but when, if at all, was this brought about by the Egyptian? The most likely account of the matter seems to be as follows. The Egyptian brought about a certain deadly conditional state of affairs in 1350 B.C. (If anyone opened the box, he would die.) Yet insofar as the Egyptian can be said to have categorically brought about the death of the archaeologist, he must have done so in 1974. He certainly couldn't have brought it about in 1350, because the archaeologist did not even exist nor was dead at that time. Of course, if this account is correct, the Egyptian did not exist at the time his criminal act was consummated, but I think that anyone who proposes a language allowing for future conditional actions

will have to learn to live with this kind of consequence.

To sum up, the first of this tensed triad, ($\delta A 2^\circ$) seems likely to be the most congenial to proposers of principles of this type, yet the first formulation is the most susceptible to the counter-example. It would be over-sanguine to expect the discussion to be conclusive, given the *dubia* we have found in the attribution of times to actions. Perhaps the best we can say is that we have found no very strong reason to believe that a useful tensed formulation of the principle can be achieved that is not open to a counter-example of this type.

3. Concluding Remarks

These two counter-examples raise serious questions about the interpretation of ($\delta A 2$). At present, however, we can hardly regard them as constituting a conclusive refutation of ($\delta A 2$) nor of L_1 . Yet plainly they do demand further refinements in our understanding of some special features of the notion of agency captured by L_1 . One alternative approach I shall only briefly present here may provide a helpful framework for a constructive redefinition of the problem. We could establish a minimal notion of direct agency using the weaker set of axioms,

$$(A \delta 1) \quad \delta_a p \supset p$$

$$(A \delta 2) \quad \delta_a (p \ \& \ q) \supset (\delta_a p \ \& \ \delta_a q)$$

plus the rule ($\delta R 1$). Fitch [3] has suggested what amounts to this minimal account of agency. Then, proceeding from this base, we can approach the problem of the consequences of actions through the following defined concepts.

$$(D 1) \quad a \text{ indirectly brings it about that } p =_{at} \delta_a q \ \& \ q \text{ causes } p$$

$$(D 2) \quad a \text{ interdirectly brings it about that } p =_{at} \delta_a q \ \& \ (q \rightarrow p)$$

$$(D 3) \quad a \text{ alterdirectly brings it about that } p =_{at}$$

$$\delta_a q \ \& \ \delta_a (q \supset p)$$

Then the various modal properties of the resultant varieties of agency can be studied and compared. This is by means the only way to proceed, but if subsequent discussion vindicates the hypothesis that ($\delta A 2$) is untenable, at least one alternate way of proceeding is available and could be pursued. For Pörn's work shows the immense promise of modal languages for the study of agency in the foundations of jurisprudence, the social sciences, (7) and any area where agency-locations and interpersonal expressions are important. System L_1 is an auspicious beginning, despite its difficulties of interpretation.

¹⁾ See Hughes and Cresswell [5], ch. 2.

⁽²⁾ This would be similar to a convention of Fitch's System DM for deontic modal logic in Fitch [2].

⁽³⁾ See Hughes and Cresswell [5], p. 39f.

⁽⁴⁾ See Pörn [6], p. 4f.

⁽⁵⁾ Alternatively, the second counter-example, below, becomes relevant.

⁽⁶⁾ For related discussion see Davidson [1].

⁽⁷⁾ See Walton [7] and Walton [8].

REFERENCES

- [1] Donald DAVIDSON, 'The Logical Form of Action-Sentences,' *The Logic of Decision and Action*, Pittsburgh, University of Pittsburgh Press, 1966.
- [2] Frederic B. FITCH, 'Natural Deduction Rules for Obligation,' *American Philosophical Quarterly*, Vol. 3, No. 1, January 1966, 27-38.
- [3] Frederic B. FITCH, 'A Logical Analysis of Some Value Concepts,' *Journal of Symbolic Logic*, Vol. 28, No. 2, June 1963, 135-142.
- [4] Jaakko HINTIKKA, *Models for Modalities*, Dordrecht, Reidel, 1969.
- [5] G.E. HUGHES and M.J. CRESSWELL, *An Introduction to Modal Logic*, London, Methuen, 1968.
- [6] Ingmar PÖRN, *The Logic of Power*, Oxford, Blackwell, 1970.
- [7] Douglas WALTON, 'Logical form and agency,' *Philosophical Studies*, forthcoming.
- [8] Douglas WALTON, 'Control,' *Behaviorism*, vol. 2, No. 2, 1974. 162-171.