ALLEN HAZEN

NATURAL DEDUCTION AND HILBERT'S ε-OPERATOR

In his recent [4], Kit Fine presents a model-theoretic account of "arbitrary" (or "variable") objects and uses it to analyze the rules of Existential Instantiation (EI), etc., found in many textbook systems of logic. Along the way he argues for the pedagogical utility of such rules and for a (not very clearly formulated) theoretical claim on their behalf: that they mirror the patterns of informal reasoning more faithfully than do the quantifier introduction and elimination rules of Gentzen-style natural deduction. In this note I wish to point out (what must have been a familiar point to many logicians) that there is an older logical tool, Hilbert's ε-calculus, providing a simple and intuitive analysis of these rules, and to express a few doubts about Fine's pedagogical and "ordinary language analytic" claims. Since, however, this will concern only the analysis of systems of first-order logic, it does not constitute a general critique of Fine's theory of arbitrary objects: he himself ([4], p. 106; all page references to Fine are to this article) feels that the advantages of his scheme are only fully apparent when we go beyond first-order reasoning. ([5], however, though defining semantic notions in great generality, does not pursue applications much beyond [4].)

The first sort of natural deduction system Fine discusses (pp. 76–78) is that typified by Gentzen's [7] NK, whose analysis carries over virtually without change to the Fitch-style systems of such texts as [6], [12] and [19]; I shall assume Fitch-style NK as a basis of comparison in what follows. As Fine remarks, his theory of arbitrary objects offers no special advantages over standard model theory in the analysis of the use of free "eigenvariables" in the Universal Quantifier Introduction (U.Q.Int) and Existential Quantifier Elimination (E.Q.Elim) rules of such systems. Fine's machinery, then, gets its first real workout in the analysis of systems with stronger quantifier rules, such as EI. The details of the analyses vary with the system, but in each

Journal of Philosophical Logic **16** (1987) 411–421.

case it is proposed to interpret the instantial letters used in connection with the rules as denoting certain arbitrary or variable objects. The soundness of the systems is proven with respect to this interpretation; since classical models are a special case of Fine's models (arbitrary-object models with a null set of arbitrary objects), this yields the soundness of the system (for theses or arguments not containing any of the special terms) with respect to the ordinary semantics of first-order logic.

The analysis I wish to describe treats the instantial letters of rules like EI as abbreviations of certain ε-terms and explains the restrictions placed on them by noting that, were the ε-terms written out in full, simple syntactic factors (notably, the asymmetry of the "contains as a proper subterm" relation among ε-terms) would lead automatically to their fulfillment. Soundness follows from (1) the fact that the restrictions on instantial letters are strong enough to allow all such letters occurring in a correct derivation to be "disabbreviated" into full ε-terms, and (ii) the soundness of the ε-calculus.

Hilbert's ε-operator is a variable-binding term operator (v.b.t.o.) in the sense of [8]. Where $A(a)$ is a formula, $(\varepsilon x)A(x)$ is thought of as denoting an arbitrarily chosen object in the domain satisfying $A(a)$ if there are any, and an arbitrarily chosen object in the domain if there are none. The metatheory of predicate calculi with ε-operators gets its classical treatment in [9]; the Germanless reader is referred to [13]. Standard inference rules for the ε-operator are (ε1) from $\exists x A(x)$ to infer $A((\varepsilon x)A(x))$, and (ε2) from $A((\varepsilon x)\neg A(x))$ to infer $\forall x A(x)$. In the present context of pure first-order logic, we may take Hilbert's second ε-theorem to be the proposition that the system formed by adding these rules to a formulation of first-order logic is conservative over first-order logic. The original proof of this is one of the triumphs of Hilbert-style finitistic metamathematics, and forbiddingly complex. There are, however, two distinct reasons for thinking that this complexity need not prevent appeal to the ε-theorem even in pedagogical contexts. One is that, for the applications described below, only ε-terms containing no variables bound to operators outside the term (*proper* ε-terms) are needed, and the finitistic proof of the ε-theorem for a system so restricted is far simpler than that for the general case. The other is an instance of the general phenomenon that the use of abstract (e.g. set-theoretic) techniques can make it possible to give proofs of elementary results that are far more readily comprehensible (and perhaps in some sense more explanatory) than the elementary proofs: we may appeal to model theory. The second ε-theorem is an immediate consequence of the soundness, with respect to the interpretation sketched above, of the rules (ε1) and (ε2), and the completeness of the predicate calculus. Another result, related to the first ε-theorem, is that when the rules of E.Q.Elim and U.Q.Int are replaced, in a complete system of first-order logic, with (ε1) and (ε2), the system remains complete: given the restrictions on premisses containing eigenvariables, we may transform a proof in the original system into one in the modified system by replacing each occurrence, in a sub-proof for one of the replaced quantifier rules, of the eigenvariable of the sub-proof with an occurrence of the appropriate (proper) ε-term (and, in a Fitch-style proof, erasing the vertical line of the sub-proof).

Suppose, now, that having replaced E.Q.Elim and U.Q.Int with the two ε rules, we abbreviate the ε-terms to single letters. The derivations in the abbreviated notation will look like those of Quine's system in [16] (discussed by Fine on pp. 78ff.); the restrictions Quine places on his rules of Universal Generalization (UG; = ε2) and EI (= ε1; his Universal Instantiation and Existential Generalization are simply U.Q.Elimination and E.Q.Introduction) can be seen as designed to ensure that all the instantial letters of a derivation can be disabbreviated to ε-terms. What Fine (p. 79) calls the *Local Restriction* amounts simply to the requirement that the bound variable of the quantified formula in the inference (the premiss in EI and the conclusion in UG) replace all occurrences of the instantial letter in the other formula. This would be implied in a careful statement of (ε1) and (ε2), and, as Fine remarks, corresponds to one of the conditions on E.Q.Elim and U.Q.Int in a Gentzen- or Fitch-style system. The remaining restrictions, *Flagging* and *Ordering*, are more interesting. Flagging requires that the same instantial letter not be used in two inferences by the new rules. One thing this rules out is, for example, inferring $\forall x A(x)$ from $A(a)$ and then, several lines later in the derivation, inferring the same conclusion from the same premiss a second time. One may occasionally want to do this if, say, the first inference is made under a hypothesis that has been discharged by the time one gets to the second, but one never *has* to use UG twice on the same premiss in this way: if nothing else, one can start one's derivation with a quick conditionalization to establish (free of any hypothesis) $(A(a) \supset \forall x A(x))$, and then use *modus ponens* rather than UG on later occasions. The important thing outlawed by the Flagging restriction is the use of the same instantial letter in two *different* inferences by UG or EI. This, though, in terms of our interpretation of instantial letters as abbreviated ε-terms, comes down to the simple injunction not to use the same letter to abbreviate different ε-terms: avoid the fallacy of equivocation.

Finally, the Ordering restriction requires that it be possible to define a linear ordering of the instantial letters occurring in a given derivation in such a way as to make every instantial letter occurring in the quantified premiss of an EI or conclusion of a UG later than the instantial term of the inference. This restriction can seem quite arcane, especially since the ordering may be quite different from the order in which the inferences occur in the derivation. When the instantial letters are seen as abbreviated ε-terms, however, all becomes clear. The ε-terms occurring in the quantified formula of an inference by (ε1) or (ε2) in which a given ε-term is the instantial term also occur as subterms within the given ε-term itself. The ordering, then, is a simple consequence of the asymmetry of the "proper subterm of" relation among ε-terms. (To be sure, such syntactic part-whole relations are in general only partial and not linear orders: an ε-term can contain two subterms which are not subterms of each other. A little reflection serves to show, however, that when the subterm relation on the ε-terms in a proof using (ε1) and (ε2) is a partial order, any of the linear orders extending it will meet the criteria of the Ordering restriction when the terms are abbreviated as Quinean instantial letters.)

In short, any derivation in Quine's system can be disabbreviated to a derivation in the ε-calculus. This, together with the model theoretic proof of the soundness of the ε-rules, gives a proof of the soundness of Quine's system which seems to be more perspicuous and perhaps in some sense more explanatory than Quine's elementary but complicated argument. The completeness of Quine's system follows from the completeness result of the system in which (ε1) and (ε2), with proper ε-terms, replace E.Q.Elim and U.Q.Int, together with the fact that any proof in that system can (after at most minimal reformulation) be abbreviated to one in Quine's system.

In a concluding discussion (p. 97) Fine argues that "it is hard to regard Quine's system either as a very satisfactory system in its own right or as a faithful representation of ordinary reasoning". The objection centers on the treatment of the instantial letters for UG. "It is a natural requirement on a derivation containing *A*-names, or any other names, that we should know what those names denote as soon as they are introduced; their interpretation should not depend upon what subsequently happens in the derivations", but Quine's system, interpreted in terms of his arbitrary object semantics, goes against "this ban on retrospective interpretation". The complaint, at least, is easily formulated in terms of our interpretation. In reading a derivation in Quine's system from the top down, one will not be able to tell, on first encountering a letter that will ultimately serve as the instantial term of a UG, what ε-term it is meant to abbreviate; for this, one will have to search down until one finds the inference where it is used.

What is less clear is the pedagogical and analytic importance of the complaint. From a pedagogical standpoint, what is perhaps most important is that students be able to construct (and verify the correctness of) derivations in the system easily. Whether the retrospective interpretability of some of the instantial letters is a hindrance in this regard I do not know: the many logic teachers who have taught from Quine's textbook over the last thirty years might have ideas on the subject. My own guess is that it wouldn't be. Students are, after all, advised to "start from both ends" in attempting to find derivations, to work up from the conclusion as well as down from the premisses. Someone might, I suppose, argue that in some more abstract sense a system of formal derivations can only represent intuitively convincing proofs if it heeds the ban on retrospective interpretation. Here, I think, Quine's system could be defended by taking seriously the idea that its instantial constants are abbreviations of ε-terms. For the convenience of students, the defense would run, we allow conventions of abbreviation which temporarily obscure the interpretation of some terms, but if the ε-terms were written out in full, each term's interpretation would be clear as soon as it was introduced.

After treating Quine's system, Fine turns to a consideration (pp. 92–98) of the system of Copi [3] and Kalish [11]. Here the individual inferences take the same form as in Quine's system, but the restrictions are very different. Fine notes that the instantial letters for EI and UG are treated quite differently, and (usefully) suggest that, in view of their different functions, separate alphabets be used for the two kinds of terms. In fact, the instantial letters for UG amount to the familiar free variables from U.Q.Int, in the sense that the restrictions limit application of the UG rule to cases where the Fitch- or Gentzen-style rule could be applied, and I will refer to it as U.Q.Int. EI, on the other hand, is like Quine's rule, with instantial letters that can be thought of as abbreviated ε-terms, and once again the restrictions (pp. 92–93) suffice to guarantee their disabbreviability in any derivation. The only subtlety is in the restriction of *Independence*. When we think of the instantial letters for EI as abbreviated ε-terms, the relation of dependence takes on a simple syntactic character: one term depends on another just in case the first (or rather, the ε-term of which it is the abbreviation) contains the second (or . . .) as a subterm. In requiring that no term dependent on the eigenvariable occur in the conclusion of a U.Q.Int, Independence keeps us from turning the free variables "invisibly" occurring in our abbreviated ε-terms into bound variables of quantification. Not that there is anything intrinsically wrong with improper ε-terms, but since the abbreviated notation does not distinguish between a "free" ε-term and one containing a variable bound by a quantifier external to it, they would tempt us to commit fallacies later. In forbidding occurrences of instantial letters dependent on the eigenvariable in the suppositions of the inference, on the other hand, Independence extends the coverage of the usual proviso on U.Q.Int – that the conclusion of the inference not depend on any undischarged hypotheses containing the eigenvariable – to cover "invisible" occurrences of the eigenvariable. (Pedagogical aside: this proviso has to be stated as a restriction in a Gentzen-style system or, as we have just seen, in a Copi-style one, but it is enforced by the restrictions on reiteration into general sub-proofs in Fitch-style ones. A hybrid system, combining a rule of EI with a Fitch-style rule of U.Q.Int, would relocate part of the Independence restriction as a tightening of the restriction on reiteration: neither formulas containing the eigenvariable nor formulas containing terms dependent on it may be reiterated into a general sub-proof.)

There is one further kind of system, closely related to systems of natural deduction, where something like an EI rule has been proposed: Beth's semantic tableaux, which, in Smullyan's vastly improved notation, have been presented in a treatise [17] and in such texts as [10]. The connection with natural deduction is perhaps most readily seen if we make a few changes in the familiar systems of natural deduction. First, incorporate into the system of classical logic the negative rules presented in [6] for the non-classical system of that book, so that negated conjunctions (disjunctions/universal quantifications/existential quantifications) become immediately equivalent to disjunctions (conjunctions/existential quantifications/universal quantifications) of negations. Second, treat the conditional as defined in terms of disjunction and negation. Third, add a propositional *Falsum* constant with the rule that *Falsum* is a direct consequence of the two formulas A and $\neg A$. Then a closed semantic tableau for a formula (argument) can be thought of as a slightly abbreviated, tree-form, presentation of a natural deduction derivation of *Falsum* from the negation of the formula (premisses and negation of conclusion of argument), in which only elimination rules are used.

The usual rules for extending tableaux with quantified formulas on them (any instance of a universal quantification may be added, but for an existential quantification a constant new to the branch must be used) correspond to U.Q.Elim and E.Q.Elim. Smullyan, however, on pp. 54–56 and 78–79 of [17], discusses a "liberalized" version of the existential quantifier rule that can shorten some tableaux in ways analogous to those in which a rule of EI can shorten natural deduction derivations. Under the liberalized rule, the constant substituted for the bound variable of the existential quantifier must either be new to the branch or fulfill the three requirements of (i) not having previously been introduced by the existential quantifier rule, (ii) not occurring in the existentially quantified formula the rule is applied to, and (iii) there not being any constant in the existentially quantified formula which was previously introduced by the existential quantifier rule.

It is fairly easy to show that dropping any of the three requirements allows the construction of closed trees validating invalid inferences, but it would be nice to have a more intuitive explanation for them. One is provided if we look on the liberalized rule as analogous to (ε1), with the constant abbreviating the ε-term. Requirement (i) amounts to the obvious requirement that the same constant not be used to abbreviate distinct ε-terms; if the liberalized rule is to be used in tableaux for translations of ordinary language arguments containing names, (i) should be supplemented with a restriction to the effect that the constant introduced should not occur in starting formulas of the tableau. Requirement (ii) simply reflects the syntactic fact that an ε-term cannot contain itself as a proper subterm. To see the function of requirement (iii), consider a tableau for the fallacious inference from $\forall x \exists y Rxy$ to $\exists x \exists y(Rxy \,\&\, Ryx)$ in which it is violated. Applying the quantifier rules to the premiss, we get in succession (1) $\exists y Rey$, (2) *Rea*, (3) $\exists y Ray$, and (4) *Rae*. The application of the existential quantifier rule to (3) to obtain (4) meets requirements (i) and (ii) but infringes requirement (iii); (2) and (4) together with the negation of the conclusion are sufficient to ensure that the tableau will close. To see what has gone wrong, try to disabbreviate the constants to ε-terms. The constant *e*, when it is introduced in (4), functions as an abbreviation for an ε-term containing the constant *a*, but *a* in turn is an abbreviation for an ε-term containing *e*: *e*, when fully disabbreviated, would have to contain itself.

Smullyan's liberalized rule could be liberalized further (his (iii) could be replaced by something like Quine's Ordering), but his requirements are at least sufficient to ensure that all the constants introduced by the rule in a given tableau can be disabbreviated to ε-terms, guaranteeing soundness.

I hope that I have at least demonstrated with my sketchy soundness proofs that reference to Hilbert's ε-operator can provide a perspicuous and intuitively satisfying account of rules like EI. It remains to consider what implications there are for the correct analysis of informal reasoning and for practice in logical pedagogy and other fields (e.g. computer program verification) where the production of long formal derivations may be called for.

First a side issue. Fine suggests that his dependency diagrams can provide a useful algorithm for checking the correctness of derivations in systems with rules like EI. Some such algorithm may indeed be necessary in fields like program verification, where derivations of several hundred or more lines may be called for. From the pedagogical standpoint, however, I have two complementary worries about dependency diagrams. To present them with Fine's theoretical justification would require a fair bit of model theory *before* the deductive system for first-order logic is introduced, and this, even if the inversion of the usual order of topics is accepted, might seem to involve devoting an inordinate amount of time to mere mechanics. On the other hand, to present them without explanation, as an additional bit of formal work to be done mechanically each time a derivation is constructed, would simply reinforce the already wide-spread opinion that introductory logic is a semester of unmotivated, pseudomathematical hen-scratching, with no detectable relevance to the rest of the philosophy curriculum. In practice, the notation of Suppes [18], where instantial letters are subscripted with the terms on which they depend, would probably not be too long for use in derivations of the size likely to be produced by students in elementary or intermediate logic courses.

But should a rule like EI be used at all? Certainly it allows some derivations to be shortened, but at the expense of some complication in the system and its metatheory. The savings in derivation length, though important for some applications, will not be enough to justify the trade-off in a *pedagogical* context unless the rule is a natural one in itself. Fine argues that EI is more faithful than E.Q.Elim to the structure of intuitive reasoning, but I am not convinced. The formalization of the argument on p. 103 in a Fitch-style system will, to be sure, involve extra lines as the nested E.Q.Elim sub-proofs are set up, but it is entirely routine (and *any* full formalization will insert steps that would be left out in ordinary mathematical exposition). I see no reason not to think that someone uttering "There is a real, c, such that c is between a and b" is performing two speech acts in uttering a single sentence: *asserting* an existential quantification and *hypothesizing* that one of the reals satisfying the condition is called c. We must look deeper for the structure of intuitive reasoning.

One of the unlovely specimens whose derivation can be shortened by EI is $\exists x(\exists y Fy \supset Fx)$. This is a monstrous formula: no one finds its validity intuitive on first encountering it. I would, at least half seriously, argue that it is evidence in favor of the superior naturalness of Gentzen- or Fitch-style natural deduction that it can only prove this horror in a longer, indirect, way. The intuitionistic invalidity of the example raises a further point. (ε1) or its abbreviated form EI cannot be used without restriction in intuitionistic logic (cf. [14]) or in the systems of weak relevant logic without excluded middle currently being explored (cf. [1], [2]) as possible ways of avoiding the logical paradoxes. (The argument of [15] is fallacious because it ignores this point.) Leading to a final pedagogical consideration: logicians who think these non-classical logics worthy of study, either for their technical mathematical interest or for broader philosophical reasons, will want their beginning students taught methods of proof that generalize to the non-classical cases.

ACKNOWLEDGEMENT

I would like to thank the University of Melbourne for the light teaching load that allowed this to be written in term time, and J. Winther for processing my scribbles into words.

BIBLIOGRAPHY

1. Brady, R., 'The simple consistency of a set theory based on the logic CSQ', *Notre Dame Journal of Formal Logic* **24** (1983), pp. 431–449.
2. Brady, R., *Universal Logic*, in preparation.
3. Copi, I., *Symbolic Logic*, New York (Macmillan), 1954; second ed. 1961.
4. Fine, K., 'Natural deduction and arbitrary objects', *Journal of Philosophical Logic* **14** (1985), pp. 57–107.
5. Fine, K., *Reasoning with Arbitrary Objects*, Oxford (Blackwell), 1985.
6. Fitch, F., *Symbolic Logic: An Introduction*, New York (Ronald), 1952.
7. Gentzen, G., 'Untersuchungen über das logische Schliessen', *Math. Zeit.* **39** (1934), pp. 176–210, 405–437; Eng. tr. 'Investigations into logical deduction', *American Philosophical Quarterly* **1** (1964), pp. 288–306, and **2** (1965), pp. 204–218; repr. M. Szabo, ed., *The Collected Papers of Gerhard Gentzen*, Amsterdam (North Holland), 1969, pp. 20–131.
8. Hatcher, W., *The Foundations of Mathematics*, Philadelphia (Saunders), 1968; second edition *The Logical Foundations of Mathematics*, Oxford (Pergamon), 1982.
9. Hilbert, D. and P. Bernays, *Grundlagen der Mathematik*, vols. I and II, Berlin (Springer), 1934 and 1939.
10. Jeffrey, R., *Formal Logic: Its Scope and Limits*, New York (McGraw-Hill), 1967.
11. Kalish, D., review of [3], *Journal of Symbolic Logic* **32** (1967) p. 254.
12. Lambert, K. and B. Van Fraassen, *Derivation and Counterexample*, Encino CA (Dickenson), 1972.

13. Leisenring, A., *Mathematical Logic and Hilbert's ε-Symbol*, London (MacDonald), 1969.
14. Mints, G. E., 'Heyting predicate calculus with epsilon symbol', (in Russian), *Zapiski* of the *Steklov Institute* **40** (1974), pp. 101 – 109; Eng. tr. *Journal of Soviet Mathematics* **8** (1977), pp. 315 – 323.
15. Priest, G., 'The logical paradoxes and the law of excluded middle', *Philosophical Quarterly* **33** (1983), pp. 160 – 165.
16. Quine, W., *Methods of Logic*, New York (Henry Holt), 1950.
17. Smullyan, R., *First-Order Logic*, Berlin (Springer-Verlag), 1968.
18. Suppes, P., *Introduction to Logic*, Princeton (Van Nostrand), 1957.
19. Thomason, R., *Symbolic Logic: An Introduction*, New York (Macmillan), 1970.

Philosophy Department,
University of Melbourne,
Parkville, VIC 3052,
Australia.