I. INTRODUCTION

This paper is the beginning of an explication of the "normative-descriptive" or "ought-is" distinction by way of the notion that our knowledge of other minds is the result of our imposition of constraints on the interpretation of events as actions by agents. My hope is that a general theory of rationality and the normative can be derived from an examination of the constraints it is rational to impose on agent-interpretation, i.e., of the fundamental knowledge we have of persons as persons.

My attempt at an explication of the "ought-is" distinction takes the following form: I want to find an absolutely general way of determining when "ought"-sentences are true. Since the extensions of the account given below to interesting cases of "ought"-sentences such as moral and prudential cases depend on relatively complicated constraints on agent-interpretation, this paper will deal only with the simplest case of "ought"-sentences, the "logical ought". If logic is thought of as a normative science of belief, it yields one of the simplest cases of the "normative-descriptive" dichotomy.

By the "logical ought" I understand what might be called consequences of the canons of obedience to the laws of thought. An instance of such an "ought" occurs in "If you believe that frogs are green, you ought to believe that anything that's not green is not a frog." The "logical ought" is, as it were, the minimal rational "ought", the one that prescribes closure of belief under logical consequence and proscribes inconsistency of belief. It should be pointed out that the principles of the "logical ought" often come into conflict with other canons of rationality, just as principles of moral "oughts" come into conflict with each other. The example above is surely true even if
believing that non-green things are not frogs is punishable by death, and so, perhaps, not to be recommended. Resolution of such conflicts is a main project of the paper.

I start with the following simple-minded presumptions:

(A) “Ought”-sentences have truth-conditions. A theory of the meaning of “ought”-sentences consists of stating these truth-conditions in a general way by finding truth-value-preserving first-order paraphrase schemes. This is a reasonable presumption, since “ought”-sentences are surface declaratives and surface declaratives usually have truth-conditions.

(B) What a person ought to believe, given that he believes that \( p \), has something to do with what a person does believe, given that he believes that \( p \). Normative theories which characterize perfect rational agents must have something to do with real imperfect rational agents. Given that we are all rational agents, theories which prescribe rational behavior must in some sense and to some degree describe actual inference behavior. I want to make precise the notion that actuality is an approximation of perfection, as it were. The term “rational” in “is perfectly rational” is presumed not to be homonymous with “rational” in “is a rational agent”.

(C) I suppose there to be only one “ought”-construction occurring in logical, moral, prudential, and other contexts. I plan to do with “ought” what anybody would do with “possibly”. Rather than propose a theory with a vast number of homonyms for all of the contexts where “possibly” occurs, any theorist tries to find a way of reducing the possibility operators to one and explaining the differences by different relativizations. (I here take, e.g., S3, S4, and S5 to be alternative theories of “possibly”, not different actually occurring senses.) If there is an adequate theory which treats all “oughts” as semantically identical, it is preferable to other adequate theories which make “ought” multiply homonymous. The main constraint this imposes on an account of the logical “ought” is that the account must not be obviously inappropriate as a theory for other “ought” contexts.

(D) A consequence of presumption (C) is that no synonymy is presupposed between “obligation”-sentences and “ought”-sentences. In fact, since I am treating all “ought”-sentences as semantically the same, a special theory would be needed to show why “obligation” sentences seem to be able to replace
"ought"-sentences without loss in some human contexts. ("Ought" and "should" and "oblighed" and "must" seem to be the really synonymous pairs. The relation between the two pairs, I would say, is that "obligation" sentences codify some "rules of thumb" for how people should or ought to behave.) The point is just that "human-oriented" "ought"-sentences do not exhaust the range of "ought"-sentences. The sentences, "If you pull out the choke, your car ought to start" and "If the bookcase is only four feet wide, it should fit on the truck" are perfectly good English. Such examples, where the "obligatory" and "must" substitutions would lead, at best, to falsehood, provide good reasons for separating the investigation of "ought"-sentences from such fields as the logic of imperatives, deontic logic, and other such investigations of obligation-systems. The proof that such an identification need not be made (and, by the way, that a unified theory of "ought"-sentences is possible) is in the working-out of the alternative accounts sketched in the final section of this paper.

Truth-preserving paraphrases and definitions require truths to be preserved. One of the expository simplifications with the logical "ought" is that there are an infinity of relatively uncontroversial truths using this "ought". If "A" is a theorem of pure first-order predicate calculus, then instances of the form "X ought to believe that A" are true. If "A" is a logical consequence of "B", then substitution instances of the schema "If X believes that B, then X ought to believe that A" are true, unless "not A" is also a logical consequence of "B".

A confusion of "ought" with "ought, all things considered" and "oughts" which imply responsibility may render some of the above instances counter-intuitive. In the case of two-year-old children who can't be expected to be very rational, there is a sense, analogous to what I call the "automotive" "ought", in which it is false that they ought to believe a long logical truth. A way of reading "X ought to believe that A", at least until the theory below which clarifies the various "senses" of "ought" is presented, is as "It is reasonable for X to believe that A". The reason we don't fault children for "violations" of the logical "ought" is that it's unreasonable to expect them to be very reasonable. Similarly for the rest of us in the case of very long theorems, it's reasonable for us to believe them, but we are not to be expected to be perfectly reasonable.
II. THE LOGICAL FORM OF SENTENCES USING "OUGHT"

In this semantical section of the paper, I argue that sentences of the form "if . . . then . . . ought . . ." are unified sentence-containing contexts, i.e., that they do not consist of some sentence-connecting "if . . . then . . ." context combined with an "ought" context in the consequent. (This view is developed in the context of an overall theory of conditionals in [8], where it is claimed that all "conditionals" are two-place predicates of sentences.) What I need to make plausible is that sentences expressing "hypothetical oughts" and sentences expressing "conditional probabilities" have parallel logical forms and parallel logical features. It's not too important that the common logical form be the one I suggest.

The argument below first shows that "if . . . then . . . ought . . ."-sentences do not "detach" and that they create referentially opaque contexts throughout the whole surface conditional. These features won't be, strictly speaking, established, since their assignment to "ought"-sentences accounts for phenomena that various special devices have typically been used to account for. My account will be established only insofar as it yields a simpler theory. With regard to familiar alternative theories, I should say briefly that the major conflict between their theories and mine is specialization in "ought"-sentences that are plausibly treated as "obligation" sentences. Thus, strictly speaking, familiar alternative theories are not competing theories of the same data. (Similar remarks apply to the account of "if . . . then . . . probably . . ."-sentences.) If it is established that "if . . . then . . . ought . . ."-sentences are non-detachable and referentially opaque in both the surface "antecedent" and the surface "consequent", then there will be good reason to suppose that "if . . . then . . . ought . . ."-sentences are unified sentence-containers.

To show non-detachability, it suffices to produce a pair of true "if . . . then . . . ought . . ."-sentences with true surface antecedents and incompatible surface consequents. Consider the following four sentences, all of which could be true together:

(a) If you pull out the choke on Jones's car, it ought to start. (This is an instance of a general principle of car-starting theory.)

(b) If Jones's car is flooded, it ought not to start. (Another such instance.)
(c) The choke is out.
(d) Jones's car is flooded.

If Jones's car ought not to start if it is not the case that it ought to start, then, if the above "if . . . then . . . ought . . ."-sentences were conditionals with detachable consequents, compatible sentences would lead to contradiction.

An example of such failure of detachment with the "logical" "ought" arises if Jones has two beliefs which are subtly (for him) incompatible and have logical consequents which are contradictories of each other. Suppose the following sentences are all true:

(e) Jones believes that all animals are dangerous.
(f) Jones believes that pets are not dangerous.
(g) Jones believes that Fred the frog is a pet animal.
(h) If Jones believes that animals are dangerous and that Fred the frog is a pet animal, he ought to believe that Fred the frog is dangerous.
(i) If Jones believes that pets are not dangerous and that Fred the frog is a pet animal, he ought not to believe that Fred the frog is dangerous. (Another instance of the same principle.)

Once again, if "if . . . then . . . ought . . ."-sentences were compounded of conditionals which detached and "ought"-sentences, a set of truths would have contradictory logical consequents.

There are two qualifying remarks I should make. In the first place, there are "if . . . then . . . ought . . ."-sentences which do detach, but these consist of a detachable conditional combined with a categorical "ought"-sentence. I argue below that categorical "ought"-sentences are hypothetical "ought"-sentences with concealed "antecedents". Thus, "if . . . then . . . ought . . ."-sentences are syntactically ambiguous between "if . . . then . . . ought . . ."- and "if . . . then . . . (if . . . then . . . ought . . .)"-sentences. This syntactical ambiguity might explain the varied reactions to sets of sentences like the above. In the second place, it's certainly possible to claim that the sentences do detach by constructing an explanation of the above phenomena and the analogous phenomenon with
"if . . . then . . . probably . . ."-sentences. Such a solution would (in the version suggested in conversation by John Troyer) claim that in hypothetical "ought"-sentences of conditional probability, the conjunct "other things are equal" is implicit. Thus, the above failure of detachment would be an illusion, since the antecedents wouldn't turn out to be true.

To show that "if . . . then . . . ought . . ."-sentences create referentially opaque contexts, the strategy is to produce a true "if . . . then . . . ought . . ."-sentence which yields a falsehood when a co-referential term is substituted for one of its singular terms. To support the claim that "if . . . then . . . ought . . ."-sentences are unified sentence-containing contexts, the surface "antecedent" of the sentence must be shown to be referentially opaque as well as the surface "consequent". Then, since the peculiar non-detaching referentially opaque "conditional" only seems to arise with "if . . . then . . . ought . . ." and "if . . . then . . . probably . . ."-sentences, the "conditional" can plausibly be treated as a pair of two-place sentence-or proposition-operators rather than as a special conditional which only has "oughts" and "probabyls" in the consequent. To argue for the opacity in "ought"-sentences, I produce two examples, one "automotive", the other "logical".

"If you pull out the choke on Jones's car, it ought to start" is true, since it is an instance of the true generalized sentence, "If you pull out the choke on a car, it ought to start". The choke on Jones's car is in fact the third knob to the right on the only flooded Ford in Jarvis, New Jersey. But the sentence, "If you pull out the third knob to the right on the only flooded Ford in Jarvis, New Jersey, then it ought to start" is false, since nothing will make a flooded car start but letting it sit a while. The consequences an event ought to have seem to depend on how the event is described. More accurately, "ought-consequences" are associated with aspects of events or with events-under-a-description, not with events themselves.

"If Jones believes that all animals are dangerous and that Fred the frog is a pet animal, he ought to believe that Fred the frog is dangerous" is true, since a person ought to believe the logical consequences of what he believes. Jones, though, is the one person who believes that pets are not dangerous. It is not true that if the one person who believes that pets are not dangerous believes that all animals are dangerous and that Fred the frog is a pet animal, then he ought to believe
that Fred the frog is dangerous. Relative to this “antecedent”, it's not the case either that such a person ought to believe that Fred is dangerous or that he ought to believe that Fred is dangerous. When both a sentence and its negation are logical consequences of a belief, neither ought to be believed.

The above example cannot be totally convincing because the theorist has a choice of principles to give up. He can decide that “ought”-sentences are transparent but that instances of true generalized “ought”-sentences are not true. A theorist can also decide that, since the instance is false in the case of Jones’s car, the generalized “ought”-sentence is also false, thus again preserving transparency. The second course would leave us with no true generalized “ought”-sentences. Analogous remarks hold of “proofs” of referential opacity in any construction. “Referential opacity”, in effect, amounts in this case to a new label for the phenomenon of moral conflict, though it is generalized somewhat. That assignment of opacity is the choice to make to account for the above sentences can only be borne out by the rest of the theory.

Persuasive examples with the “logical” “ought” are somewhat hard to construct, since the only uncontroversial singular terms refer to believers. Also, given the above-mentioned syntactical ambiguity between “if . . . then . . . ought . . .”- and “if . . . then . . . (if . . . then . . . ought . . .)”-sentences, a kind of “all things considered” reading can lead the hearer to reject one or the other of the premises which would show the opacity rather than to accept the opacity. In cases like the above, people are even willing to give up the principle that a person ought to believe the logical consequences of what he believes. As Davidson has argued in [3], such cases cause no more difficulty than this case with a die with five “1”s and one “2”: This die is such that if Jones throws it for the only time in his life, it will probably come up “1”, but the die is also such that if the only person who will throw the die only once and get a “2” throws the die, he will probably not get a “1” even though he is identical with Jones. Just as a two-place predicate or operator theory of “if . . . then . . . probably . . .”-sentences solves the paradox in [5] for Hempel, so a relativized two-place predicate or operator theory of “if . . . then . . . ought . . .”-sentences will save such principles as that a person ought to believe the logical consequences of his beliefs.
A factor which discourages acceptance of the above results is the identification of "ought" in some cases with the corresponding "it is obligatory that" construction. "It is obligatory that" seems to create contexts which are transparent. Further, conditionals with such contexts seem to "detach" in much the same way that "it is necessary that" consequents detach. I claim that the general identification of "ought"-sentences and "it is obligatory that"-sentences is a mistake. "Ought"-sentences are formally like sentences of conditional probability, whereas "it is obligatory that"-sentences are analogous to "it is necessary that"-sentences. It is only in rather special cases that "ought"-sentences and "it is obligatory that"-sentences coincide in truth-value. Because of this, there is no argument here with theories of obligation or logic.

If "if . . . then . . . ought . . ."-sentences consist of a single operator which ranges over the entire apparent conditional and governs the content of both the surface "antecedent" and the surface "consequent", then there are two tactics a theorist can adopt in dealing with "categorical" "ought"-sentences. Either he can treat the "oughts" as homonyms, or he can find ways of building in concealed or understood "if"-clauses. These "if"-clauses will turn out to be understood first arguments of two-place predicates, much analogous to the "for a man" second argument that is filled in in understanding "Jones is tall". I choose the latter, theoretically preferable, tactic.

Given the opacity, the unity, and the non-detachability of the "if . . . then . . . ought . . ." construction, my first-order rendering of "If Jones believes that A, he ought to believe that B" is "(ES)(ES')(O(SS') & S means that Jones believes that A & S' means that Jones believes that B)".5 "O" is taken to be a metalinguistic predicate. Nothing much changes if S and S' are taken to be propositions or some intensional entity rather than sentences, except that "O" is treated as a relation between meanings of sentences rather than between sentences which have meanings. With propositions, the sentence would have the form "O(that Jones believes that A, that Jones believes that B)". Either way, non-detachability and referential opacity are accommodated, since predicates don't "detach" and since either named sentences or propositions create referentially opaque contexts. The question of the homonymy of occurrence of "ought" in automotive, logical, and moral contexts then becomes the question of whether there are several or only
one "O" predicates in English deep structures.

III. PRINCIPLES OF BELIEF

If the above or something equally unilluminating gives the logical form of sentences using "ought", then knowing the logical form of such sentences isn't going to be of much use in finding out when the two-place "O" predicate is true of a pair of sentences. To develop the part of the theory I need for the "logical" "ought", I state some principles of belief, interpreted both as principles of agent-interpretation and as a priori truths about believers.

The a priori principles of belief are "most of the time" principles and are false in stronger formulations. No principle is, as it were, absolutely binding on an interpreter in every case. Given sufficient simplification of the overall theory of the person in question or that the violation of the constraint for the particular sentence makes the believer satisfy other principles more often, any constraint can be violated in accounting for an action. For every person and for every principle, that person must believe in accord with that principle most of the time, but no principle is so powerful that any person must believe in accord with it all the time. The theory, as a theory of person-interpretation, tells us to maximize several on occasion incompatible features in systems of belief. As a theory of belief, it tells us that systems of belief and instances of belief satisfy statistical regularities.

I state the theory as a set of quantified conditional probability claims about belief:

(1a) If 'A' is a short contradiction, then if B is a believer, then B probably doesn't believe that 'A'.

(2a) If 'A' is inconsistent with 'C', then if B believes that 'C', then B probably doesn't believe that 'A'.

(3a) If 'A' is a short logical truth, then if B is a believer, then B probably believes that 'A'.

(4a) If 'A' and 'C' are sentences the first of which is a conditional, '[D] only if [E]', and the second of which is its antecedent, '[D]', then if B assents to 'A' and to 'C', then B probably believes that 'E'.

(5a) If a sentence 'A' is a logical truth and 'D' is a substitution
instance of \( A \) which doesn't differ much in length from \( A \), then if \( B \) believes that \( A \), then \( B \) probably believes that \( D \).

The above principles about belief are stated in terms of conditional probability. That is, if \( A \) is a short logical truth, then given that \( B \) is a believer, \( B \) probably believes that \( A \). Relative to a conjunction of \( B \) is a believer" with further information, e.g., that \( B \)'s guru says that \( A \) is false, it may be improbable that \( B \) believes that \( A \). I should further point out that these principles are vastly weaker than the principles which generate the infinity of logical "ought"-structures. The next section's task is to generate those principles out of these.

"If . . . then . . . probably . . ." is an "evidence-giving" relation (so that, e.g., logical truths aren't second arguments for any "If . . . then . . . probably . . ." truths). I leave aside non-formal questions about probability. Formally, I take "If . . . then . . . probably . . ."-sentences to be two-place predicates with sentences as arguments. Once again, many other entities will do as arguments.

The standard formal theory of conditional probability with which my theory begins, treats \( \text{pr} \) as a two-place function of some kind of entity as arguments and a number as value. The form of "The probability that the car will start, given that you pulled out the choke is 0.7" is \( (ES)(ES')(\text{pr}(S,S') = 0.7 & S \text{ means that you pulled out the choke & } S' \text{ means that the car will start}) \). (From now on written as \( \text{pr} \text{ (you pulled out the choke, 'the car will start') } = 0.7 \).) The "If . . . then . . . probably . . ." construction is then defined as follows:

\[ (S)(S')(P(S,S') \iff \text{pr}(S,S') \text{ is large for a pr-value}) \].

The logical form of "If you pull out the choke, your car will probably start" is:

\[ (ES)(ES')(P(S,S') & S \text{ means that you pull out the choke} & S' \text{ means that the car will start}) \].

In the above definition, I leave the vague predicate "is large" primitive, since "if . . . then . . . probably . . ." is vague. There is no sharp boundary between cases where "\( P(S,S') \)" is true and "not-(\( P(S,S') \))" is true in terms of \( \text{pr} \)-values, just
as there is no sharp boundary between \textit{pr}-numbers which are large and ones which are not large for \textit{pr}-numbers.

By taking \textit{“P”} to be a metalinguistic predicate of pairs of sentences, I take \textit{“if . . . then . . . probably . . .”} sentences to be opaque and non-detachable. Exactly the considerations which showed \textit{“if . . . then . . . ought . . .”} to be a single unitary operator apply to \textit{“if . . . then . . . probably . . .”} as well. The reader can recall the examples with Jones’s automobile and replace \textit{“ought”} with \textit{“probably”}. Once again, an ambiguity between \textit{“if . . . then . . . probably . . .”} and \textit{“if . . . then . . . (if . . . then . . . probably . . .)”}, that is, between conditional probability readings and readings as conditionals with \textit{“all things considered, probably”} consequents, can cause confusion and misinterpretation of examples. The reader is referred to Hempel [5] for a fuller account of these phenomena.

The fundamental theory of believers built into common sense and codified by Quine could now be stated as quantified \textit{“P”}-sentences. As I have written them below, the quantifications are, strictly speaking, nonsense, since variables range over both things and expressions. This flaw can be repaired at great cost in legibility with a concatenation function \textit{(C)} and by the use of a denotation relation \textit{(Den)} which relates words and what they name. For example, \textit{“(x) (x assents to ‘x is here’)”} becomes, in a conscientious notation, \textit{“(x) (Ey) (Den(y) = x & x assents to C(y, “is here”))”}. For the sake of legibility, I have stuck to sloppy quantification:

\begin{align*}
(1b) & \quad (x)(S)(S \text{ is a short contradiction only if } P(\neg x \text{ is a believer}^1, \neg x \text{ doesn’t believe } S^1)). \\
(2b) & \quad (x)(S)(S') (S' \text{ is inconsistent with } S \text{ only if } P(\neg x \text{ believes } S'^1, \neg x \text{ doesn’t believe } S^1)). \\
(3b) & \quad (x)(S)(S \text{ is a short logical truth only if } P(\neg x \text{ is a believer}^1, \neg x \text{ believes } S^1)). \\
(4b) & \quad (x)(S)(S')(S'' \text{ is the conditional } S' \text{ only if } S''^1 \text{ only if } P(\neg x \text{ believes } S \& x \text{ believes } S'^1, \neg x \text{ believes } S''^1)). \\
(5b) & \quad (x)(S)(S') (S \text{ is a logical truth } \& S' \text{ is not much different in length from } S \& S' \text{ is a substitution-instance of } S \text{ only if } P(\neg x \text{ believes } S^1, \neg x \text{ believes } S'^1)).
\end{align*}
IV. THE LOGICAL “OUGHT”

As a preliminary to giving a definition of “ought”, I describe in rough terms the relation between rational agents (which we all are) and perfect rational agents (believers who believe a complete set of axioms and believe logical consequences of what they believe). A logically perfect believer is a believer who, in every step of a chain of changing beliefs, believes what any believer would probably believe, if he had arrived at the previous point in a chain of changing beliefs. For the perfect believer (structurally speaking), the “P” principles which characterize all believers can be replaced by universally quantified conditionals (or “is large” can be replaced by “= 1”). A person who always believed in accordance with the constraints maximization of which we impose on the interpretation of others would believe all logical truths and all logical consequences of what he believed.

To see this, consider a very long theorem of the first-order functional calculus. By the Gödel completeness theorem, there is a proof of the theorem. Consider a proof of that theorem, in which each step is either an axiom or derived from an earlier step by modus ponens or by application of a substitution rule. By principle (3), a perfect believer believes all the short axioms of the first-order functional calculus, since for any believer and any short logical truth, that believer probably believes that logical truth. If the transition from one step of the proof to the next is by modus ponens, principle (4) says that the perfect believer will believe the succeeding step, since any person who believed all previous steps (which the perfect believer does) would probably believe the next. If the transition is by the application of the substitution rule, then there is a series of applications of the substitution rule such that the result of each successive application of the rule is near enough in length to the previous result that principle (5) applies and such that it ends with the succeeding step in the original proof. In this series of intermediate steps, the perfect believer will believe every step, since any believer who believed all previous steps would probably believe the succeeding step. Since every proof can be thus broken down into steps such that anyone who believed the contents of all previous steps would probably believe the next, the perfect believer—the person who, structurally speaking, always believes what everyone would probably
believe given his beliefs—believes all theorems and, therefore, by conditionalization and modus ponens, all logical consequences of his beliefs. With any very long proof, it will be very improbable that any actual rational agent will get through the whole proof, i.e., end up believing the theorem, even though, if he had gotten to any particular point in the proof (i.e., if he believed everything in the proof so far), he would probably have gotten to the next step. We would expect habitual failure to believe long theorems and distant logical consequences of beliefs in actual cases because of the principle of multiplication of probabilities. When there is a chain of contents of beliefs such that anyone probably holds the beginning belief and anyone who held all the previous beliefs in the chain would probably hold the next, it is true that anyone ought to believe that last element of the chain, given that he is a believer.

In showing the relation between the person who believes what he ought to believe, in the “logical” sense, and real believers, I need to make the above sketch precise by giving a definition of “O” in terms of “P”. To be completely general and not restricted just to “logical” “oughts”, the definition must be more complicated than it strictly needs to be to accommodate just the logical case. The logical case is easier because it is much more obvious that the definition is truth-preserving. When the definition has been stated and explained, I will indicate how other “senses” of “ought” can be generated by changing the first argument of the “O” relation, given that the appropriate extensions and strengthenings of the theory of agent-interpretation can be justified.

V. THE DEFINITION OF THE PREDICATE “O” IN TERMS OF THE PREDICATE “P”

I will first define the “O” predicate in terms of probability-chains and then show how the definition restricts the probability-chains which can justify the various clauses in the definition. In the following definition, “PC(A,B)” is defined just like “O(A,B)” except that clause (g) is deleted. “L” abbreviates “the length of”, a purely syntactical notion, and is used with the number 11,562 in order to set an arbitrary upper bound on how many consequences of the first argument of the “O”-sentence are drawn. For the purposes of the following definition, a sentence which doesn’t have the form of a conjunction is considered to be a one-conjunct conjunction, and the
single sentence is both the first and last conjunct. I letter the clauses of the definition for ease of reference:

"O(A,B)" is true if and only if either there is an S such that:

(a) ((S is a conjunction) &
(b) (y)(y is a conjunct in S only if y has the form P(C,D)) &
(c) the first argument in the first conjunct of S is A &
(d) the second argument in the last conjunct of S is B &
(e) (y)(y is a conjunct of S only if the first argument of y is the conjunction of the previous arguments occurring in previous conjuncts of S and A) &
(f) (C)((L(C) is less than 11,562 times L(A) & P(A,C)) only if C occurs as an argument in a conjunct if it can before any D such that not-(P(A,D)) occurs in S) &
(g) not-(PC(A, not-B)) &
(h) S is true)

or "O(A,B)" is an instance of a true generalized "O"-sentence.

For sentences of the form "(x_1) . . . (x_n)(O(F(x_1, . . . , x_n), G(x_1, . . . , x_n)))", the modification that Scan be a quantified conjunction is required. Since names of open sentences will be occurring in the chain, clauses of the definition have to be altered for the general case as well.

I show how this definition works, that is, how it restricts probability-chains so that only those corresponding to true "if . . . then . . . ought . . ."-sentences fit the definition, by considering the following truth: "Jones ought to believe that if whoever saw Lola in Harry's tavern with Joe and Fred last Saturday night will surely see Lola in Mike's garage with Phil; then if it's not the case that whoever saw Lola in Harry's tavern with Joe and Fred last night will surely see Lola in Mike's
garage with Phil, then Lola, Joe, Fred, Mike, and Phil cannot all be freshmen." This is true just because the content of the supposed belief is a logical truth. The sentence has the form \( O(A,B) \), where "A" is the sentence, "Jones is a believer", the suppressed first argument. In passing, the sentence "If Jones believes that whoever . . ., then Jones ought to believe that if it's not the case . . ." will be dealt with and shown to be justified by an analogous chain.

The chain will start with some substitution instance of the conjunct, "\( P("Jones is a believer", "Jones believes X") \)", where \( X \) is some sentence that a person probably believes, given that he is a believer. Any short logical truth, for instance, "Fred floats and doesn't float only if grass grows" ("\(((p \& \sim p) \rightarrow q)\)"), would give us a truth if inserted for \( X \). The second conjunct might be "\( P("Jones is a believer and Jones believes ((p \& \sim p) \rightarrow q)\), "Jones believes that if Fred floats and grass grows only if rabbits run, then if Fred floats only if grass grows, then Fred floats only if rabbits run")". ("\(((p \& q) \rightarrow r) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow r))\)".) Both of these conjuncts have second arguments which are immediate "\( P\)-consequences of "Jones is a believer", by the principle that if a sentence is a short logical truth, any believer probably believes it. The first argument of the second conjunct is a conjunction because of clause (e). The application of clauses (a), (b), and (c), so far should be obvious.

Because of clause (f), the next several million conjuncts will have second arguments which are short logical truths. Clause (f) requires that every direct "\( P\)-consequence of "Jones is a believer" that can be built into the chain be built into the chain. Since previous second arguments of conjuncts in the chain are now conjuncts in the first argument of every succeeding conjunct, it should be obvious that not every direct "\( P\)-consequence will be built in, given a way of starting the chain. Some things are probable given \( A \) which are not probable given \( A \& B \& C \& D \& E \), even though each of \( B, C, D, \) and \( E \) are probable relative to \( A \). The idea here is to get a "complete content" of "Jones is a believer", as will be explained in the particular account of clause (f), below. The "11,562" section of the clause is an arbitrary device to capture a full content of "Jones is a believer" without requiring the chain to run on forever.

So far, then, we have "\( P("Jones is a believer", "Jones believes
that \(((p \& \sim p) \rightarrow q)\)\) and \(\ldots \& P\) ("Jones is a believer and
Jones believes that \(((p \& \sim p) \rightarrow q)\) and Jones believes that
\(((p \& q) \rightarrow r) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow r))\) and Jones believes
that \((q \rightarrow (p \rightarrow q))\) and \(\ldots\).\), where in place of the first
set of dots are several million conjuncts and in place of the
second set are all of the second arguments of those conjuncts.
Now the chain can start to contain second arguments which
are not direct "\(P\)"-consequences of "Jones is a believer". One
of these might be "Jones believes that \(((p \& \sim p) \rightarrow q) \rightarrow
(p \rightarrow ((p \& \sim p) \rightarrow q)))\), in which the content of belief isn't
much longer than "\(q \rightarrow (p \rightarrow q)\)". By principle (5b) of person
interpretation, a person who believes that \(q \rightarrow (p \rightarrow q)\), then,
probably believes that \(((p \& \sim p) \rightarrow q) \rightarrow (p \rightarrow ((p \& \sim p)
\rightarrow q))\). The next conjunct of the chain then will contain in
its first argument conjuncts to the effect that Jones believes
that \(((p \& \sim p) \rightarrow q)\) and that Jones believes that \(((p \& \sim p)
\rightarrow q) \rightarrow (p \rightarrow ((p \& \sim p) \rightarrow q))\). By principle (4b), then, Jones
probably believes that \(p \rightarrow ((p \& \sim p) \rightarrow q)\), given that he
has all the beliefs attributed to him in the first argument.

Since the first argument in the next conjunct contains "Jones
believes that \(((p \& q) \rightarrow r) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow r)))\), and
since "\(((p \& (p \& \sim p)) \rightarrow q) \rightarrow ((p \rightarrow (p \& \sim p)) \rightarrow (p
\rightarrow q)))\)" is not much longer than "\(((p \& q) \rightarrow r) \rightarrow ((p \rightarrow q)
\rightarrow (p \rightarrow r)))\), the second argument of that conjunct can
be "Jones believes that \(((p \& (p \& \sim p)) \rightarrow q) \rightarrow ((p \rightarrow (p
& \sim p)) \rightarrow (p \rightarrow q)))\)\), by principle (5b) again. It is obvious
that succeeding second arguments will get us to a second
argument, "Jones believes that \((\sim p \rightarrow (p \rightarrow q))\) or "Jones
believes that if it's not the case that Fred floats then Fred
floats only if grass grows". To get our original "ought"-sentence's
content as a second argument, it suffices to make longer
substitution-instances of this logical truth, each substitution-in-
stance not much longer than the previous one, until we get
the second argument we need.

The existence of such a chain of true "\(P\)"-sentences asserted
by the "\((ES)\)" is what it takes for the original "\(O\)"-sentences
to be true. This doesn't mean that in asserting the "\(O\)"-sentence
we claim to know the chain that makes it true. It is important
to distinguish claims about logical form from claims about the
truth-conditions of particular predicates. This is a claim about
the truth-conditions of the predicate "\(O\)". The rest of this
section will consist of a clause-by-clause account of the definition
and its truth-preserving proclivities:

(a) "S is a conjunction".

The last clause, "S is true", must have the result that "O(A,B)" is true only if each "P"-clause in the sentence which shows the "O"-sentence to be true is true.

(b) "(y) (y is a conjunct in S only if y has the form P(C,D))".

This clause requires every conjunct to be a conditional-probability sentence.

(c) "The first argument in the first conjunct is A".

This clause makes the relativization in the "O"-predication a relativization in every "P"-clause. In showing that Jones ought to believe a theorem, "A" is "Jones is a believer", an "understood" relativization. In showing that when "D" is a logical consequence of "C", if Jones believes that C, he ought to believe that D, "A" is "Jones believes that C".

(d) "The second argument in the last conjunct is B".

This clause says that B is something that is probable given A, given what is probable given A, given what is probable given what is probable given A, etc., as far as the chain goes on. Any appropriate conjunction will have as many true "O"-predications corresponding to it as there are second arguments in it.

(e) "(y)(y is a conjunct of S only if the first argument of y is the conjunction of the previous arguments occurring in previous conjuncts of S and A)".

This clause makes three contributions to the adequacy of the definition.

Firstly, it keeps the conjunction from going off on side-tracks, as it were. No conjunction which meets the definition, given this clause, selects one thing L which one argument makes probable, then takes one thing M out of the things L makes probable, and so on. This clause requires that each new second
argument be probable relative to everything that has come before it in the conjunction. So one class of counterexamples designed to show that the definition makes "O" hold between sentences A and B such that it is not the case that if A then B ought to be true, is eliminated.

Secondly, this clause keeps extraneous material out of the conjunctions. New truths can't be brought into first arguments, since a new truth won't be A and won't have occurred in previous conjuncts. This clause helps keep "ought"-sentences strictly relative to A when A is the first argument. Bringing in extraneous truths would produce counterexamples by yielding true conjunctions which correspond to false "O"-sentences. If other truths could be inserted in first arguments, then, since \( P(A \& B_1 \& \ldots \& B_n, A) \), it would follow that if Jones believes that \( 2 + 2 = 4 \), Gibralter ought to belong to the British.

Thirdly, this clause specifies the form of the conjunctions which make "O"-sentences true. In conformity with this clause, a five-member conjunction making "O(A,B)" true must have the form, "\( P(A, C) \& P(A \& C, D) \& P(A \& C \& D, E) \& P(A \& C \& D \& E, F) \& P(A \& C \& D \& E \& F, B) \)". At each succeeding conjunct, the first argument becomes more complex and the things the argument makes probable become more restricted.

\[ (f) \quad "(C)((L(C) is less than 11,652 times L(A) \& P(A, C)) only if C occurs as an argument in a conjunct if it can before any D such that not-(P(A, D)) occurs in S)". \]

This clause is also designed to avoid sidetracks which would lead to counterexamples and to make the outcomes of chains more determinate, given an argument A. What it says roughly is that as many of the really different parts of the "content" of the first argument as possible, given a way of starting to draw consequences, must be in the chain before any indirect "P"-consequences of A, (that is, "P"-consequences of "P"-consequences of A which are not "P"-consequences of A) are drawn. In general, given a way of starting to draw "P"-consequences, not every direct "P"-consequence of A can go in a chain. That is, if we have already built in the direct consequences C and D, then the only way we can build in E is if \( P(A \& C \& D, E) \). But "\( P(A, E) \)" is compatible with "not-(P(A & C \& D)"
Thus, this clause allows as \( \text{"P"-consequences} \) of the continuing chain only ones which are also \( \text{"P"-consequences} \) of \( A \), until we run out of direct \( \text{"P"-consequences} \) of \( A \) which are also \( \text{"P"-consequences} \) of the continuing chain.

The restriction on available \( C \)'s that they be less than 11,652 times as long as \( A \) sets an arbitrary upper bound within which all really different parts of the content of \( A \) can be stated. The idea of this restriction is that any direct \( \text{"P"-consequence} \) of \( A \) which cannot be stated in any sentence less than 11,652 times the length of \( A \) is going to come to roughly the same thing as some sentence which is also a direct \( \text{"P"-consequence} \) of \( A \) and which is less than 11,652 times the length of \( A \).

There is still some freedom in constructing chains, because it matters what direct \( \text{"P"-consequence} \) of \( A \) the chain starts with and what the next one is, etc., since some ways of starting will allow a direct consequence of \( A \), say \( C_{12} \), where other ways won't, if \( C_{12} \) isn't a \( \text{"P"-consequence} \) of the conjunction of \( A \) and the \( \text{"P"-consequences} \) with which the chain started.

\( \text{(g) \ "not-(PC(A, not-B))"} \).

This clause excludes the possibility that we could get two chains with the same first argument in the first \( \text{"P"-sentence} \) showing true both \( \text{"O(A,B)"} \) and \( \text{"O(A, not B)"} \). \( \text{PC(A,B)} \), or \( \text{there is a probability-chain from A to B} \), is defined just like \( \text{"O(A,B)"} \) except that its definition doesn't have this clause. Thus, \( \text{"O(A,B)"} \) can be defined as \( \text{PC(A,B) & not-(PC(A, not-B))} \). What the clause says is that if a pair of chains can be produced from the same first argument, one ending in a sentence and the other in the negation of that sentence, then neither \( \text{"O(A,B)"} \) nor \( \text{"O(A, not-B)"} \) is true. If we can get both results, then neither is something that ought to be relative to \( A \). This explains why nothing follows about what a person ought to believe given his belief in a contradiction. (The supplementary step, \( \text{P("Jones believes not-A", "not-(Jones believes A)")"} \) is required for this explanation).

\( \text{"... or \ "O(A,B)" is an instance of a true generalized "O"-sentence.} \)

That is, \( \text{"O(A,B)"} \) is an instance of a sentence of the form \( \text{"(x_1) ... (x_n) (O(C,D))"} \) which is true. This clause makes
one distinction between “ought” and “ought, all things considered”. Suppose we have a special theory about Jones; we know his situation and beliefs. These beliefs can be such that even though $B$ is not a logical consequence of $A$, if Jones believes $A$, he ought to believe $B$. This could happen if Jones is in a peculiar epistemic situation. In speech, the distinction is made by emphasizing “Jones”, as in “Jones ought to believe $A$” or “if Jones believes $B$, he ought to believe $A$".

The sentence “$O(\text{"Jones believes } A\text{"}, \text{"Jones believes } B\text{"})" has two ways of being true which don’t always work together. There can be a true generalized “ought”-sentence, “$(x)(O(\text{"x believes } A\text{"}, \text{"x believes } B\text{"}))$”, of which “$O(\text{"Jones believes } A\text{"}, \text{"Jones believes } B\text{"})$” is an instance and so true by this clause. But there can be “$P$”-truths of the form “$P(\text{"Jones believes } A\text{"}, F\text{"Jones"})$” (where $F$ is an open sentence), while there is no “$P$”-truth of the form “$(x)(P(\text{"x believes } A\text{"}, Fx))$”. The “oughts” in these cases may often be justified by one-conjunct conjunctions and so be equivalent to an “if . . . then . . . probably . . .”-sentence in much the way that automotive “ought”-sentences are.

Thus, a sentence like “If Jones believes $A$, he ought to believe $B$” is several ways ambiguous, though many of these ways are pragmatically, not syntactically or semantically, differentiated. Much long explanation could be given of how we are able to “tell what someone means” by one of these sentences in a context but won’t be here.

VI. OTHER FIRST ARGUMENTS FOR “O”-SENTENCES

A claim of this paper is that the above definition of the “$O$”-predicate gives the truth-conditions of every sentence using “ought” in English. I can’t demonstrate this in a paper of this length, but I can say roughly how it is done and what other a priori principles of person-interpretation there have to be for the definition to be truth-preserving for cases other than the logical “ought”.

The differences among the “rule of thumb”, logical, epistemic, prudential, and moral “oughts” are differences in their first arguments and differences in the kind of “$P$”-principles, i.e., constraints on person-interpretation, that their justifications appeal to.

The first arguments of logical “oughts” are belief-sentences, as we have seen, and the second argument is also a belief-sen-
The relevant “P”-principles for the logical “ought” are all structural principles which say what the interrelations among elements of a system of beliefs must be.

The first arguments of epistemic “oughts” are either ascriptions of knowledge or belief or sentences about the location and circumstances and history of the agent; e.g., “If he’s been to Ohio, he ought to know where Granville is.”

In the case of both prudential “oughts”, the first argument includes a want-sentence, i.e., either something like “Jones wants a peach” or “Jones is a wanter” (“(E.S)(Jones wants-true S)”).

The first argument of a moral “ought”-sentence is an ascription of agenthood or personhood; e.g., “Jones is a person.” In the moral case, that is, no special content about the particular interests or beliefs occurs in the first argument, so that nothing but “P”-consequences of Jones being a person qua person can start the chain.

Both moral and prudential “ought”-sentences typically have intentional action sentences as second argument. Intentional action sentences I take to be analyzed as the propositional attitude “makes-true”, so that “Jones bites his sister” has the form “(E.S)(Jones makes-true S & S means that Jones bites his sister)”. Roughly, “makes-true” or “does” in the intentional sense is to “wants” as “knows-true” or “knows” is to “believes”. Both knowing and doing have both external and internal truth conditions in that truth, plus possibly some causal chain, is required in addition to the agent being in a propositional state.

“Rule of thumb” “ought”-sentences are in a way the most general case of “ought”-sentences. One way of phrasing my general claim about intentional systems and their application is that our predictive theories about agents framed in terms of beliefs, desires, and actions are mere “rule of thumb” theories at best and essentially unrefinable to anything predictively much better. Wherever there is a “rule of thumb” body of knowledge, applications of that body of knowledge are often in the form of “ought”-sentences. From basically statistical principles about cars, we know that if you pull out the choke, the car ought to start. It is not hard to see how these “ought”-sentences work and how our definition is truth-preserving for these cases, if we allow that pragmatic grounds can stop us from saying automotive “ought”-sentences which are made true by very long conjunctions. Just about any first argument can occur in the “rule of thumb” “ought”-sentence, since so much of
our knowledge of the world is "statistical". What function analogously to belief-constraints in these cases are just generalized probabilistic, "ceteris paribus" true beliefs.

For the intentional system "ought"-sentences, the ones which interest us, the argument that my definition is truth-preserving must take two stages. I first have to show that there are true "P"-principles about wanting and being a person, and then I have to show that these principles are adequate to generate all and only the right "P"-chains. I will sketch briefly the first stage. The cases I will deal with are the epistemic, "structural"-prudential, and moral "ought"-sentences. What I will be limited to doing is saying what kind of constraints on agent-interpretations have to be shown to be rational to show that my definition applied to these types of "ought"-sentences come out true when and only when they do.

Truths with the epistemic "ought" are truths about what a person ought to believe which are made true by "P"-principles other than (1a)–(5a) and which have a belief-sentence or a situation-sentence as first argument. Constraints on agent-interpretation which are relevant to the epistemic "ought" will be constraints or truths about knowledge and belief, at least. As an example of such a constraint, there is Quine's dictum ([6], Ch. 2) that people usually have to hold correct beliefs (have knowledge) about what's happening in stimulus situations, but not much else. When a complete set of such constraints are found, they will amount to a codification of basic perception-theory and inductive logic. The "P"-principles required for the truth of epistemic "ought"-sentences will be internal structural principles as well as at least situation or truth-principles (e.g., "(x)(P("x is a believer & x is in front of a cow", "x knows he is in front of a cow"))").

By the "structural"-prudential "ought" I have in mind "O"-sentences that follow from basic decision theory. The first arguments of such "ought"-sentences consist of a conjunction of a set of want-sentences and a set of belief-sentences. Decision-theory will say what is a rational desire relative to those wants and beliefs. Several people have argued that decision-theoretic consistency and rationality is imposed on agent-interpretation in roughly the way that first-order logical consistency and rationality is. In this case, like the case with the logical "ought", we know what the constraints should be like, since decision theory is already codified. All that needs to be done to find
basic truths about wanters is to take the axioms and rules of the calculus of probabilities and convert them to \("P"-sentences. The truth of "structural"-prudential "ought"-sentences depends entirely on structural principles about desires.

The prudential "ought" differs from the "structural"-prudential "ought" in that, in this case, the first argument consists of want-sentences alone, either the ascription of a particular want or set of wants to a person or the statement that the entity in question is a wanter (i.e., \((E_S)\text{ (Jones wants-true } S)\)). The knowledge that is required to yield a rational action relative to the person's wants will be knowledge that anyone probably has, given that he has those wants. In order to get truth-preservation for real prudential "ought"-sentences, all relevant knowledge has to come into the \("P"-\)chain as something any person with those wants ought to know. Basically, this means that constraints on agent-interpretation have to be strong enough to enable us to form a \("P"-\)-chain starting just with the sentence that a thing has a particular set of wants and ending with any truth. Thus, in addition to structural constraints on wants, prudential "ought"-sentences require for their truth very strong content-constraints on assignment of beliefs and knowledge.

The question whether the above definition is truth-preserving in the case of the prudential "ought" turns out to be the question of whether there is a truth-constraint among the epistemic constraints. Davidson, in arguing that seeking agreement with another person is a rational procedure in translation ([2]), is basically claiming that it is rational in agent-interpretation to make as many of the beliefs of the other true as possible, and thus claiming that most of the beliefs of anything translatable are true. If he is correct, then for any truth \(S\), \((x)(P("x \text{ is a believer } \& S \text{ is true}", "x \text{ believes } S"))\). To avoid the somewhat counterintuitive sound of this principle, it's worth remembering that it doesn't imply that \((x)(P("x \text{ is a believer } \& S \text{ is a newly discovered truth of quantum mechanics}", "x \text{ believes } S"))\). The generalized principle is a conditional probability claim relative just to \(S\) being a truth and \(x\) being a believer. Since what is needed for the prudential "ought" are truths to apply the person's given wants to, the only way I can see to get the prudential "ought" is by an agreement-constraint (which is a "truth-to-the-best-of-our-knowledge" constraint) or some other constraint which will have the effect of being a truth-constraint.
An agreement-constraint or a truth-constraint (of which Quine's stimulus-situation correctness constraint used for the epistemic "ought" is a weak example) is a departure from logical and decision-theoretic constraints in that it is a content constraint rather than a structural constraint. I should make this distinction explicitly. Structural constraints codify knowledge to the effect that any agent's system of wants and beliefs must have certain internal inter-relationships and constrain agent-interpretation in accordance with this knowledge. Content constraints codify knowledge about what any agent probably believes and probably wants, and constrain agent-interpretation so that most such wants and beliefs are imposed on others. Content-constraints, roughly, rule out the possibility of their being agents with structurally sound but very stupid systems of wants and beliefs.

In the case of the moral "ought", there have to be content-constraints on both wants and beliefs for my definition to be truth-preserving. Moral "ought"-sentences take nothing but personhood as a starting point and thus must have want-content principles in order that "P"-chains end up with actions. ("Hypothetical" moral "ought"-sentences are genuine conditionals. Moral "ought"-sentences about what Jones ought to do given his particular situation depend on being able to build truths, particularly truths about Jones, into the "P"-chains by means of the belief content-constraints. The moral "It ought to be the case that . . ." has the form "(x)(O("x is a person", "x wants that . . ."))", i.e., anyone ought to want that . . .) The perfect rational agent, morally speaking, has the correct basic wants and true beliefs and calculates correctly. I have already indicated how truths can be introduced into "P"-chains by agreement-constraints; the case of correct basic wants is similar, given that the notion of a basic want, or a want which is not derived from beliefs, can be made clear. Roughly, there has to be a basic-want agreement constraint, so that any desire that is a basic desire of mine is, to the best of my knowledge, probably shared by any other. This will yield "What a person, in my opinion, ought to do." To get what a person really ought to do, morally speaking, we need to find truths of the form "(x)(P("x is a person", "x wants A'"))".

That it is rational to apply some such constraints in agent-interpretation can be seen from the following kind of example: A person drops a rock on his toe and suffers great pain. An explanation which we know to be wrong, even though
it is structurally sound and involves no false beliefs, is that
the person wanted pain and thought that dropping a rock
on his toe would be a good way to get it. That is, we know
the following about agents: they probably don't want pain.
Which moral "ought"-sentences are true depends on what is
true and what content-constraints there are on wants. In the
case of the moral "ought", there are, unfortunately, fewer
undisputed truths to be preserved by definition. A correct theory
of "ought"-sentences will, I hope, be confirmed by sentences
with first arguments other than "is a person", and will then
be applied to help us discover what the true moral "ought"-sen-
tences are.7

REFERENCES

[7] Samuel C. Wheeler III, The Logical Form of Ethical Language: Radical Translation,

NOTES

1 These extensions are stated and argued for in [7].
2 Thus, the views we express about "ought"-sentences aren't really in conflict
with the work of Von Wright or with Castañeda [1].
3 The non-detachability of "ought"-sentences is well-established by Donald
Davidson in [3]. What follows is my rendering of what I take to be his conclusions.
4 E.g., on the part of writers like W. D. Ross and Castañeda in [1]. Some
of the alternative reaction seems to come from concentration on "oughts" which
can be read as "obligeds".
5 Means that" can be read by propositionists as a two-place relation between
a sentence and a proposition. My own version of "means that" is based on Davidson's
view that a truth-definition is a theory of meaning, stated in [2]. I analyze ". . .
S means that the Earth is round" as ". . . 'S is true if the Earth is round' is
a logical consequence of the-best-theory-from me to the utterer of S". "The
best-theory-from" is a two-place function with ordered pairs of people as arguments
which yields a conjunction of sentences (a finitely stated truth-definition) as value.
We can refer to such theories without knowing what they are or how to find
them. "Is a logical consequence of" is a two-place relation between sentences.
6 This point is made in [4].
I would like to thank Clark Glymour, Gilbert Harman, and John Wallace for their help with the dissertation from which this paper is derived, as well as John Troyer and the referees for advice and comments on the paper itself.