# New Essays on the Knowability Paradox

Edited by

JOE SALERNO

OXFORD
UNIVERSITY PRESS

# *Contents*

PART I: EARLY HISTORY

PART II: DUMMETT'S CONSTRUCTIVISM

PART III: PARACONSISTENCY AND PARACOMPLETENESS

## PART IV: EPISTEMIC AND TEMPORAL OPERATORS: ACTIONS, TIMES AND TYPES

## PART V: CARTESIAN RESTRICTED TRUTH

## PART VI: MODAL AND MATHEMATICAL FICTIONS

## PART VII: KNOWABILITY RECONSIDERED

# Introduction

*Joe Salerno*

## The Knowability Paradox

In his seminal paper A Logical Analysis of Some Value Concepts (1963; reprinted, Chapter 2 of this volume), Frederic Fitch articulates an argument that threatens to collapse a number of modal epistemic distinctions. Most directly, it threatens to collapse the existence of fortuitous ignorance into the existence of necessary unknowability. For it shows that there is an unknown truth, only if there is a logically unknowable truth. Fitch called this 'Theorem 5', which usually is represented formally as follows:

$$(\textit{Theorem } 5) \quad \exists p(p \;\&\; \neg Kp) \vdash \exists p(p \;\&\; \neg\Diamond Kp),$$

where $p$ holds a place for sentence letters; $\Diamond$ is normal possibility, read 'it is possible that'; and $K$ is the epistemic operator, 'it is known (by someone [like us] by some means or other at some time) that'.

The theorem rests on tremendously modest modal epistemic principles, which we will turn to shortly. The converse of Theorem 5 is modest as well. So Theorem 5 does the interesting work in erasing the logical difference between there being truths forever unknown and there being truths logically unknowable.

The contrapositive of Theorem 5 is better known as the knowability paradox:

$$(\textit{Knowability Paradox}) \quad \forall p(p \rightarrow \Diamond Kp) \vdash \forall p(p \rightarrow Kp).$$

If each truth is knowable in principle, then it follows logically that each truth is at some time known. That's the result. It is thought to be paradoxical for a number of related reasons. First, it refutes all too easily interesting brands of anti-realism which are committed to the knowability principle, $\forall p(p \rightarrow \Diamond Kp)$. It refutes them since the knowability principle entails the obviously false omniscience principle, $\forall p(p \rightarrow Kp)$. The knowability principle has been claimed for a number of historic non-realisms, among them Michael Dummett's semantic anti-realism, Hilary Putnam's internal realism, the logical positivisms of the Berlin and Vienna Circles, Peirce's pragmatism, Kant's transcendental idealism, and Berkeley's metaphysical idealism. How strange that the knowability principle, and every brand of non-realism that avows it, are only as plausible as the exceedingly implausible, and obviously false, omniscience principle. An extension of Fitch's result, found in Williamson (1992: 68), shows that a traditional strengthening of the knowability principle forecloses on the very distinction between what is possible and what is actual. Roughly, if truth is possible knowledge then possibility is actuality.[1] The paradoxicality is that sophisticated forms of anti-realism could be so easily refuted.

A second reason to regard the proof as paradoxical is that it threatens to erase the logical distinction between the knowability principle and the omniscience principle. More specifically, the proof logically collapses the relatively moderate and plausible claim that each truth *can* be known into the apparently stronger and unbelievable claim that each truth is *in fact* known. The claims seem to carry distinct logical commitments, but they do not if Fitch's result is valid.

Fitch's result presupposes the following principles.

*Knowing a conjunction requires knowing each of the conjuncts:*
(A) $K(p \,\&\, q) \vdash Kp \,\&\, Kq$
*Knowing entails truth:*
(B) $Kp \vdash p$
*Theorems are necessarily true:*
(C) If $\vdash p$, then $\vdash \Box p$
And, *a necessarily false proposition is impossible:*
(D) $\Box\neg p \vdash \neg\Diamond p$

The proof may be characterized this way:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{K(p \,\&\, \neg Kp)}\,(1)}{Kp \,\&\, K\neg Kp}\,(\mathrm{A})}{Kp \,\&\, \neg Kp}\,(\mathrm{B})\ \&\ \text{trivial logic}}{\neg K(p \,\&\, \neg Kp)}\,(1)}{\Box\neg K(p \,\&\, \neg Kp)}\,(\mathrm{C})}{\neg\Diamond K(p \,\&\, \neg Kp)}\,(\mathrm{D})$$

At the top of the tree we suppose for *reductio* that the Fitch-conjunction, $p \,\&\, \neg Kp$, is known. By (A), it follows that each conjunct is known. The third line demonstrates an application of factivity, (B), to the right conjunct of the second line. In the face of the ensuing contradiction, we discharge and deny our only assumption. By necessitation, (C), and then by (D), we conclude with the impossibility of our initial assumption—giving, $\neg\Diamond K(p \,\&\, \neg Kp)$.

Now suppose the knowability principle, $\forall p(p \rightarrow \Diamond Kp)$, and take the following instance: $(p \,\&\, \neg Kp) \rightarrow \Diamond K(p \,\&\, \neg Kp)$. This together with the above

[1] More carefully, Williamson shows this: if necessarily something is true if and only if it is knowable, then necessarily p is possible if and only if p.

theorem, $\neg\Diamond K(p \,\&\, \neg Kp)$, entails $\neg(p \,\&\, \neg Kp)$, which may be generalized to $\forall p\neg(p \,\&\, \neg Kp)$. The classical equivalent is the omniscience principle, $\forall p(p \to Kp)$. At a glance:

$$\dfrac{\dfrac{}{\neg\Diamond K(p \,\&\, \neg Kp)} \qquad \dfrac{\forall p(p \to \Diamond Kp)}{(p \,\&\, \neg Kp) \to \Diamond K(p \,\&\, \neg Kp)}}{\dfrac{\neg(p \,\&\, \neg Kp)}{\dfrac{\forall p\neg(p \,\&\, \neg Kp)}{\forall p(p \to Kp)}}}$$

In sum, if all truths are knowable, then all truths are known:

$$\forall p(p \to \Diamond Kp) \vdash \forall p(p \to Kp).$$

## The Generalized Paradox

Fitch generalized the knowability result, showing that any operator $O$ that is both factive and closed under conjunction-elimination, generates the following aporia:

$$\begin{array}{c}\underbrace{\forall p(p \to \Diamond Op) \qquad \exists p(p \,\&\, \neg Op)} \\ \vdots \\ \bot\end{array}$$

To prove this Fitch begins with Theorem 1, which holds of any factive operator $O$ that is closed under conjunction-elimination:

$$\text{(Theorem 1)} \vdash \neg\Diamond O(p \,\&\, \neg Op).$$

Theorem 1 generates the above aporia.[2] Others have noted that it is not just factive, conjunction-distributive operators that validate Theorem 1 and generate the aporia. Belief, for instance, is closed under conjunction-elimination but is not factive. Yet arguably a belief-instance of Theorem 1 is provable, giving

$$\neg\Diamond B(p \,\&\, \neg Bp).^3$$

In this way the corresponding aporia is generated for the belief operator:

$$\begin{array}{c}\underbrace{\forall p(p \to \Diamond Bp) \quad \exists p(p \,\&\, \neg Bp)} \\ \vdots \\ \bot\end{array}$$

[2] To see how, substitute $p \,\&\, \neg Op$ for $p$ in $\forall p(p \to \Diamond Op)$. By Theorem 1, it follows that $\neg(p \,\&\, \neg Op)$. This in turn may be may be generalized, giving $\forall p\neg(p \,\&\, \neg Op)$, or equivalently $\neg\exists p(p \,\&\, \neg Op)$.

[3] See, for instance, Linsky (1986; and Ch. 11 of this volume).

That is, the plausible notion that any truth could be believed is inconsistent with the truism that some truths are not ever believed. Such proofs about belief avoid unrestricted factivity principles in favor of restricted principles about the transparency of beliefs about one's own beliefs. To take another example, a knowledge-version of the result may be derived without the conjunction-distributivity principle (Williamson 1993).

Most generally, then, a Fitch-aporia, or Fitch paradox, is generated for any operator $O$ just when

1. The conjunction $p \,\&\, \neg Op$ is un-$O$-able: $\forall p \neg \Diamond O(p \,\&\, \neg Op)$;
2. The $O$-ability principle, $\forall p(p \rightarrow \Diamond Op)$, is plausible; and
3. Clearly, some truths are un-$O$-ed: $\exists p(p \,\&\, \neg Op)$.

Operators that seem to generate Fitch-aporias include

It is written truthfully on the board that

Somebody brought it about that

God brought it about that

The laws of nature made it the case that

It is believed that

It is thought that

So, for instance, the paradox of omnipotence may be seen, logically, as a special case of Fitch's paradox. It says, roughly, that God can do anything that is in fact done, but only if God does in fact do everything. Another example: any truth can in principle be thought, but only if every truth is (at some time) thought. This latter result, like the result about belief, requires (in lieu of factivity) a principle that avows some minimal transparency of one's thoughts about one's thoughts.

## The Volume, Contributions and Literature

We here turn to some traditional and developing treatments of the paradox. The earliest version of the knowability proof appears in a 1945 referee report for the *Journal of Symbolic Logic* (printed here as Chapter 1). Its author, Alonzo Church, anonymously conveyed the proof to Fitch. The proof had the effect of undermining a certain definition of 'value' that Fitch was articulating—a definition that is trivialized if there are unknowable truths. So the proof originates in a context that is very different from the one in which we discuss the proof today. We think of the knowability paradox today either as an all-too-quick refutation of anti-realism or as a logical collapse of apparently distinct philosophical commitments. More on the more recent debate in a moment. Church offers a number of potentially promising ways to block the proof. He is most sympathetic to a rejection of closure principles for knowledge and belief, and a fortiori the principle that knowledge is closed under conjunction-elimination. This is principle (A) in our earlier presentation of the proof. So Church ultimately takes the knowability proof to be invalid—dare I say, paradoxical. However, Church's proposal does not help Fitch, since Fitch is deeply committed to necessary logical connections between the relevant propositional attitudes. Church considers that one may alternatively appeal to Russell's theory of logical types, which would have the effect of blocking special instances of the conjunctive distributivity principle—principle (A). The appeal to types foreshadows Linsky (Chapter 11 of this volume) and Hart (Chapter 19 of this volume). However, Church notes that the type-theoretic approach, like the rejection of closure principles, is antithetical to the goals of Fitch's manuscript.

Fitch had a very different kind of reply in mind. He responds to the referee report with a letter to the editor, in which he restricts the relevant class of true propositions to ones that it is 'empirically possible' to know. Fitch's definition of value is thus resuscitated. His restriction strategy foreshadows Neil Tennant (1997), where we find an analogous restriction to the class of true propositions that it is logically possible for somebody to know. We will discuss Tennant's restriction in a moment. It should be noted here that Church was unimpressed with Fitch's restriction strategy, and in a subsequent referee report (also in Chapter 1) attempted a version of the knowability proof that respects Fitch's restriction. The debate between Fitch and Church is tracked in Salerno (Chapter 3). Church's argument against Fitch's restriction strategy, I argue, is critically flawed.

Part I of the volume is dedicated to this, the early history of the Church–Fitch paradox of knowability. Chapter 1 is the pair of referee reports from 1945. They record one side of a dialog between Fitch and the referee regarding the paper submitted by Fitch to JSL. Chapter 2 is Fitch's seminal 1963 paper, shaped in no small part by those reports. Fitch's paper has been the logical fuel or foil for the literature on the knowability paradox. Essay 3 is my understanding of the first two essays. It offers an account of why Fitch included the knowability result in the 1963 paper.

Part II is about Michael Dummett's semantic anti-realism. The first wave of reactions to Fitch's 1963 paper, including Hart and McGinn (1976), Hart (1979), Mackie (1980), and Routley (1981), had a common theme. They all aimed to use Fitch's proof to discredit various forms of verificationism, the view that all meaningful statements (and so all truths) are knowable.[4] The knowability principle is commonly taken throughout the literature as a particularly clear expression of Hilary Putnam's internal realism (1981) and Michael Dummett's

[4] An exception is Walton (1976), whose aim was to draw lessons in the philosophy of religion. For related discussion, see Plantinga (1982); Humberstone (1985); MacIntosh (1991); Kvanvig (1995, and 2006); Rea (2000); Wright (2000); Cogburn (2004); Bigelow (2005); Brogaard and Salerno (2005); and Moretti (2008).

anti-realism (1959b; 1973, and elsewhere). The Fitch paper then threatens these forms of anti-realism. Since Williamson (1982) and Rasmussen and Ravinkilde (1982), however, we find various proposals to vindicate at least Dummettian anti-realism. Fitch's reasoning is classically, but not intuitionistically, valid. Specifically, the move from $\neg(p \ \& \ \neg Kp)$ to $p \rightarrow Kp$ (i.e., the final step in our version of the proof) is intuitionistically unacceptable, since it harbors an application of double-negation elimination—i.e., $\neg\neg p \vdash p$. Leading developments in Dummettian anti-realism favor intuitionistic revisions to classical logic.[5] As such, Dummettian anti-realism is said to evade the unwelcome classical consequences of Fitch. The proposal is further developed in Williamson (1988b; 1990; and 1992).

An objection to the intuitionistic strategy is found in the view that the intuitionistic consequences of Fitch's reasoning are as bad, or almost as bad, as the classical consequences. The objection is developed in Percival (1990).[6] The main intuitionistic consequence is $p \rightarrow \neg\neg Kp$, which says that no truths are forever unknown. Some equivalent formulas include $\neg(p \ \& \ \neg Kp)$, which denies that there are unknown truths, and $\neg Kp \rightarrow \neg p$, which says that anything forever unknown is false, and $\neg(\neg Kp \ \& \ \neg K\neg p)$, which denies that there are any forever undecided statements. The potentially irksome consequence, which is a focus of Wright (1993a: 426–7) and Williamson (1994a), can be put this way. The intuitionistic anti-realist lacks the resources to express the apparent truism that there may be truths that never in fact will be known, formally $\exists p(p \ \& \ \neg Kp)$. That is because the inconsistency derivable from the joint acceptance of the knowability principle and $\exists p(p \ \& \ \neg Kp)$ is intuitionistically acceptable.

In Chapter 4 Dummett embraces the intuitionistic consequences without regret. The paper defends $p \rightarrow \neg\neg Kp$ as the best expression of semantic anti-realism.[7] In a letter to the editor of this volume, Dummett explains that the intuitionistic anti-realist

> as I conceive of him or her, does not think it irksome that the notion 'never in fact' cannot be expressed by the use of the intuitionistic logical constants. Rather he or she thinks that the only meaning that can be given to 'never' is that expressible by the intuitionistic logical constants. So there is no worry and no frustration. (Letter: September 27, 2005)

For insightful discussion of the intuitionistic use of 'never', see Williamson (1994a).

Incidently, Dummett does not endorse the position articulated in his (2001), which proposes a restriction of the knowability principle to 'basic' or atomic

[5] For alternative formulations of the anti-realist argument against classical logic, see Tennant (2000); Salerno (2000); and Wright (2001).

[6] Important further discussion and a reply appears in DeVidi and Solomon (2001).

[7] Cf., DeVidi and Solomon (2001), which offers a defense of this very position on behalf of the Dummettian anti-realist. Dummett embraces the truth of $p \rightarrow \neg\neg Kp$ in much earlier work, including (1977: 339 [2000: 236]).

sentences. Dummett tells me that he wrote that paper to dispel the myth that Fitch's paradox is an objection to *any* form of anti-realism.

In Chapter 5 Stig Rasmussen further investigates and defends Dummett's newly favored knowability principle, $p \rightarrow \neg\neg Kp$. The centerpiece of the discussion is the 'mapping objection,' which points out that Gödel's 1933 mapping of intuitionistic logic into S4 fails to preserve the original formulation of the knowability principle, and that this fact counts against the original formulation as an expression of intuitionistic anti-realism.

In Chapter 6 José Bermúdez argues that the Dummett (2001) position is well-motivated. The position restricts the knowability principle to atomic statements, and defines intuitionistic truth inductively from there. Bermúdez offers an instructive account of Dummett's development in (1990) and (1996). There Dummett attempts to clarify the notion of indefinite extensibility of such concepts as set, natural number, and real number, and argues that only intuitionistic logic can illuminate a proper understanding of the notion. It is argued that if this is correct, then the Dummett (2001) theory of truth is well-motivated, and so, we have a principled solution to the knowability paradox.

Part III is dedicated to paraconsistency and paracompleteness. The paraconsistent approach to the paradox is first suggested in Richard Routley (1981). While considering the liar ('This very statement is not true'), the knower ('This very statement is not known') and Fitch's proposition, $\Diamond K(p \,\&\, \neg Kp)$, Routley entertains, but does not endorse, a uniform treatment:

> What the hardened paraconsistentist says is that [the liar] and $\Diamond K(p \,\&\, \neg Kp)$, though inconsistent, are nonetheless coherent, that this is how things are: some (but not too many) inconsistencies hold true. (1981: 112, n. 26)

Routley does not endorse the approach. His actual position is that Fitch's result is valid and that it indicates a *necessary* limitation of human knowledge. Fitch's result shows us that if there is in fact an unknown truth then there is a logically unknowable truth. On the assumption that our actual ignorance is a contingent matter, it is unclear whether the resulting unknowability is contingent or necessary. However, if *necessarily* we actually fail to know some truths, as Routley argues, then it follows by Fitch's main argument and the closure of necessity (over necessary implication) that, necessarily, some truths are unknowable. The passing insight about paraconsistency emerges in the context of Routley's more central discussion of the necessary limits of knowledge.

The paraconsistent approach is first defended in Beall (2000), where it is argued that the knower sentence provides independent evidence that knowledge is inconsistent. For the concept entails that $Kp \,\&\, \neg Kp$, for some $p$. Further, it is argued that without a solution to the knower we should accept contradictions of this form and go paraconsistent. To this end Wansing (2002) defines a paraconsistent, constructive, relevant, modal, epistemic logic that evades Fitch.

The section of this volume on paraconsistency constitutes the most recent developments of the paraconsistent treatment of the problem. In Chapter 7, Graham Priest develops the Routley/Beall proposal by countenancing the mere possibility of truth-value gluts and appealing to a paraconsistent logic with excluded middle. Beall, in Chapter 8, compliments the development by exploring alternatives, some of which avoid the epistemic oddities of Priest's framework. Beall's centerpiece is a semantic framework that is paracomplete, but not paraconsistent, and avoids a commitment even to the mere possibility of truth-value gluts.

Part IV is an exploration of temporal and epistemic analogs of Fitch's reasoning. The strategy is to translate the modalities in the knowability principle into a favored temporal or epistemic logic, and to draw lessons from there about the plausibility of the knowability principle and the result in which it figures. Johan van Benthem (Chapter 9) does this by placing the result in a dynamic epistemic setting—a setting in which the truth values of our epistemic attributions vary over time with the performance of various actions, such as announcements. The essay is a more thorough development of van Benthem (2004). John Burgess (Chapter 10) translates the Fitch modalities into various Priorian temporal modalities. Each of these two approaches offers, not a rejection of Fitch's proof, but an investigation of the problematic nature of the corresponding knowability principle.

Bernard Linsky (Chapter 11) proposes that we block Fitch's result by appealing to a theory of types in our account of epistemic and doxastic reasoning. Interestingly, this is one of the proposals that Alonzo Church (in Chapter 1) considers when entertaining objections to the knowability proof. Linsky shows that the theory of types systematically treats a wide variety of contemporary paradoxes of knowledge and belief.

Part V is dedicated to Neil Tennant's Cartesian restriction strategy. Tennant's position is that intuitionistic logic alone will not free anti-realism from the grips of Fitch. His well-discussed proposal is to restrict the knowability principle to *Cartesian propositions*, that is, propositions that it is not provably inconsistent to know. Objections to the proposal include Hand and Kvanvig (1999), Williamson (2000b), and DeVidi and Kenyon (2003). For replies see Tennant (2000a; and 2000b). Further motivation for Tennant's proposal can be found in Jon Cogburn (2004) and Igor Douven (2005).

In Chapter 12, Williamson continues the debate, specifically against Tennant (2001a), and renews his pessimism about the prospects for a successful defense of semantic anti-realism (Cf. Williamson 2000b). Debate with Williamson continues in Tennant (forthcoming). Chapter 13 is Kvanvig's renewed discontent with Tennant's (and any other) restriction to the knowability principle. As he sees it, the real paradoxicality is not that Fitch's result threatens anti-realism, but that it threatens to collapse the very distinction between the existence of unknown truth and the existence of unknowable truth. The section is completed by Chapter 14, which is Tennant's current position—a modification of the Cartesian restriction strategy.

Part VI is about modal and mathematical fictionalism. We learn in Brogaard (Chapter 15) that modal fictionalism is threatened by Fitch's paradox. Otávio Bueno (Chapter 16) evaluates the relevance of Fitch's paradox in the epistemology of mathematics. He argues that the mathematical fictionalist must contend with the unwelcome consequences of Fitch.

Part VII, Knowability Reconsidered, includes papers that reconsider the anti-realist thesis about the knowability of truth. There is a history of attempts to either refute or reformulate anti-realism in reaction of Fitch. I mentioned some refuters earlier. The reformulater is one who rejects, or offers an alternative to, the knowability principle as a characterization of anti-realism. They include Edgington (1985), Melia (1991), Wright (2000), Hand (2003), and Jenkins (2005), among many others. Michael Hand (Chapter 17) further develops his 2003 proposal that Dummett's anti-realist is not committed to the knowability principle, owing to the fact that it carelessly blurs semantic conditions about verification procedures with pragmatic conditions about the performance of such procedures. C. S. Jenkins (Chapter 18) agrees that the knowability principle fails as an expression of anti-realism. Her own statement of anti-realism (2005) is echoed here, but her primary concern is to take issue with Kvanvig (2006), in which it is argued that the real paradoxicality of Fitch's proof is the modal collapse that occurs in the reasoning from the knowability principle to the omniscience principle. W. D. Hart (Chapter 19) takes Fitch's proof to be evidence for realism. He argues that the prospects are not good for a solution coming from the theory of types. Christoph Kelp and Duncan Pritchard (Chapter 20) offer some hope for an anti-realism that endorses a justified believability principle in place of the knowability principle. They evaluate the thesis that, for all true propositions, it must be possible to justifiably believe them. An alternative weakening of the knowability principle is proposed by Greg Restall (Chapter 21). His principle states that, for every truth $p$, there is a collection of truths, such that (i) each of them is knowable and (ii) their conjunction is equivalent to $p$. Restall proves that this formulation evades the paradox, and draws lessons about the operant notion of possibility.

I regret that the volume is incomplete. It includes no extensive discussion of Dorothy Edgington's important 1985 proposal, in which the knowability principle is reformatted as a thesis about the knowability of *actual* truth. Important criticisms are found in Wright (1987a (2nd edn., 1993: 428–32)), Williamson (1987a; 1987b; 2000a, ch. 12), and Percival (1991). Developments of Edgington's proposal are found in Rabinowicz and Segerberg (1994), Linström (1997), Rückert (2004), Fara (forthcoming), and Murzi (manuscript). Related proposals, that focus on the modal semantics of Fitch's paradox, include Kvanvig (1995; 2006), Brogaard and Salerno (2006) and Costa-Leite (2006). These latter three approaches diagnose various modal fallacies. Kvanvig appeals to issues about when we are licensed to substitute into modal contexts. Brogaard and Salerno appeal to Stanley and Szabo's (2000) theory of quantifier domain restriction, according to which there is hidden structure in quantified noun phrases. Costa-Leite appeals to the fusion of Kripke frames—the insight being that knowability is not to be understood compositionally out of one-dimensional possibility and knowledge operators.

# 1

# Referee Reports on Fitch's "A Definition of Value"

*Alonzo Church*

The referee reports transcribed below were handwritten by Alonzo Church to his co-editor, Ernest Nagel, of the Journal of Symbolic Logic. They were issued in 1945 in response to a paper by Frederic Fitch, "A Definition of Value," which was not published. They contain the earliest formulations of the modal epistemic result today known as "Fitch's knowability paradox." The bracketed numerals, [n.], indicate the original page numbers. Our appendix is a list of the cited principles. They either originate in Lewis and Langford (1932) or are hypothesized by us to be the principles from Fitch's original manuscript, which was not found. The original reports are located in the Ernest Nagel Papers, Box 1, Arranged Correspondence, Church, Alonzo. Rare Book and Manuscript Library, Columbia University. They are printed here in full by their permission and by kind permission of Alonzo Church, Jr. We are grateful to Nick Zavediuk for his assistance in the transcription process.

## First Referee Report

[1.] It seems to me that the role and meaning of Professor Fitch's 'SC' is seriously in need of clarification. It is not sufficient merely to take 'SC' as primitive and undefined. It must be contemplated that ultimately there is either a definition of 'SC' or an elaborate set of empirical postulates about it; otherwise particular empirical necessitations such as "brakeless trains are dangerous" could not be decided.

Perhaps Fitch means to say that there is one absolutely determined set of "all the valid laws of empirical sciences," such that the currently accepted laws of the currently known empirical sciences are an approximation to a certain subset of this set, and that SC is something like the conjunction of all the laws of this set. But this belief in an ultimate set of absolutely valid empirical laws is held by hardly any contemporary empirical scientist. And I think that the recent history of the empirical sciences, especially physics, renders such a belief indefensible.

The only alternative I see is to take a particular formalized system of empirical science (say a system which unifies the empirical sciences as they are known today), and to take SC to be the conjunction of the primitive propositions of this system. This is acceptable, but it must be realized that it makes empirical necessitation relative to a particular system.

Accepting this emendation, I go on to Fitch's Def. 3.

[2.] Following the line of Fitch's thought, let me call a proposition empirically impossible if SC strictly implies its negation. (This makes empirical impossibility equivalent to the negation of empirical possibility and is therefore consistent with Fitch's Def. 6.) Then it may plausibly be maintained that if $a$ is not omniscient there is always a true proposition which it is empirically impossible for $a$ to know at time $t$. For let $k$ be a true proposition which is unknown to $a$ at time $t$, and let $k'$ be the proposition that $k$ is true but unknown to $a$ at time $t$. Then $k'$ is true. But it would seem that if $a$ knows $k'$ at time $t$, then $a$ must know $k$ at time $t$, and must also know that he does not know $k$ at time $t$. By Def. 2, this is a contradiction.

Now an empirically impossible proposition empirically necessitates every proposition. Therefore, the argument runs, by taking $q$ in Def. 3 to be $k'$, it may be inferred that everything is of value to $a$ at time $t$. Thus Def. 3 is reduced to a triviality.

In spite of the plausibility of the preceding argument I think Fitch has a good defense (but only one). This defense is that there is no law of psychology according to which one who believes a proposition must believe all its logical consequences; on the contrary, historical counter-examples are well known. To be sure, one who believes a proposition without believing its more obvious consequences is a fool; but it is an empirical fact that there are fools. It is even possible [3.] that there might be so great a fool as to believe the conjunction of two propositions without believing either of the two propositions; at least, an empirical law to the contrary would seem to be open to doubt. On this ground it is empirically possible that $a$ might believe $k'$ at time $t$ without believing $k$ at time $t$ (although $k'$ is a conjunction one of whose terms is $k$).

Unfortunately this defense compels Fitch to abandon his Ax. 1. And, what is more serious, it lights the way to a second and opposite objection to Def. 3.

If there is no empirical law according to which one who believes a proposition must believe its logical consequences, it would seem that by the same token there is no empirical law according to which a person's desires must be in reasonable accord with that person's beliefs. If someone desiring to recover from a certain disease, and knowing the one and only course of action which will lead to recovery, nevertheless does not desire that course of action, we may call that someone a fool; but again the fact is that fools there be. It is a historical fact that there have been persons who desired to avoid smallpox, who knew the medical efficacy of vaccination as a preventive, and who nevertheless violently resisted vaccination (therefore presumably did not desire it). I conclude [4.] that there is no valid law of psychology according to which anything whatsoever about my desires may be inferred from the fact that I know so-and-so. It follows by Def. 3 that nothing is of value to $a$ at time $t$, and again Def. 3 is reduced to a triviality. This is my second objection to Def. 3, and the one to which I attach the greater weight.

Since Def. 3 is the core of the paper, I hold that the entire paper is therefore untenable.

By this it is not to be understood that I disapprove of the idea of applying symbolic logic to value theory. However severe my criticism of Fitch for what I hold to be logical flaws, my criticism is still more severe of those philosophers who offer similar definitions of value in vague verbal form without the attempt at even so much accuracy as may be attained by the use of certain elementary notations of symbolic logic. For these latter escape the kind of criticism I level against Fitch only by making their statement so vague as to render all criticism uncertain. The very fact that an attempt by Fitch to state formally what I take to be a rather ordinary sort of definition of value leads to these logical difficulties is an indication of the need for at least some elementary symbolic logic here.

Let me say also that, in order to meet Fitch on his own ground, I have accepted uncritically what seems to be his notion of [5.] proposition, although it is well known that the notion of proposition is uncertain and in need of clarification. I am willing to concede, at least as a possibility, that one way to obtain clarification of the notion is to plunge directly into the use of propositions and to clear up individual difficulties as they arise.

Finally, I note that Fitch makes a medical error on page 4 of the manuscript, in implying that quinine is the only cure for malaria. An entirely different drug, atabrine, is as a matter of fact also an effective cure.

It seems to me that according to ordinary usage it would be said that quinine is valuable to the malaria sufferer, even if there does exist an alternative method of cure by means of atabrine. This observation may reveal another deficiency in Def. 3. But in view of more serious objections it seems unnecessary to go into this.

## Second Referee Report

1. It is not clear to me why Fitch thinks that, to quote his letter, "In order to show that $a$'s ignorance of $k'$ is empirically necessary, he would first have to show that $a$'s ignorance of $k$ is empirically necessary." The fact is that the quoted statement is false. To enforce my point, let me put the matter quite formally:

Assume: $k$.

Assume also: $\sim(a \text{ KN}t\ k)$.

Def.: $k' = (k \,\&\, \sim(a \text{ KN}t\ k))$.

By the foregoing Def., and Fitch's Th. 3: $(a \text{ KN}t\ k')$ EN $(a \text{ KN}t\ k)$.

By Fitch's Def. 2, and Lewis-Langford 11.2: $(a \text{ KN}t\ k')$ SI $k'$.

Hence by the Def. of $k'$, and Lewis-Langford 11.2, 11.6: $(a\ \mathrm{KN}t\ k')\ \mathrm{SI} \sim(a\ \mathrm{KN}t\ k)$.

Hence by Lewis-Langford 12.42: $(a\ \mathrm{KN}t\ k)\ \mathrm{SI} \sim(a\ \mathrm{KN}t\ k')$.

Hence by Fitch's Th. 1: $(a\ \mathrm{KN}t\ k)\ \mathrm{EN} \sim(a\ \mathrm{KN}t\ k')$.

The transitive law for EN follows from Lewis-Langford 15.1, 16.2, 11.6. Hence from the foregoing step and step 4 above we get: $(a\ \mathrm{KN}t\ k')\ \mathrm{EN} \sim(a\ \mathrm{KN}t\ k')$.

Hence by the law of double negation and Fitch's Def. 5: $\sim((a\ \mathrm{KN}t\ k')\ \mathrm{EC}\ (a\ \mathrm{KN}t\ k'))$.

Hence by Fitch's Def. 6: $\sim\mathrm{EP}(a\ \mathrm{KN}t\ k')$.

However, it follows from our assumptions: $k'$.

From this now it follows (for the reason explained in my previous report, and as I [2.] understand Fitch to admit) that Fitch's Def. 3 is untenable as it now stands, and must be altered if the paper as a whole is to be maintained.

Now let us consider the revised form of Def. 3 which Fitch proposed in his letter of February 26 and afterwards abandoned. As I understand it this is:

Def. 3R. $(a\ \mathrm{VL}t\ p) = (\mathrm{Eq})\ [q\ \&\ \mathrm{EP}(a\ \mathrm{KN}t\ q)\ \&\ [(a\ \mathrm{KN}t\ q)\ \mathrm{EN}\ (a\ \mathrm{DS}t\ p)]]$.

I think that a *reductio ad absurdum* of Def. 3R is possible along the same lines as that I have given for Def. 3. At least, let me attempt it, and leave it to Fitch to say where if anywhere I have assumed something he would not admit.

I shall show as a consequence of Def. 3R that instant death is of value to $a$ at time $t$. In other words, if $p'$ is "$a$ dies at time $t$," I shall show that $a\ \mathrm{VL}t\ p'$.

Assume that $a$ does not desire instant death at time $t$ (because in the contrary case, if we assume that it is empirically possible for one to know one's own desires, the conclusion $a\ \mathrm{VL}t\ p'$ is obvious). Nevertheless it is empirically possible for $a$ to desire instant death at time $t$, both (1) because it is empirically possible that $a$ should be insane, and (2) it is empirically possible that $a$'s external circumstances [3.] at time $t$ might have been so dreadful as to compel even a sane man to desire instant death.

Assume also that $a$ is not omniscient, and let $k$ be something which is true but unknown to $a$ at time $t$. Define $k'$ as before. Then as before $k'$ is true but $a$'s ignorance of $k'$ is empirically necessary.

Let $q'$ be the disjunction $(a\ \mathrm{DS}t\ p') \vee k'$. Then $q'$ is true.

I suppose Fitch would admit

$(a\ \mathrm{KN}t\ p)\ \mathrm{EN}\ (a\ \mathrm{KN}t\ (p \vee q))$.

At least this seems to be entirely in the spirit of his Th. 3, and it is hard to see how he could maintain one and deny the other. Also I suppose that logical consequences of the empirically possible are empirically possible, and that it is empirically possible for one to know one's own desires.

Thus because $a$ DS$t$ $p'$ is empirically possible, therefore $a$ KN$t$ ($a$ DS$t$ $p'$) is empirically possible, and therefore $a$ KN$t$ $q'$ is empirically possible.

Because ($a$ DS$t$ $p'$) EN ($a$ DS$t$ $p'$) and $k'$ EN ($a$ DS$t$ $p'$), therefore $q'$ EN ($a$ DS$t$ $p'$). Hence because ($a$ KN$t$ $q'$) EN $q'$, it follows that ($a$ KN$t$ $q'$) EN ($a$ DS$t$ $p'$).

Finally, taking, in Def. 3R, $p$ to be $p'$, and $q$ to be $q'$, we get the conclusion that $a$ VL$t$ $p'$.

The case of instant death is of course [4.] chosen only for illustration. In general, under Def. 3R, everything is of value to $a$ at time $t$ which it would be empirically possible for $a$ to desire at time $t$ under any empirically possible circumstances (however remote from the actual circumstances). This is little if any less disastrous than the situation under Def. 3, that everything whatever is of value to $a$ at time $t$.

Of course the foregoing refutation of Fitch's definition of value is strongly suggestive of the paradox of the liar and other epistemological paradoxes. It may therefore be that Fitch can meet this particular objection by incorporating into the system of his paper one of the standard devices for avoiding the epistemological paradoxes. If this is possible it will involve a drastic rewriting of the paper, not just a footnote here and there.

To my further objection—that there is no law of psychology according to which it can be inferred from the fact that $a$ knows something that therefore $a$ desires something—Fitch replies by pointing out that $a$ might know that $a$ desires $p$. If, however, Fitch consents to adopt one of the standard devices for avoiding the epistemological paradoxes, this reply will no longer be open to him. For example, on the basis of Russell's original [5.] theory of types, "$a$ desires $p$" is of higher order than $p$, whereas the two "something" 's in my assertion must of course be understood as of the same order. On the basis of Tarski's resolution of the epistemological paradoxes, the distinction between language and meta-language has roughly the same effect.

I insist therefore that there is no known law of psychology according to which "$a$ desires $p$" is ever a necessary consequence of "$a$ knows $q$." Moreover, in the light of every-day experience (summed up in the commonly heard conclusion, "Some people are utterly unreasonable"), it seems unlikely that there is a valid law of psychology of that sort remaining to be discovered.

If some of us think that there is a notion of value in spite of the fact that some people are utterly unreasonable, it is because we think we know how to distinguish between what is reasonable and what is unreasonable. The problem is whether this distinction between reasonable and unreasonable can be defined in non-valuational terms, or whether this or some like value-theoretic concept must be accepted as primitive (undefined). I do not think that Fitch has solved the problem.

The assumption that there is an absolute set of valid empirical laws, SC, to which the accepted laws of the empirical sciences are [6.] approximations in some sense, is of course a piece of metaphysics. I have no objection to metaphysics per se.

But this particular piece of metaphysics has more opponents than adherents, and it seems that a definition of value which presupposed it would be of interest only to a very restricted circle. I do not understand why Fitch objects to avoiding the metaphysical issue by making his definition of value relative to a particular (comprehensive) system of empirical science. But this is a side-issue, in view of the existence of more serious objections.

As to the matter of quinine and atabrine: it seems that, according to Fitch, if *a* is a malaria sufferer who has equal access to quinine and atabrine, then both "quinine is of value to *a*" and "atabrine is of value to *a*" are false.

Moreover it may be that there is some drug which is a quicker and more certain cure for malaria than either quinine or atabrine, which is easily accessible to *a* (if he only knew), but whose properties in this respect are still undiscovered. If so, then not even the disjunction, quinine or atabrine, is valuable to *a*.

It seems to me that such a notion of value departs so far from the everyday notion that it is hardly justified to use the same word for it.

Finally, Fitch's plan of adding [6.] short postscripts to his paper commenting on particular objections by the referee does not seem to me a good one. So far as the objections either are valid or represent misunderstandings likely to be duplicated by others, they should be met (if that is possible) by alterations in the body of the paper.

# Appendix

Joe Salerno and Julien Murzi

## Fitch's Operators

p SI q = p $\prec$ q = p strictly implies q

p EN q = p empirically necessitates q

EPp = p is empirically possible

p EC q = p is empirically consistent with q

*a*KN*t*p = *a* knows at time *t* that p

*a*B*t*p = *a* believes at *t* that p

*a*VL*t*p = *a* values at *t* that p

*a*DS*t*p = *a* desires at *t* that p

## Lewis and Langford (1932) Definition and Theorems

$(p \prec q) =_{df} \sim\Diamond(p \mathbin{\&} \sim q)$

11.2 $(p \mathbin{\&} q) \prec p$

11.6 $((p \prec q) \prec (q \prec r)) \prec (p \prec r)$

12.42 $(p \prec \sim q) \prec (q \prec \sim p)$

15.1 $((p \supset q) \,\&\, (q \supset r)) \prec (p \supset r)$

16.2 $((p \prec q) \,\&\, (p \prec r) \,\&\, T) \prec (p \prec (q \,\&\, r))$: $T = ((q \,\&\, r) \prec (r \,\&\, q))$

## Fitch's Definitions[1]

### Def. 2*

Def. 2 is Fitch's definition of knowledge. All we know from Church's use is that it justifies the principle that *a*'s knowing at time *t* that p strictly implies p: $a\mathrm{KN}t\mathrm{p} \prec \mathrm{p}$.[2]

### Def. 3*

$$a\mathrm{VL}t\ \mathrm{p} =_{\mathrm{df}} \exists \mathrm{q}(\mathrm{q} \,\&\, (a\mathrm{KN}t\mathrm{q}\ \mathrm{EN}\ a\mathrm{DS}t\mathrm{p})).$$

Value is what one would desire given sufficient knowledge: it is valuable to *a* at *t* that p if and only if there is a true proposition q, such that *a*'s knowing at *t* that q empirically necessitates *a*'s desiring at *t* that p.[3]

### Def. 3R

$$a\mathrm{VL}t\ \mathrm{p} =_{\mathrm{df}} \exists \mathrm{q}(\mathrm{q} \,\&\, \mathrm{EP}(a\mathrm{KN}t\mathrm{q}) \,\&\, (a\mathrm{KN}t\mathrm{q}\ \mathrm{EN}\ a\mathrm{DS}t\mathrm{p})).$$

Value is what one would desire given sufficient knowledge: it is valuable to *a* at *t* that p if and only if there is a truth q that it is empirically possible to know and *a*'s knowing at *t* that q empirically necessitates *a*'s desiring at *t* that p.[4]

### Def. 5*

$$(\mathrm{p\ EN} \sim \mathrm{p}) \prec \sim (\mathrm{p\ EC\ p}).$$

Necessarily, if p empirically necessitates ~p, then p is not (empirically) consistent with itself.[5]

### Def. 6*

$$\sim (\mathrm{p\ EC\ p}) =_{\mathrm{df}} \sim \mathrm{EPp}.$$

p is not (empirically) consistent with itself just in case p is not empirically possible.[6]

[1] An asterisk, '*', indicates that the principle does not appear explicitly in the reports, and therefore, that we have hypothesized its content.

[2] Church's applications appear in Report 1: 2 and Report 2: 1.

[3] Our formulation of Def. 3 is based on Church's trivialization argument against it. Compare Report 1: 2 and Report 2: 1–2.

[4] Report 2: 2.

[5] Report 2: 1.

[6] Report 1: 2, and Report 2: 1.

## Fitch's Axioms and Theorems

**Ax. 1***

$$(a\mathrm{B}t\mathrm{p}\ \ \&\ (\mathrm{p\ EN\ q})) \prec\ a\mathrm{B}t\mathrm{q}$$

Belief is closed under "empirically necessary" implication: necessarily, if *a* believes at *t* that p and p empirically necessitates q, then *a* believes at *t* that q.[7]

**Th. 1***

$$(\mathrm{p} \prec \mathrm{q}) \prec (\mathrm{p\ EN\ q})$$

Strict implication strictly implies empirical necessitation: necessarily, if p strictly implies q then p empirically necessitates q.[8]

**Th. 3***

$$a\mathrm{KN}t(\mathrm{p}\ \&\ \mathrm{q}) \prec (a\mathrm{KN}t\mathrm{p}\ \&\ a\mathrm{KN}t\mathrm{q})$$

Knowing a conjunction strictly implies knowing the conjuncts: necessarily, if *a* knows at *t* that both p and q, then *a* knows at *t* that p and *a* knows at *t* that q.[9]

[7] The discussion at Report 1: 2–3 suggests that Ax. 1 is this closure principle for belief. Alternatively, it is an unrestricted closure principle for knowledge (viz., knowledge is closed under necessary empirical implication).

[8] See for instance, Church's use in Report 2: 1.

[9] Report 2: 1.

# 2

# A Logical Analysis of Some Value Concepts[1]

*Frederic B. Fitch*

The purpose of this paper is to provide a partial logical analysis of a few concepts that may be classified as value concepts or as concepts that are closely related to value concepts. Among the concepts that will be considered are *striving for*, *doing*, *believing*, *knowing*, *desiring*, *ability to do*, *obligation to do*, and *value for*. Familiarity will be assumed with the concepts of logical necessity, logical possibility, and strict implication as formalized in standard systems of modal logic (such as S4), and with the concepts of obligation and permission as formalized in systems of deontic logic.[2] It will also be assumed that quantifiers over propositions have been included in extensions of these systems.[3]

There is no intention to provide exhaustive logical analyses, or to provide logical analyses that reflect in detail the usage of so-called ordinary language. This latter task seems impossible anyhow because of the ambiguities of ordinary language and the obvious inconsistencies and irregularities of usage in ordinary language. Furthermore, the term 'ordinary language' is itself rather vague. Whose ordinary language? Should English be preferred to Chinese? Various arguments that invoke English or Latin grammatical usage are seen to be without foundation from the standpoint of Chinese.

Just as the concepts of necessity and possibility used in so-called ordinary language correspond in some degree to the concepts of necessity and possibility

Chapter 2 was first published in the *Journal of Symbolic Logic* 28/2, 135–42 (1963) and is reproduced by permission of the Association for Symbolic Logic.

[1] An earlier draft of this paper was presented as a retiring presidential address to the Association for Symbolic Logic; read before the Association at Atlantic City, New Jersey, December 27, 1961.

[2] For example see A. R. Anderson, *The formal analysis of normative systems*, Technical Report No. 2, Contract No. SAR/Nonr-609(16), Office of Naval Research, Group Psychology Branch, 1956; also, by the same author, *A reduction of deontic logic to alethic modal logic*, ***Mind***, n.s. vol. 67 (1958), pp. 100–3.

[3] Such quantifiers can be introduced by methods analogous to those used in R. C. Barcan (Marcus), *A functional calculus of first order based on strict implication*, this Journal, vol. 11 (1946), pp. 1–16; *The deduction theorem in a functional calculus of first order based on strict implication*, ibid., pp. 115–18; and F. B. Fitch, *Intuitionistic modal logic with quantifiers*, *Portugaliae mathematica*, vol. 7 (1948), pp. 113–18. See also, R. Carnap, *Modalities and quantification*, this Journal, vol. 11 (1946), pp. 33–64.

used in modal logic, so too it is to be hoped that the ordinary language concepts of striving, doing, believing, desiring and knowing will correspond in some degree to the concepts that we will partially formalize here. Also, just as there are various slightly differing concepts of possibility and necessity corresponding to differing systems of modal logic, so too there are presumably various slightly differing concepts of striving, doing, believing, and knowing, having differing formalizations.

We begin by assuming that striving, doing, believing, and knowing all have at least some fairly simple properties which will be described in what follows, and we leave open the question as to what further properties they have.

First of all, we assume that striving, doing, believing, and knowing are two-termed relations between an agent and a possible state of affairs. It is convenient to treat these possible states of affairs as propositions, so if I say that *a* strives for *p*, where *p* is a proposition, I mean that *a* strives to bring about or realize the (possible) state of affairs expressed by the proposition *p*. Similarly, if I say that *a* does *p*, where *p* is a proposition, I mean that *a* brings about the (possible) state of affairs expressed by the proposition *p*. We do not even have to restrict ourselves to *possible* states of affairs, because impossible states of affairs can be expressed by propositions just as well as can possible states of affairs. In the case of believing and knowing, there is surely no serious difficulty in regarding propositions as the things believed and known. So we treat all these concepts as two-termed relations between an agent and a proposition. In a similar way, the concept of *proving* could also be regarded as a two-termed relation between an agent and a proposition.

For purposes of simplification, the element of time will be ignored in dealing with these various concepts. A more detailed treatment would require that time be taken seriously. One method would be to treat these concepts as a three-termed relation between an agent, a proposition, and a time. Another method would be to avoid specifying times explicitly, but rather to use a temporal ordering relation between states of affairs. This latter method might be more in keeping with the theory of relativity, in either its special or general form.

As a further step of simplification we will often ignore the agent and thus treat each of the concepts under consideration as a *class* of propositions rather than as a two-termed relation. For example, by ‘striving’ we will mean the class of propositions striven for (that is, striven to be realized), and by ‘believing’ we will mean the class of propositions believed, relativizing the whole treatment to some unspecified agent. But the agent can always be specified if we wish to do so, and we can replace classes by two-termed relations.

A class of propositions (in particular such classes of propositions as striving, knowing, etc.) will be said to be *closed with respect to conjunction elimination* if (necessarily) whenever the conjunction of two propositions is in the class so are the two propositions themselves. For example, the class of true propositions is closed with respect to conjunction elimination because (necessarily) if the conjunction of two propositions is true, so are the propositions themselves. If $\alpha$ is a class closed with respect to conjunction elimination, this fact about $\alpha$ can be expressed in logical symbolism by the formula,

$$(p)(q)[(\alpha[p \,\&\, q]) \prec [(\alpha p) \,\&\, (\alpha q)]],$$

where '$\prec$' stands for strict implication.

We assume that the following concepts, viewed as classes of propositions, are closed with respect to conjunction elimination:

striving (for),
doing,
believing,
knowing,
proving.

For example, in the case of believing we assume:

$$(p)(q)[(\text{believes}[p \,\&\, q]) \prec [(\text{believes}\, p) \,\&\, (\text{believes}\, q)]].$$

Here are some further concepts which are evidently closed with respect to conjunction elimination:

truth,
causal necessity (in the sense of Burks),[4]
causal possibility (in the sense of Burks),
(logical) necessity,
(logical) possibility,
obligation (deontic necessity),
permission (deontic possibility),
desire for.

A class of propositions will be said to be *closed with respect to conjunction introduction* if (necessarily) whenever two propositions are in the class, so is the conjunction of the two propositions. If $\alpha$ is a class closed with respect to conjunction introduction, this fact about $\alpha$ can be expressed in logical symbolism by the formula,

$$(p)(q)[[(\alpha p) \,\&\, (\alpha q)] \prec (\alpha[p \,\&\, q])].$$

Except for causal, logical, and deontic possibility, all the concepts so far regarded as closed with respect to conjunction elimination could perhaps also be regarded as closed with respect to conjunction introduction, or some varieties of them could. For present purposes, however, we do not need to commit ourselves on this matter except to say that truth and causal, logical, and deontic necessity are all indeed closed with respect to conjunction introduction.

A class of propositions will be said to be a *truth class* if (necessarily) every member of it is true. If $\alpha$ is a truth class, this fact about $\alpha$ can be expressed

[4] A. W. Burks, *The logic of causal propositions*, ***Mind***, n.s. vol. 60 (1951), pp. 363–82.

in logical symbolism by the formula, $(p)[(\alpha p) ⥽ p]$. The concepts truth, causal necessity, and logical necessity are clearly truth classes. It also seems reasonable to assume that doing, knowing, and proving are truth classes, and so we make this assumption. Thus, whatever is true or causally or logically necessary is true; and (as we assume) whatever is done, known, or proved is also true.

The following two theorems about truth classes will be applied to some of the above-mentioned truth classes in subsequent theorems.

Theorem 1. If $\alpha$ is a truth class which is closed with respect to conjunction elimination, then the proposition, $[p \ \&\ {\sim}(\alpha p)]$, which asserts that $p$ is true but not a member of $\alpha$ (where $p$ is any proposition), is itself necessarily not a member of $\alpha$.

*Proof.* Suppose, on the contrary, that $[p \ \&\ {\sim}(\alpha p)]$ is a member of $\alpha$; that is, suppose $(\alpha[p \ \&\ {\sim}(\alpha p)])$. Since $\alpha$ is closed with respect to conjunction elimination, the propositions $p$ and ${\sim}(\alpha p)$ must accordingly both be members of $\alpha$, so that the propositions $(\alpha p)$ and $(\alpha({\sim}\ (\alpha p)))$ must both be true. But from the fact that $\alpha$ is a truth class and has ${\sim}(\alpha p)$ as a member, we conclude that ${\sim}(\alpha p)$ is true, and this contradicts the result that $(\alpha p)$ is true. Thus from the assumption that $[p \ \&\ {\sim}(\alpha p)]$ is a member of $\alpha$ we have derived contradictory results. Hence that assumption is necessarily false.

Theorem 2. If $\alpha$ is a truth class which is closed with respect to conjunction elimination, and if $p$ is any true proposition which is not a member of $\alpha$, then the proposition, $[p \ \&\ {\sim}(\alpha p)]$, is a true proposition which is necessarily not a member of $\alpha$.

*Proof.* The proposition $[p \ \&\ {\sim}(\alpha p)]$ is clearly true, and by Theorem 1 it is necessarily not a member of $\alpha$.

Theorem 3. If an agent is all-powerful in the sense that for each situation that is the case, it is logically possible that that situation was brought about by that agent, then whatever is the case was brought about (done) by that agent.

*Proof.* Suppose that $p$ is the case but was not brought about by the agent in question. Then, since doing is a truth class closed with respect to conjunction elimination, we conclude from Theorem 2 that there is some actual situation which could not have been brought about by that agent, and hence that the agent is not all-powerful in the sense described.

Theorem 4. For each agent who is not omniscient, there is a true proposition which that agent cannot know.[5]

*Proof.* Suppose that $p$ is true but not known by the agent. Then, since knowing is a truth class closed with respect to conjunction elimination, we conclude from Theorem 2 that there is some true proposition which cannot be known by the agent.

[5] This theorem is essentially due to an anonymous referee of an earlier paper, in 1945, that I did not publish. This earlier paper contained some of the ideas of the present paper.

Theorem 5. If there is some true proposition which nobody knows (or has known or will know) to be true, then there is a true proposition which nobody can know to be true.

*Proof.* Similar to proof of Theorem 4.

Theorem 6. If there is some true proposition about proving that nobody has ever proved or ever will prove, then there is some true proposition about proving that nobody can prove.

*Proof.* Similar to the proof of Theorem 4, using the fact that if $p$ is a proposition about proving, so is $[p \,\&\, \sim(\alpha p)]$.

This same sort of argument also applies to the class of logically necessary propositions, since this is a truth class closed with respect to conjunction elimination. Thus by Theorem 1 we have the result that every proposition of the form $[p\& \sim \Box p]$ is necessarily not logically necessary, and hence necessarily possibly false, where '$\Box$' denotes logical necessity. In other words, the proposition $\Box \sim \Box[p \;\&\; \sim \Box p]$ is true for every proposition $p$.[6] In particular, if $p$ is a true proposition which is not necessarily true, then $[p \;\&\; \sim \Box p]$ is a true proposition which is necessarily possibly false.

I now wish to describe a relation of *causation*, or more accurately, *partial causation*, which will be used in giving a definition of doing in terms of striving and a definition of knowing in terms of believing, as well as some other definitions.

I will assume that partial causation, expressed by 'C', satisfies the following axiom schemata C1–C4:

C1. $[[p \text{ C } q] \;\&\; [q \text{ C } r]] \strictif [p \text{ C } r]$. (transitivity)
C2. $[p \;\&\; [p \text{ C } q]] \strictif q$. (detachment)
C3. $[p \;\&\; [[p \;\&\; q] \text{ C } r]] \strictif [q \text{ C } r]$. (strengthening)
C4. $[[p \text{ C } q] \;\&\; [p \text{ C } r]] \equiv [p \text{ C } [q \;\&\; r]]$. (distribution)

Here '$p \equiv q$' is defined as '$[p \strictif q] \;\&\; [q \strictif p]$'.

I will also employ an identity relation among propositions and will employ the following axiom schemata I1–I9 for this identity relation:[7]

I1. $[[p = q] \;\&\; (\ldots p \ldots)] \strictif (\ldots q \ldots)$.
I2. $p = p$.

[6] This result in slightly different form is to be found in the two papers by Anderson cited above. He uses it in constructing a model of deontic logic in alethic modal logic and attributes it to W. T. Parry, *Modalities in the survey system of strict implication*, this Journal, vol. 4 (1939), pp. 137–54, Theorem 22.8.

[7] It is interesting to observe that I2–I9 may be used to serve as postulates for an algebra like Boolean algebra but somewhat weaker, provided that the identity symbol is regarded as a symbol for equality in such an algebra and that (in place of I1) there are added postulates to the effect that equality is symmetrical and transitive, and that the negates, conjuncts, and disjuncts of equal elements of the algebra are equal. Also, there should be a postulate to the effect that there are at

I3. $p = \sim\sim p$.
I4. $p = [p \,\&\, p]$.
I5. $[p \,\&\, q] = [q \,\&\, p]$.
I6. $[p \,\&\, [q \,\&\, r]] = [[p \,\&\, q] \,\&\, r]$.
I7. $[p \,\&\, [q \vee r]] = [[p \,\&\, q] \vee [p \,\&\, r]]$.
I8. $p = [[p \,\&\, q] \vee p]$.
I9. $[\sim p \,\&\, \sim q] = \sim [p \vee q]$.

Notice that we do not have such theorems as $p = [p \,\&\, [q \vee \sim q]]$ and $p = [p \vee [q \,\&\, \sim q]]$.

Only a few of the axiom schemata listed above will be directly relevant in what follows. The ones most relevant are C2, C4, I1, and I6. The property expressed by C3 reflects the fact that C is only *partial* causation. If C were total causation, then C3 would clearly be unacceptable. It should also be remarked that C need not be regarded as restricted to relating states of affairs that have space-time location, but may relate any state of affairs (e.g., a mathematical truth) to other suitable states of affairs. Otherwise, the sort of knowledge defined below would be knowledge only of states of affairs that have space-time location.

Using the relation C, a definition of doing in terms of striving will now be given. It is perhaps best to regard this definition merely as an axiom schema that provides a necessary and sufficient condition for doing, and similarly in subsequent definitions. As before, reference to the agent and to time are omitted for simplicity.

D1. (does $p$) $\equiv \exists q$[(strives for $[p \,\&\, q]$) & [(strives for $[p \,\&\, q]$) C $p$]].

This means that an agent does $p$ if and only if there is some (possible or impossible) situation $q$ such that the agent strives for $p$ and $q$, and a result of this striving is that $p$ takes place. Using I1, I6, C4, and properties of existence quantification, it is easy to show that this definition gives the result that doing is closed with respect to conjunction elimination.

A definition of knowing in terms of believing is now given:

least two unequal elements of the algebra. Such an algebra provides an algebraic formulation for the Anderson–Belnap system of first degree entailments with quantifiers omitted (A. R. Anderson and N. D. Belnap, Jr., *First degree entailments*, Technical Report No. 10, ibid., 1961, since the assertion that $p$ entails $q$ can be defined as the assertion that $p$ equals the conjunction of $p$ with $q$, or equivalently as the assertion that $q$ equals the disjunction of $q$ with $p$. This algebra was suggested to me by a list of theorems on page 21 of my paper, *A system of combinatory logic*, Technical Report No. 9, ibid., 1960, and in part also by some discussions with Anderson. It also bears a close relation to the system of my paper, *The system C$\Delta$ of combinatory logic*, Technical Report No. 13, ibid., 1962 (also forthcoming in this JOURNAL). The system of first degree entailment including quantifiers was also arrived at independently by Miss Patricia A. James and myself as a modified form of the system of my book ***Symbolic logic*** (New York, 1952) prior to the Anderson–Belnap formulation of that system. This alternative approach to the system of first degree entailment is sketched on p. vii of Miss James's doctoral dissertation, *Decidability in the logic of subordinate proofs* (Yale University, 1962).

D2. $(\text{knows } p) \equiv \exists q[p \;\&\; q \;\&\; [[p \;\&\; q] \text{ C } (\text{believes } [p \;\&\; q])]]$.

This means that an agent will be said to know *p* provided that *p* and some (possibly other) situation *q* are both true, and provided that the fact that they are both true causes the agent to believe the fact that they are both true. Thus the known fact *p* must be causally efficacious (as part of the conjunction [*p* & *q*]) in bringing about the agent's belief that [*p* & *q*] is the case, and hence that *p* itself is the case, since belief is assumed closed with respect to conjunction elimination. It is easy to show that knowing, as thus defined, is a truth class closed with respect to conjunction elimination.

Ability to do can be defined in the following way:

D3. $(\text{can do } p) \equiv \exists q[(\text{strives for}[p \;\&\; q])\text{C}p]$.

This definition can be shown to give the result that ability to do is closed with respect to conjunction elimination.

Obligation to do can be defined in terms of doing and the concept of obligation as expressed by the operator '0' of deontic logic, as follows:

D4. $(\text{should do } p) \equiv 0(\text{does } p)$.

Obligation to do, as thus defined, can be shown to be closed with respect to conjunction elimination and also with respect to conjunction introduction.

I now wish to propose a definition of desire, as follows:

D5. $(\text{desires } p) \equiv \exists q[(\text{believes}(\text{can do}[p \;\&\; q]))\text{C}(\text{strives for}[p \;\&\; q])]$.

This means that an agent desires a situation *p* if his belief that he can achieve the conjunction of *p* with some (possibly other) situation causes him to strive for that conjunction of situations. Desire as thus defined can be shown to be closed with respect to conjunction elimination.

A concept of *value*, which I now wish to consider, can be defined in the following way:

D6. $(\text{value } p) \equiv \exists q \exists r[q \;\&[(\text{knows } q)\text{C } (\text{strives for } [p \;\& r])]]$.

This means that a situation *p* is a value for an agent if (and only if) there is an actual situation *q* and situation *r* such that if the agent knows *q* then he will strive for the conjunction of *p* and *r*. In knowing *q* the agent may be supposed to have all the knowable relevant information concerned with the effect of his striving for the conjunction of *p* and *r*, and if this knowledge causes him to strive for this conjunction, it must be because this conjunction, and in particular *p* itself, is of value to him. To see why *q* may be supposed to contain all the knowable relevant information for the purpose at hand, let us suppose, on the contrary, that *q* does not contain all such relevant information. Then there might be some additional information *s* such that knowledge of the conjunction of *q* and *s* would cause the agent not to strive for any conjunction of the form [*p* & *t*]. But in the hypothetical case that the agent knew [*q* & *s*], he would also know *q* because of the fact that knowing is closed with respect to conjunction elimination, and this knowledge of *q*, by assumption, would cause him to strive for [*p* & *r*]. Thus he would be caused to strive for [*p* & *r*] and also caused not to strive for [*p* & *r*], and the assumption that he could know such a proposition as [*q* & *s*] leads to an absurdity. Hence *q* may be regarded as containing all the knowable relevant information. It can be shown easily that value as thus defined is closed with respect to conjunction elimination.

The objection might be raised against the above definition of value that the agent must be assumed to be rational, since otherwise he might have all the relevant knowledge to enable him to make a choice in his own interest, and yet, being irrational, he would be caused by this knowledge to make some other choice and to strive for some outcome that would not be of value to him. One way, and perhaps the only way, to attempt to meet this objection is to maintain that all irrationality is due to lack of sufficient knowledge, so that the having of sufficient relevant knowledge already rules out any relevant amount of irrationality. According to this view, any sort of insanity would be curable simply by giving the patient sufficient knowledge of himself and of the world around him. This view would not deny that in practice there might be insuperable obstacles that prevent the communication of this knowledge to the patient, but the existence of such obstacles would not prove that irrationality was not essentially a lack of knowledge.

This definition of value of course does not guarantee that there are any values in this sense, though it seems to me not unreasonable to assume that there may be values in this sense.

A more difficult problem is the problem of the comparison of values, that is, the problem of greater and less among values. This problem will not be dealt with here.

YALE UNIVERSITY