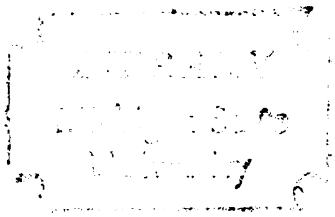
THESIS



This is to certify that the

dissertation entitled

Moore's Problem and the Prediction Paradox:
New Limits for Epistemology

presented by

Roy A. Sorensen

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Philosophy

Herbert E. Hendry
Major professor

Date 18 May 1982

O-12771

MOORE'S PROBLEM AND THE PREDICTION PARADOX:

NEW LIMITS FOR EPISTEMOLOGY

By

Roy A. Sorensen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Philosophy

1982

ABSTRACT

MOORE'S PROBLEM AND THE PREDICTION PARADOX:

NEW LIMITS FOR EPISTEMOLOGY

By

Roy A. Sorensen

Ludwig Wittgenstein once exclaimed that the most important philosophical discovery made by G. E. Moore was of the oddity of sentences like 'It is raining but I do not believe it'. This dissertation can be viewed as a partial vindication of Wittgenstein's enthusiasm.

However, my direct target is the prediction paradox. In the first chapter, the history of the prediction paradox is covered in detail. With the help of some new variations of the prediction paradox, I then argue in Chapte II that the paradox has not yet been solved. Chapter III contains my solution to Moore's problem. My concept of an epistemic blindspot emerges from this chapter and is used to establish new kinds of limits on knowledge in Chapter IV. In the following chapter I argue that the prediction paradox is a symptom of our unfamiliarity with these limits. Thus the prediction paradox is part of a general epistemological problem rather than an isolated logical problem. I try to make this claim more plausible in Chapter VI by applying the lessons learned about these new limits for epistemology to the more traditional philosophical problems associated

with predictive determinism. Along the way I show that disagreement amongst ideal thinkers is possible. I use this possibility to argue against emulation theories of moral problem solving, like ideal observer theories, conventionalism, and the Rawlsian appeal to the original position. I conclude the chapter by using epistemic blindspots as counterexamples to predictive determinism and retrodictive determinism. Having shown how pre-decisional blindspots have been illicitly employed to support the thesis that decisions are uncaused in Chapter VI, I argue in Chapter VII that post-decisional blindspots are involved in Newcomb's problem. A solution to this problem is then proposed. In my concluding remarks I provide a general characterization of my approach to the philosophical problems that have concerned me in this dissertation and a brief summation of its results.

For Julia Lynn Driver

ACKNOWLEDGMENTS

I thank the many people who have helped me become better at philosophy through their conversations with me. Of all these people, Herbert E. Hendry has helped me the most, and so it is to him that I am the most grateful.

TABLE OF CONTENTS

INTRODUCTION

The most popular variation of the prediction paradox involves a teacher who tells his students that there will be a suprise test next week. A clever student objects that the test is impossible. He first notes that the test cannot be given Friday since the students would then know on Thursday evening that that test must be on Friday. The test cannot be given on Thursday since the student would then know on Wednesday evening that the test is either on Thursday or Friday, and they have already eliminated Friday. In a like manner, the remaining days of the week are eliminated thereby "proving" that the test cannot be given.

The first two commentators on the prediction paradox agreed with the clever student and considered the paradox to be veridical. Following commentators were more sophisticated. Most have either thought that the clever student's argument contains an equivocation or have thought that the teacher's announcement is, contrary to appearances, self-referential. A recent few have thought that the prediction paradox shows that we must reject the principle that one knows only if one knows that one knows. Still others have tried to place the paradox in the same family as Moore's problem.

Moore's problem is the problem of explaining the oddity of sentences like 'It is raining but I do not believe it'. Since I agree with those who think that the prediction paradox is related to Moore's problem, I try to solve the latter in the hope of solving the former. My analysis of Moore's problem yields a definition of an

epistemic blindspot. Roughly, an epistemic blindspot is a consistent propostion which cannot be know by a certain people at certain times. We all have epistemic blindspots though they have been almost entirely unnoticed even by philosphers.

Epistemic blindspots are counterexamples to the principle that I can know whatever you can know and to the principle that if I can know something at a certain time, then I can know it at another time. Thus these blindspots show that there are unfamiliar limits to knowledge. I argue that the prediction paradox is a symptom of our unfamiliarity with these new limits for epistemology.

To further support my claim that our unfamiliarity with these limits are responsible for some philosophical problems, I try to show that much of the work done on the topic of predictive determinism is flawed by this unfamiliarity. One example of an epistemic blindspot is that a person cannot know what his decision is immediately before he makes the decision. This blindspot has been used to support the thesis that decisions are uncaused. Roughly, the argument is that if decisions are caused, then they are in principle predictable in which case it would be possible to know what one's decision will be immediately before making it. Since it would then be the case that one can know something that one cannot know, we must reject the supposition that decisions might be caused. Although I try to refute this argument, I do use blindspots as counterexamples to predictive determinism and for that matter, retrodictive determinism. In addition, I show how the sentences Moore was interested in suggest a way for ideal thinkers to disagree. The possibility of this kind of

disagreement undermines emulation theories of moral problem solving. emulation theory of moral problem solving is a theory which implies that there is an agent or group of agents such that for any moral question, one can correctly answer the question by agreeing with the answer of that agent or group of agents. Ideal observer theories, conventionalism, and Rawls' original position device are examples.

In addition to the pre-decisional blindspot mentioned above, there is an interesting post-decisional blindspot involved in Newcomb's problem. This problem involves a chooser and a predictor. The chooser is shown in two boxes. One box is transparent and contains one thousand dollars. The other box is opaque and contains one million dollars if and only if the predictor has predicted that the chooser will decide to take only the opaque box. Newcomb's problem is the problem of determining whether one should take only the opaque box or both boxes. I argue that once the role of this post-decisional blindspot is understood, Newcomb's problem is solved.

The general theme of this work is that several recent philosophical problems are due to our unfamiliarity with certain peculiar epistemological limits. As we gain familiarity with these limits, these problem are solved and we are given reason to hope that contributions to other philosophical problems can be made by further study of these limits.

CHAPTER I

HISTORY OF THE PREDICTION PARADOX

Although Quine reports that the prediction paradox had some currency from 1943 onward, it first appeared in philosophical literature in 1948 in D.J. O'Connor's "Pragmatic Paradoxes".

> The military commander of a certain camp announces on a Saturday evening that during the following week there will be a "Class A blackout". The date and time of the exercise are prescribed because a "Class A blackout" is defined in the announcement as an exercises which the participants cannot know is going to take place prior to 6:00 pm on the evening in which it occurs. It is easy to see that it follows from the announcement of this definition that the exercise cannot take place at all. It cannot take place on Saturday because if it has not occurred on one of the first six days of the week it must occur on the last. And the fact that the participants can know this violates the condition which defines it. Similarly, because it cannot take place on Friday last available day and is, therefore, invalidated for the same reason as Saturday. And by similar arguments, Thursday, Wednesday, etc., back to Sunday are eliminated in turn, so that the exercise cannot take place at all.[1]

O'Connor considers the argument cogent. He points out that the definition of a "Class A black-out" is consistent but goes on to claim that it is pragmatically self-refuting. He compares the definition to the following sentences:

(1) I remember nothing at all.

(2) I am not speaking now.

(3) I believe there are tigers in Mexico but there aren't any there at all.

Although (1)-(3) are consistent, they "could not conceivably be true in any circumstances".[2] Further, (1)-(3)

> are all statements in the first person which refer to the contemporary behaviour or state of mind of the speaker. In other words, they are all statements involving what Russell calls "egocentric particulars" and Reichenbach calls "token reflexive" words. That their peculiarities are closely connected with this can be seen from the fact that the peculiarties disappear if we substitute "you" or "he" for "I" or allow the statement to refer to past or future conditions of the speaker. But not all pragmatic paradoxes are of this kind,...[3]

In "Mr. O'Connor's 'Pragmatic Paradoxes'," L. Jonathan Cohen argues that pragmatic paradoxes are consistent propositions which are falsified by their own utterance. Public announcement of

(4) A "Class A blackout" will take place during the following week,

makes it false. In a footnote Cohen adds:

> If the camp commander intended to stage a suprise exercise on one day during the week and yet wanted to warn his troops of his intention, he would have to make an announcement somewhat like one or other of the following: Either "One day next week there will be a surprise exercise. A surprise exercise is an exercise about which, unless it takes place on the last day of the period for which you are warned, you will be in doubt as to when it is to happen until 6:00 pm on the evening in which it occurs" Or "One day next week there will be an exercise. Unless it take place on Saturday you will be in doubt as to when it is to happen until 6:00 pm on the evening in which it occurs." In the former case he utters a prediction and a definition, in the latter two

> predictions. Owing to the irreversibility of the time series, if it is known that an event will take place on either t1 or t2 or...tn-1.[4]

In "Pragmatic Paradoxes", Peter Alexander objects to Cohen's treatment of (1) and (4). (4) is not paradoxical at all since

> any announcement of an intention is implicitly recognised to be conditional on the possiblity of carrying out that intention. Even if I make a simple statement like "I will go to the cinema tomorrow" I mean, although I do not state, that I shall do so if I am not in any way prevented. Thus Professor O'Connor's statement, which can be abbreviated to read "A 'Class A Blackout' will be carried out next week" ought for completeness, to read "If the conditions of a 'Class A Blackout' can be realized, a 'Class A Blackout' will be carried out next week." Now this seems to raise no other difficulties than are raised by any conditional statement whose condition is unrealisable, like, for instance, "If I can live without air I will not breathe all day tomorrow but, similarly, men might cease next week to be able to realize that if the blackout had not occurred by Friday it must occur on Saturday, and then the condition would realizable. Any problems raised by these statements do not appear to be similar to those raised by the other statements with which I have dealt (1)-(3) nor to be properly called "paradoxical".[5]

The first publication devoted exclusively to the prediction paradox was Michael Scriven's "Paradoxical Announcements" in 1951. Whereas O'Connor regarded the paradox as rather frivolous and Alexander considered it interesting but of no great concern, Scriven is deeply impressed by the prediction paradox. Scriven puts (4) in the same class as (5) and (6).

(5) You are going to have a surprise at lunch-time tomorrow. You are are going to have steak and eggs.

(6) I'll wage you can't find the roots of the equation $x^2+5x-24=0$ within thirty seconds. The roots are 3 and -8.

Although the person who says (5) or says (6) does not contradict himself in the usual sense, his saying (5) or (6) is pointless since he has undermined part of what he says. Scriven goes on to insist that the unexpectedness of the exercise be given a logical rather than a psychological interpretation. The drill is unexpected by the participants in the sense that they cannot produce a proof that it will occur on a given day. Scriven argues that a solution to the paradox requires that one distinguish between publicly uttered statements and ordainments. Ordainments are guarantees, as when the dates of performances and meetings are announced. As a private prediction

(7) There will be a Class A Blackout next Saturday,

is proper, but it cannot be used as an ordainment for the drill participants. Construed as an ordainment, (7) guarantees a blackout which will on the one hand have an unspecified date, and on the other hand, have a specified date. This incompatibility forces one to conclude that either the blackout will occur on Saturday and not be Class A, or it will not occur Saturday and will be Class A. Neither conclusion is proper. We would only be led to these conclusions if we inferred from the self-refuting character of the announcement that there was a mistake. Since no proper conclusion can be drawn from (7)

as an ordainment, a Saturday blackout will be a Class A blackout, making (7) correct. Scriven next considers

(8) There will be a Class A blackout next week.

He claims that this announcement is also self-refuting since if the blackout does not occur before Saturday, it will be equivalent to (7) on Saturday morning.

> Saturday is therefore not a real possibility or else [(8)] is self-refuting. In general, a Class-A blackout cannot occur on the last day of any sequence of nights during which it is ordained or else the governing announcement will be self-refuting. The first five nights of the week now form such a sequence: at the next stage, the next four. An thus the nights of the reversed week fall one by one: falling with the last is the point of the ordainment.
>
> Now if the governing announcement is [(8)] which is self-refuting, and a blackout occurs on any night of the week, the statement [(8)] will be verified. And if publicly stated, it would still be correct.
>
> Conclusion. At first we thought that the reductive proof showed a Class-A blackout to be impossible while in fact any blackout that took place was a Class-A blackout. Now we have come to see that the suicide of the announcement as an ordainment is accompanied by its salvation as a statement.[6]

Scriven's proposal deviates sharply from the proposals of his predessesors. Whereas O'Connor and Cohen held that the paradox was veridical, Scriven classifies it as falsidical. In the next issue of Mind, O'Connor reported that he converted to Alexander's view . Since Alexander believed that the alleged paradox is dissolved by his paraphrase, O'Connor's conversion deepens his disagreement with Scriven. Scriven believes that there is a paradox.

Apparently in the hope of undermining Alexander's proposed dissolution, Paul Weiss reformulated O'Connor's paradox.

> A headmaster says, "it is an unbreakable rule in this school that there be an examination on an unexpected day." The students argue that the examination cannot be given on the last day of the school year, for if it had not been given until then, it could be given only on that day and would then no longer be unexpected. Nor, say they, can it be given on the next to the last day, for with the last day eliminated, the next to the last day will be the last, so that the previous argument holds, and so on and so on. Either the headmaster gives the examination on an expected day or he does not give it at all. In either case he will break an unbreakable rule; in either case he must fail to give an examination on an unexpected day.[7]

Weiss explains that O'Connor's formulation makes it possible for the announcement to be rescinded, so that the nonoccurrence of the blackout can be predicted. Weiss' stipulation that the rule is unbreakable corrects this flaw. In addition, Weiss believes that it is more appropriate to call the paradox "the prediction paradox". Since this name has a plurality of users, I have adopted it as well.

Weiss attempts to solve the paradox by assimilating it to the problem of logical fatalism. By the law of excluded middle, all proposition about the future, it is now either true or false. But then there is nothing one can do to change the truth values of these propositions. Thus the law of excluded middle seems to imply that we are not free. For example, tomorrow I will either eat cereal or not. But given that either 'I will eat cereal tomorrow' is true now or false now, there is nothing I can do to avoid eating cereal if it is

true now that I will and there is nothing I can do which will bring it about that I eat cereal tomorrow it is now false. According to Weiss and many others, Aristotle tried to avoid logical fatalism by denying that (9) implies (10).

(9) It is true that p or not-p.

(10) Either it is true that p or it is true that not-p.

According to this view, contingent propositions about the future lack truth values. Weiss then claims that the prediction paradox arises from confusing the collective and the distributive senses of 'or'. (9) is an example of the collective 'or' while (10) is an example of the distributive 'or'.

> When we predict we refer to a range of possibilities which are as yet undistinguished one from the other. They are connected by means of a collective "or", prohibiting the separation of any one of them from the others, without the introduction of some power or factor not included in the concept of the range. Since predictions always refer to a range and never to the specific determinations of it produced in fact, the predictions must be supplemented by history or the imagination if we are to select and eliminate first one and then another alternative. What is selected and eliminated in history or in the imagination will be something distinct, focused on, actualized, connected with others by means of a distributive "or". If we avoid confusing these two meanings of "or", our paradox, I think, will disappear.[8]

This distinction is obscure but the basic outline of Weiss' solution can be discerned. When we are asked to consider whether the examination could be given on the last day, we imagine ourselves in the future and thus shift from the realm of the possible to the realm of the actual. The disjunction of examination dates is distributive

meanings of "or", our paradox, I think, will disappear.[8]

This distinction is obscure but the basic outline of Weiss' solution can be discerned. When we are asked to consider whether the examination could be given on the last day, we imagine ourselves in the future and thus shift from the realm of the possible to the realm of the actual. The disjunction of examination dates is distributive in the realm of the actual but is collective in the realm of the possible. The shuttling back and forth in time invites confusion between realms and thus confusion between kinds of disjunctions.

Another popular version of the prediction paradox is the Hangman. A man is sentenced to hang on one of the following seven noons but must be kept in ignorance until the morning before the execution. The man argues that he cannot be hung on the last day since he would know after the penultimate noon. Having eliminated the last day, the rest are eliminated in the familiar way. In his "On a so-called Paradox", W. V. Quine blocks the elimination by showing that the announcement corresponding to the one-day case is not self-contradictory. Given that the judge says

(11) You will be hanged tomorrow noon and will not know the date in advance,

Quine claims that the man should reason as follows:

> "We must distinguish four cases: first, that I shall be hanged tomorrow noon and I know it now (but I do not); second, that I shall be unhanged tomorrow noon and know it now (but I do not): third, that I shall be unhanged tomorrow noon and do not know it now; and fourth, that I shall be hanged tomorrow noon and do not know it now. The latter two alternatives are the open possibilities, and the last of all would fulfill the decree.

> Rather than charging the judge with self-contradiction, therefore, let me suspend judgement and hope for the best.[9]

Since the base step of the induction is fallacious, Quine concludes that there is no paradox.

The first attempt to assimilate the prediction paradox to the self-referential paradoxes appeared in R. Shaw's "The Paradox of the Unexpected Examination". Shaw insists that "'knowing' that the examination will take place on the morrow" must be 'knowing' in the sense of "being able to predict, provided the rules of the school are not broken".[10] Shaw complains that "If instead one adopted a vague common-sense notion of 'knowing', then one could perhaps agree with Professor Quine that an unexpected examination could take place even in a one-day term; but to my mind, this would be evading the paradox rather than resolving it."[11] Given that 'unexpected' means 'not deducible from certain specified rules of the school', Shaw believes he can formulate two rules for the school described in Weiss' prediction paradox.

> Rule 1: An examination will take place on one day of next term.
> Rule 2: The examination will be unexpected, in the sense that it will take place on such a day that on the previous evening it will not be possible for the pupils to deduce from Rule 1 that the examination will take place on the morrow.[12]

Although a last day examination can be eliminated since it would violate Rule 2, an examination on any other day would satisfy Rules 1 and 2. By adding a third rule, the possibility of an examination on the last two days can be eliminated.

Rule 3: The examination will take place on such a day that on the previous evening it will not be possible for the pupils to deduce from Rules 1 and 2 that the examination will take place on the morrow.[13]

If only two days remain in the term, the pupils can deduce by Rule 1 that the examination is on one of the two remaining days. By Rule 2, they can eliminate the last day, leaving the next to the last day as the only possibility. Since this deduction would violate Rule 3, the last two days are not possible examination days. However, an examination on any other day of the term would satisfy Rules 1, 2, and 3. In general, the last n days of the term are eliminated by appealing to Rule 1 and n additional rules of the form

Rule n + 1: The examination will take place on such a day that on the previous evening it will not be possible for the pupils to deduce from the conjunction of rules 1, 2, . . ., n, that the examination will take place on the morrow.

The n + 1 rules are incompatible with an n + 1 day term.

Shaw concludes that original paradox arose by taking in addition to Rule 1,

Rule 2*: The examination will take place on such a day that on the previous evening the pupils will not be able to deduce from Rules 1 and 2* that the examination will take place on the morrow.[14]

By applying rules 1 and 2*, one can eliminate every day of the term. Once we realize that 2* is self-referential, the paradox is resolved.

Ardon Lyon complains that Shaw's choice of the rules for the school is an evasion rather than solution of the paradox. Lyon points out that mere self-referentiality is not sufficient for paradox. For

example, 'This sentence is written in black ink' is perfectly all right. Lyon reject's Quine's analysis on the grounds that Quine's criterion for knowing implies that we cannot know anything about the future. According to Lyon, the paradox rests on an equivocation. Shaw's Rule 2* can mean either S1 or S2, but not both.

> S1 The examination will be unexpected in the sense that . . . it will not be possible for the pupils to deduce from Rules 1 and S1 that the examination will take place on the morrow, unless it takes place on the last day.
>
> S2 The examination will be unexpected in the sense that . . . it will not be possible for the pupils to deduce from Rules 1 and S2 that the examination will take place on the last day.[15]

Lyon argues that if one reads Rule 2* as S1, like a sensible person should, then the clever student's argument is fallacious. And even if one reads it as S2

> . . . it can have no possible application, must always remain false, for nothing, including setting the examination earlier, would make it true that the boys would be unable to deduce on the eve of the last day that it would occur on the morrow, <u>if</u> the master were to wait that long. For R1 and S2 applied together on the eve of the last day give us:
>
> (1) The examination must take place tomorrow.
>
> (2) (The examination will be unexpected in the sense that) it is not possible to deduce from (1) and (2) that it will take place on the morrow.
>
> clearly contradict each other, as opposed to Quine's solution.[16]

Shaw concludes that the paradox arises from taking Rule 2* to mean S1 and S2 at the same time.

In 1960, David Kaplan and Richard Montague published "A Paradox Regained" in the Notre Dame Journal of Formal Logic, the first publication on the paradox to appear outside of Mind. They begin their rigorous development of Shaw's self-referential approach by letting M, T, and W respectively stand for 'K is hanged on Monday', 'K is hanged on Tuesday', and 'K is hanged on Wednesday'. '$K_s(x)$' stands for 'K knows on Sunday afternoon that sentence x is true'. 'K_m', 'K_t', and 'K_w' are treated analogously. The variable 'x' takes names of sentences as substituends. So Kaplan and Montague introduce a system of names of expressions. If $\overline{E}$ is any expression, then $\overline{E}$ is the standard name of E, constructed according to one of various alternative conventions. They suggest that one might either construe $\overline{E}$ as the result of enclosing E in quotes, or identifying $\overline{E}$ with the numeral corresponding to the Godel number of E, or regarding $\overline{E}$ as a structural-descriptive name of E.

Kaplan and Montague are now in a position to express the judge's decree, D1:

$$M \,\&\, \neg T \,\&\, \neg W \,\&\, \neg K_s(\overline{M}) \;\vee$$
$$\neg M \,\&\, T \,\&\, \neg W \,\&\, \neg K_m(\overline{T}) \;\vee$$
$$\neg M \,\&\, \neg T \,\&\, W \,\&\, \neg K_t(\overline{W})$$

They use the sentence '$I(S_1, S_2)$' to indicate that S_1 logically implies S_2. Kaplan and Montague then express the principles K appeals to for the impossibility of D_1, as:

(A_1) $(\neg M \,\&\, \neg T) \rightarrow K_t\ (\overline{\neg M \,\&\, \neg T})$

(A_2) $[I(\overline{\neg M \,\&\, \neg T}, \overline{W}) \,\&\, K_t(\overline{\neg M \,\&\, \neg T})] \rightarrow K_t(\overline{W})$

They use the sentence '$I(\overline{S}_1, \overline{S}_2)$' to indicate that S_1 logically implies S_2. Kaplan and Montague then express the principles K appeals to for the impossibility of D_1, as:

(A_1) $(-M \ \& \ -T) \rightarrow K_t \ (\overline{-M \ \& \ -T})$

(A_2) $[I(\overline{-M \ \& \ -T}, \ \overline{W}) \ \& \ K_t(\overline{-M \ \& \ -T})] \rightarrow K_t(\overline{W})$

(A_1) and (A_2) are special cases of the principles of knowledge by memory and the deductive closure of knowledge, respectively. Although dubious in full generality, "we can hardly deny K the cases embodied in (A_1) and (A_2), especially after he has gone through the reasoning above."[17] We can also assume that K knows (A_1) and (A_2).

(A_3) $K_m(\overline{A_1 \ \& \ A_2})$

Since K assumes that (A_1) and (A_2) logically imply $-W$, he tries to argue that he cannot be hanged on Wednesday noon.

(A_4) $[I(\overline{A_1 \ \& \ A_2}, \ \overline{-W}) \ \& \ K_m(\overline{A_1 \ \& \ A_2})] \rightarrow K_m(\overline{-W})$

To exclude Tuesday, K uses the following analogues of (A_1) and (A_2):

(A_5) $-M \rightarrow K_m \ (\overline{-M})$,

(A_6) $[I(\overline{-M \ \& \ -W}), \ \overline{T}) \ \& \ K_m(\overline{-M}) \ \& \ K_m(\overline{-W})] \rightarrow \ K_m(\overline{T})$

Additional analogues to (A_1) and (A_2) are used to eliminate Monday, and thus to show that D_1 cannot be fulfilled.

Kaplan and Montague note that K has committed the fallacy Quine pointed out when applying (A_2). (A_2) only implies $\overline{W}$ when conjoined with D_1, so (A_2) must be replaced by

$(A_{2'})$ $[I(\overline{-M \ \& \ -T \ \& \ D_1}, \ \overline{W}) \ \& \ K_t(\overline{-M \ \& \ -T}) \ \& \ K_t(\overline{D_1})] \rightarrow K_t(\overline{W})$

Thus we need to also assume $K_t(\overline{D_1})$. But this seems unreasonable, especially in light of K's attempt to prove that the decree will not be fulfilled.

However, Kaplan and Montague argue that Quine's formulation fails to capture the self-referential aspect of the decree. For the sake of brevity, Kaplan and Montague use the two day version of the hangman paradox to show how the self-referential aspect can be expressed.

(i) $D_3 \equiv [[M \ \& \ -T \ \& \ -K_s(\overline{D_3 \rightarrow M})] \ v \ [(-M \ \& \ T) \ \& \ -K_m(\overline{D_3 \rightarrow T})]]$

K excludes Tuesday and then Monday be appealing to the following analogues of (A1)-(A4):

(B1) $-M \rightarrow K_m(\overline{-M})$

(B2) $-[[I(-\overline{M}, \overline{D_3 \ T}) \ \& \ K_m(\overline{-M})] \rightarrow K_m(\overline{D_3 \ T})$

(B3) $K_s(\overline{B_1 \ \& \ B_2})$

(B4) $[I(\overline{B_1 \ \& \ B_2}, \overline{D_3 \rightarrow M}) \ \& \ K_s(\overline{B_1 \ \& \ B_2})] \rightarrow K_s(\overline{D_3 \rightarrow M})$,

According to Shaw, the decree is genuinely paradoxical, not merely incapable of fulfillment. However, Kaplan and Montague argue that the decree is merely incapable of fulfillment since the supposition that D_3 can be fulfilled leads to absurdity.

> Suppose as before that K is hanged on Tuesday noon and only then. In this possible state of affairs, -M and T are ture. The hangman must now establish $-K_m(\overline{D_3 \rightarrow T})$. To apply his earlier line of reasoning, he must show that $D_3 \rightarrow T$, considered on Monday afternoon, is a non-analytic sentence about the future. But $D_3 \rightarrow T$ is in fact analytic, for as K has shown, $-D_3$ follows logically from general epistemological principles, and hence so does $D_3 \rightarrow T$.[18]

A paradoxical decree would result if the judge tried to make the decree capable of fulfillment by adding a stipulation:

> Unless K knows on Sunday afternoon that the present decree is false, one of the following conditions will be fulfilled: (1) K is hanged on Monday noon' is true, or (2) K is hanged on Tuesday noon but not on Monday noon, and on Monday afternoon K does not know on the basis of the present decree that 'K is hanged on Tuesday noon' is true.[19]

Kaplan and Montague are able to show that this version is a complicated variation of the Liar paradox leading to the conclusion that the decree can and cannot be fulfilled. They go on to consider a one-day version of this variation:

> Unless K knows on Sunday afternoon that the present decree is false, the following condition will be fulfilled: K will be hanged on Monday noon, but on Sunday afternoon he will not know on the basis of the present decree that he will be hanged on Monday afternoon.[20]

Finally, they consider a version in which "the number of possible dates of execution can be reduced to zero". Here the judge asserts:

> K knows on Sunday afternoon that the present decree is false.[21]

Another branch of the self-referential approach was introduced by G. C. Nerlich in his "Unexpected Examinations and Unprovable Statements". After expressing his view that the prediction paradox is neither trivial nor easy to solve, Nerlich suggests that

> . . . it is a quite unique kind of ordinary language problem, having some connection with the situation posed by Goedel's famous sentence, to the effect that the sentence itself cannot be proved.

> It will be clear, when I have dealt with the paradox, why I think it is of some importance to logic—of more importance than the comparatively simple Grelling paradox, for example.[22]

Nerlich reviews Shaw's treatment of the paradox. Shaw provided a non-self-referential formulation of the school rules and a self-referential formulation. He then argued that the first formulation is not paradoxical since no unexpected examination can be given during the term and that the second formulation is paradoxical. Nerlich insists that both formulations are paradoxical. After all, if an examination is given Wednesday, it would not be expected. Thus Shaw's first formulation shows that self-reference is not an essential feature of the prediction paradox.

Nerlich next considers Lyon's claim that the paradox rests on an equivocation. Lyon argued that the sensible interpretation of the announcement is (a) rather than (b):

(a) it will not be possible to deduce from the statement when the examintion will occur at any time prior to its occurrence, unless it occurs on the last day.

(b) it will not . . . , whether or not it occurs on the last day.

Nerlich objects that the announcement cannot mean (a) since there is a perfectly proper and strict sense of 'unexpected' in which (a) is equivalent to the 'the examination will occur unexpectedly, unless it occurs expectedly on the last day.' Since the announcer can plainly mean strictly what he said, that the examination will be unexpected, the equivalence of (a) and the above ensures that the announcer does not mean (a).

On the other hand, denying that the announcement means (a) is not tantamount to asserting that it means (b). One is only denying that the examination will occur on any day such that on a previous day, the examination date could be deduced. Nerlich further argues that (a) is not equivalent to the announcement because

> . . . there are tests which actually require the rejection of the "unless" clause and such tests occur daily. The trial emergency stop in every driving test is a case in point. The trial is improper if the order does not take the candidate unawares, so it cannot be allowed to occur expectedly even at the end of the test. Yet proposing such a trial is not proposing anything contradictory.[23]

Nerlich admits that his own solution is "rather bizarre". He first points out that at each stage of the student's argument a negation of a statement of the form 'Examination on -day' is derived. But after deriving a negation for each of the alternatives, the students have no basis for thinking one day rather than another is the examination date. So if the examination is given on one of the days, it will be unexpected. To falsify the announcement, the students must derive a statement which excludes

> . . . "a day such that it is not possible to deduce from the head's statement, at any time prior to the day, that the examination has been arranged for that day.[24]

Since only negations are derived, the announcement is not falsified. The possibility of deriving an examination date from a contradiction should be ignored since it would be of no use to the pupils.

> So due to the fact that it entails not, e.g., Examination on Wednesday, but something else (a contradiction), the statement is self-consistent.
>
> This is a hard saying. However, let us look again at the curious logical features of this everyday remark. The statement is partly about an examination and partly about its own logical consequences, viz. that the examination date is not among them . . . . The only way in which this metalogical statement can be falsified is by proving that the examination has been arranged for a certain day. It is this that the students attempt to do but fail to do, producing only days on which it seems not possible to hold it. And that is because in the attempt, they are forced to use the very premise (or set of premises) which they hope to falsify.[25]

Nerlich admits that this alone is insufficient to account for the odd state of affairs since reductio ad absurdum arguments also use the premises the arguer hopes to falsify. He claims that the oddity is due to the fact that the key premise states that it cannot be used that way, for it says that only false statements can be deduced. Nerlich goes on to further claim that in so far as it is about provability, the prediction paradox resembles Goedel's incompleteness proof. Central to the proof is a sentence, G, which is true only if G is not provable. If the logical system is consistent, then G must be undecidable. For if G is proved, then G is also unproved, and if the negation of G is proved, then G is proved. So here consistency is incompatible with completeness. Nerlich claims that the same holds true for the announcement in the prediction paradox. By implying that

there is a true but unprovable alternative, the announcement is, as it were, describing itself as incomplete.

> But just that remark about incompleteness seems to make the system now complete, and therefor contradictory. Yet, as we have seen, it is really neither complete nor inconsistent.[26]

Nerlich concludes that when one's sole source of information seems to impeach himself, one does not know what to make of it. This is just what the teacher wants. He manages to say nothing by contradicting himself.

In The British Journal for the Philosophy of Science, Martin Gardner compared the prediction paradox to Langford's Visiting Card Paradox. Langford's paradox consists of a visiting card on the front of which is written 'The assertion on the other side of this card is true' while on the back is written 'The assertion written on the other side of this card is false'. To show the analogy between the prediction paradox and the Langford paradox, Gardner constructs a "New Prediction Paradox". Here, one puts a card in an envelope and instructs the receiver to send it to a mutual friend only after writing on its (as yet blank) back 'Yes' or 'No' according to whether the receiver feels justified in predicting that the mutual friend will find that 'No' has been written on its back. In "A Comment on the New Prediction Paradox", Karl Popper agrees that Gardner has established a close analogy between the two paradoxes. As a friendly amendment, however, Popper argues that Gardner's paradox can be formulated in such a way that it is free of the idea of negation (common to the Liar

and Langford paradoxes). Here, one instructs the receiver to write 'Yes' in a blank rectangle to the left of one's signature if, and only if, the receiver feels justified in predicting that when it is sent back, the rectangle will still be blank.

In the first issue of the American Philosophical Quarterly, Brian Medlin first expresses disappointment with all of the previous contributions to the problem except Shaw's. Nerlich is first criticized for offering a solution which merely reformulates the paradox. Medlin then moves on to formalize the paradox. Although he never mentions Montague and Kaplan, his approach and major results duplicate their work. However, Medlin does defend the stronger thesis that the prediction paradox rather than an offshoot of it is a paradox of self-reference. I will return to Medlin shortly when I describe Jonathan Bennett's criticisms of the self-referential approach.

In the next issue of American Philosophical Quarterly, Frederic Fitch's "A Goedelized Formulation of the Prediction Paradox" appeared. Fitch first argues that the announcement is merely self-contradictory. He then modifies the prediction paradox by weakening the notion of surprise so that an expected last day examination counts as a surprise examination. Fitch shows that this prediction is consistent and considers it a resolution of the paradox. Third, Fitch develops Nerlich's suggestion by modifying the prediction in the prediction paradox so that it is an undecidable proposition equivalent to Goedel's.

The first general criticism of the self-referential approach appeared in Jonathan Bennett's review of the articles written by Shaw, Lyon, Nerlich, Medlin, and Fitch. Bennett's first criticism of the attempt to solve the prediction paradox by showing that it has an element of self-referentiality is that Nerlich's objections have not been satisfactorily answered. Nerlich first argued that Shaw illegitimately assumed that all self-reference is improper. Medlin conceded that some cases of self-reference may be proper, citing R. M. Smullyan's "Languages in which Self-Reference is Possible" (The Journal of Symbolic Logic, 1957), but denies that self-reference is proper in the case in question. Medlin formulates the announcement as:

> (M) The information concerning dx [the day on which the examination occurs] is not sufficient to allow determination of x at any stage before the examination is actually given.[27]

The impropriety of (M) is then argued for on the grounds that

> The proposition (M) says something about the propostions in a non-empty set S; namely, that the conjunction of all these propositions does not constitute a premiss of sufficient power to permit the determination of x at any stage before the examination is given. . . . But if (M) is in S, then what (M) says is (roughly) that (M) does not permit us to determine x. This kind of self-reference is circular. It invites us the question, "What does not permit us to determine x?" We do not understand (M) until we know what (M) is about, which set S happens to be. If (M) is itself in S, then we shall never know this and never understand (M).[28]

Bennett objects to this argument since if it is valid, one could prove that 'No universal proposition entails that all men are mortal' is unintelligible. Nerlich's second objection was that self-reference is not essential to the paradox. Medlin formulates Nerlich's objection with the help of the following (using '—' standing above symbol for propositional negation, and letting p_i be the propostion 'The examination occurs on the i^{th} day').

(I) $(p_1 \vee p_2 \vee p_3) \ \& \ p_i \ \& \ p_j \ (i \neq j;\ i \leq j,\ j \leq 3)$

(M_1) From (I) it is not possible to determine x, even given as additional information one of $\overline{p_1}$, $\overline{p_1 \vee p_2}$.

(M_2) From (I) & (M_1) it is not possible to determine x, even given as additional information $\overline{p_1}$

(M_3) From (I) & (M_1) & (M_2) it is not possible to determine x.

(C) (M_1) & (M_2) & (M_3).

Nerlich argues that self-reference is not essential to the prediction paradox because (I) & (C) imply a contradiction by steps parallel to the self-referential cases. In reply, Medlin argues that Nerlich's own proposed solution keeps the paradox alive with the help of self-reference. Medlin explains that Nerlich argues in favor of the compatibility between (I) & (C) and p_2 on the grounds that there is no sound deduction from the former to the latter.

> But if this is to be taken as providing a model for (I) & (C), then we must interpret (C) as saying of <u>itself</u> that it does not, with (I), constitute sufficient information for the determination of x. The statement for which p_2 does provide a model is

(M_4) The conjunction (I) &(C) does not constitute sufficient information for the determination of x.

Unlike (C), the statement (M_4) is true. It is true because (C) is false. Nerlich confuses (M_4) with (C). He is then led to say that (C) is true because it is false. We should note in passing that the case p_1 provides a model for (M_4). So does the case p_3: that is why Nerlich finds that even an examination on d_3 is unexpected.[29]

Despite Medlin's report that Nerlich agrees with all of Medlin's comments about Nerlich's analysis, Bennett dismisses Medlin's attempt to meet Nerlich's objection as ad hominem. Even if the above criticism of Nerlich's constructive analysis succeeds, it does not show that Nerlich's destructive analysis fails. After noting the common diagnosis that the Lyon ambiguity in the announcement is the source of our puzzlement, Bennett concludes:

> Perhaps there is that ambiguity and perhaps it might puzzle someone; but it has nothing to do with the fact which makes the announcement teasing to everyone, namely the fact -- noted by Fitch on page 161 -- that "in practice the event may nevertheless occur on some one of the specified set of days, and when it does occur it does constitute a sort of surprise." But <u>that</u> puzzle cannot be handled by someone who thinks that "the Prediction Paradox can be formulated in a . . . way that makes no use of epistemological or pragmatic concepts" (p. 161).[30]

Bennett's review is followed by James Cargile's review of Kaplan and Montague, Gardner, and Popper. Cargile dismisses Gardner's "new prediction paradox" as not being a genuine paradox. Although Langford presents some similar paradoxes, the visiting card paradox is due to

Jourdain. Cargile concludes that Langford has already shown that the alleged paradoxes of Gardner and Popper have already been dealt with in Lewis and Langford's chapter on logical paradoxes in their Symbolic Logic.

Cargile summarizes "A Paradox Regained" as variations on the theme:

A: "K knows that A is false."

He points out that this is an old theme appearing in Buridan's Sophismata. In Cargile's opinion,

> . . . These "knower"-type paradoxes are just Liar-family paradoxes in which knowing is involved only in that it entails truth. "K knows that p is false" is logically equivalent to "p is false and K knows it." So A is fundamentally the same as
> B: "B is false and K knows it."
> B is just a case of the Conjunct-Liar, "This conjunction is false and q," which makes possible a semblance of proving the falsity of any q you please. Similarly with
> C: "K does not know that C is true,"
> which appears to be true but unknowable by K.
> It is fundamentally the same as
> D: "Either D is false, or D is true but K does not know it,"
> which is a case of the Disjunct-Liar.[31]

So unlike Bennett, Cargile is quite sympathetic to the self-referential approach.

According to R. A. Sharpe, the prediction paradox arises if both parties know and apply the rules set by the teacher's announcement. For then, all the days are eliminated. If the rules excluded all but one day, no paradox would arise.

> Since the rule here excludes all days in the week as possible days for the examination, to choose a day at all will be a surprise in the sense of displaying ignorance of or a deliberate breaking of the rule. An element of self-reference arises from the fact that on the terms by which the paradox can occur, the master must take into account the boy's own prediction before choosing a day. Since he cannot choose days which they have predicted, they negatively affect the choice and if they have played a part in making the choice it is difficult to see how it can surprise them.[32]

Sharpe points out that an announcement which only excluded one day would still be self-referential but no paradox would arise. He therefore concludes that self-reference is not a sufficient condition for the paradox. It is interesting to note that the "element of self-reference" to which Sharpe alludes, is not the kind of self-reference Bennett and Cargile considered. Sharpe's conception of self-reference seems to be game-theoretic.

J. M. Chapman and R. J. Butler in their "On Quine's 'So-called Paradox'", propose a "perverse solution" taking Quine's rejection of the base step of the induction as their inspiration. Like others, they argue that if the examination has not been given by Thursday, the students can deduce that the examination has not been given by Thursday, the students can deduce that the examination is on Friday and deduce that it is not on Friday.

> The conclusion that the examination must be held on the last day is just as warranted as the conclusion that it cannot be held. Therefore the boys cannot predict, by a valid process of logical argument and without laying themselves open to contradiction, that

> the examination will be held on the last day. Therefore the examination will be held on the last day. Therefore the examination, even if it is held on the last day, will be unexpected in the required sense.[33]

Another proposal put forth by commentators professing sympathy with Quine is "The Prediction Paradox Again". Here, James Kiefer and James Ellison first insist that the problem can only be made interesting and precise if surprise is defined in terms of deducibility.

> Let us use "deduce$_1$" to mean "deduce, using as premises the nonoccurrence of the examination up to the moment of deduction, plus the truth of this announcement". Let us use "deduce$_2$" to mean "deduce, using as the premise the nonoccurrence of the examination up to the moment of deduction". Let us define "surprise$_1$" and in terms of "deduce$_1$" and "deduce$_2$" respectively.[34]

Kiefer and Elison interpret the prediction paradox as showing that the announcement is contradictory if given the 'surprise$_1$' reading. However, if the announcement will true if the examination is given on any day of the week. Once this ambiguity is noted, the authors claim the paradox is resolved. They claim that if they have correctly understood Quine, he has largely anticipated their solution.

Quine's suspicions about the base step of the induction are shared in Judith Schoenberg's "A Note on the Logical Fallacy in the Paradox of the Unexpected Examination". The elimination argument begins with

the last day: if the examination has not been given by the penultimate day, then . . . . Schoenberg claims that the antecedent of this conditional illegitimately assumes that conditions laid down by the teacher have already been violated. The rest of the student's argument is "merely a verbal play". So although Schoenberg agrees with the student that the examination cannot be given on the last day, she believes that the student is arguing fallaciously when he begins his argument with the conditional 'If the examination has not been given by the penultimate day, then it must be given on the last'.

> . . . the premise entertains a condition under which the event cannot occur as defined, and thus cannot serve as the point of departure for a line of reasoning about the event's possibility. All it can lead to deductively is a clarification of the conditions under which the event cannot occur by definition.[35]

In "The Surprise Exam: Prediction on the Last Day Uncertain", J. A. Wright launches another attack on the base step of the induction. He suggests that the usual interpretation of the announcement is to the effect that:

> (1) A test will be held, and any one day of a given finite set of days is possible for it.
> (2) It will not be possible to predict the test, with logical necessity, on the morning of that day.

Wright then suggests that the paradox can be avoided by reading the announcement as saying

the last day: if the examination has not been given by the penultimate day, then . . . . Schoenberg claims that the antecedent of this conditional illegitimately assumes that conditions laid down by

the teacher have already been violated. The rest of the student's argument is "merely a verbal play". So although Schoenberg agrees with the student that the examination cannot be given on the last day, she believes that the student is arguing fallaciously when he begins his argument with the conditional 'If the examination has not been given by the penultimate day, then it must be given on the last'.

> . . . the premise entertains a condition under which the event cannot occur as defined, and thus cannot serve as the point of departure for a line of reasoning about the event's possibility. All it can lead to deductively is a clarification of the conditions under which the event cannot occur by definition.[35]

In "The Surprise Exam: Prediction on the Last Day Uncertain", J. A. Wright launches another attack on the base step of the induction. He suggests that the usual interpretation of the announcement is to the effect that:

> (1) A test will be held, and any one day of a given finite set of days is possible for it.
> (2) It will not be possible to predict the test, with logical necessity, on the morning of that day.

Wright then suggests that the paradox can be avoided by reading the announcement as saying

> (A) Any one of a finite set of days is a possible day for the test to be <u>planned</u>.
> (B) It will be cancelled if it is <u>acutally</u> predicted on the morning of that day.[36]

The change from "the test will be held on" to "the test is planned for" allows (A) to be cancelled rather than contradicted by (B). The change from possibility of prediction to actual prediction undermines

the base step of the induction according to Wright. The teacher will only refuse to consider a Friday examination if he is certain that a student will come to him with a prediction. Although it is highly probably that a student will do this, it is not certain. Thus the students cannot eliminate a Friday examination with certainty.

Later in 1967, M. J. O'Carroll published "Improper Self-Reference in Classical Logic and the Prediction Paradox" in Logique et Analyse. O'Carroll claims that although the prediction paradox has received much attention, it has not been correctly formulated. He argues that the teacher is really claiming that the students cannot deduce the day of the examination without there also being a counterdeduction that is not that day. O'Carroll also claims that the conclusion to be drawn is

> . . . either it is not true that there is an exam on one and only one afternoon "next week" or the the teacher's statement . . . falls outside the field of valid application of two-valued, non-levelled logic.[37]

Two years after his sympathetic review of the self-referential approach to the prediction paradox, James Cargile rejected this approach in favor of a game-theoretic approach. He conceives the problem as involving rational agents, one of which is trying to make a choice that cannot be predicted by others even though all the rational agents have the same relevant information. Cargile stipulates that the teacher has no means of randomizing his choice and that this is common knowledge. Besides knowing that the teacher prefers to give a surprise test, the students know that it is common knowledge that both

teacher and students are ideally rational agents. Since Cargile is interested in the two day version of the prediction paradox, it is also common knowledge that the test must take place either Thursday or Friday. Cargile believes this situation leads to a puzzle because

> . . . the following would appear to be an essential truth about ideal rationality: If two ideally rational agents are asking independently whether a give proposition is true and if both have exactly the same relevant data and exactly the same knowledge about what is relevant, then they will both reach the same conclusion. The conclusion may be "Yes" or "No" or "insufficient data to determine" or "the question is unclear," etc., but it must be the same for both. For suppose that the two agents arrive at different answers, X and Y. The X cannot be a better answer than Y on the information given, since that would contradict the assumption that both agents are ideally rational -- that is, think as well as is possible in every case. But then the answer "X" is no better than answer than "Y" is determinable on the information given and is clearly a better answer than X of Y, which contradicts the assumption that both agents will give the best possible answer on the information available to them.[38]

The teacher will think that the students might be surprised by a Thursday test just in case the students will. If the teacher thinks that there is no chance that a Thursday test will be surprising, then the students will know this as well, because they will have arrived at the same conclusion. If the teacher concludes that he cannot know whether there is a chance, then the students will know this. Cargile's point is that someone can surprise someone else only if the surpriser and the surprisee disagree about something at some time. SInce the teacher and the students are ideally rational agents with

the same relevant information, such a disagreement is impossible (given the principle mentioned above).

Cargile tries to solve the problem by introducing a third ideally rational agent; a judge to adjudicate the students' claim. Cargile argues that the students know that the test will be given on Thursday only if the judge would agree that they know. The students cannot know that the test will be given on Thursday because the teacher will only give the test on Thursday if he knows that the students do not know that the test will be given Thursday. If the judge ruled in favor of the students, he would be ruling against the judgment made by an ideally rational agent, the teacher. Since the students cannot satisfy this criterion of knowledge, they do not know. Cargile concedes, however, that the students can have justified confidence in the test being held Thursday. Indeed, since the standards for certainty fluctuate from context to context, Cargile is willing to allow that the students are certain that the test will be on Thursday in other, less stringent contexts.

Six months after publishing Cargile's article, the Journal of Philosophy published Robert Binkley's "The Surprise Examination in Modal Logic". Binkley's main achievement was to provide a rationale for Quine's claim that the prisoner cannot eliminate the last day because he does not know that the announcement is true. Binkley points out that the announcement corresponding to the single day version,

(11) You will be hanged tomorrow noon and will not know the date in advance, resembles the sentences G. E. Moore was so puzzled by:

(12) It is raining but I believe it is not raining,

(13) It is raining but it is not the case that I believe it is raining.

In Knowledge and Belief, Jaakko Hintikka argued that these sentences cannot be believed by perfect logicians even though (12) and (13) are consistent. Since the prediction paradox is a paradox for perfect logicians, Binkley points out that Hintikka's explanation of the incredibility of (12) and (13) can be extended to the question of why the prisoner cannot know (11). The prisoner cannot believe, and therefore, cannot know (11) because (11) is logically incredible to the prisoner. By appealing to the principle that if a perfect logician believes p, then he believes that he will believe p thereafter, Binkley is able to demonstrate that the announcement corresponding to the n + 1 day case are so incredible to the prisoner. So Binkley concludes that the prediction paradox is in the same family as Moore's paradox. In "Believing and Disbelieving", Kathleen Johnson Wu arrives at much the same conclusion as Binkley, differing only in that she sees no need to restrict the analysis to perfect logicians. About ten years after Binkley's article appeared, Igal Kvart published "The Paradox of Surprise Examination". Kvart never mentions Binkley or Wu but aside from greater care in formalization, does little else but duplicates Binkley's results.

When most people learn about the prediction paradox, they are inclined to accept the base step of the induction more readily than the induction step. The first commentator to plausibly follow this inclination was Craig Harrison in "The Unanticipated Examination in View of Kripke's Semantics for Modal Logic". Harrison considered the paradox as it arises in the following form. Student a is told by his instructor that there will be a test on either the second or the fourth of the month. The test will be unforeseen in the sense that if the test is given on the fourth of the month, then a will not know so on the third, and if the test is given on the second, a will not know so on the first. Although Harrison provides a formalization of the prediction paradox as it arises in this form, I prefer another formalization for reasons which will become apparent in the next chapter. Let 'Ka_ip_k' read 'a knows at day i that the test occurs on day k'. Where i<j, let 'S' be the conjunction: $((p_2$ $-Ka_1p_2)$ & $(p_4$ $(-Ka_3p_4$ & $Ka_3-p_2))$ & $(p_2$ v $p_4)$ & $(Ka_ip$ $Ka_jp))$. S represents the situation a is in. The second conjunct of S states that if the test should be given on the fourth, a will not know it on the third but will realize that no test has been given on the second. The fourth conjunct is the temporal retention principle; anything a knows, he knows thereafter. Although this Although this principle does not hold for ordinary people since they forget, die, go insane, etc., it does seem that these disturbances can be stipulated away, so that the principle holds for the ideal thinker a in an ideal epistemological environment. In addition to the standard rules of inference (TF), I shall appeal to the following rules.

KD: $\frac{K(A\&B)}{KA\&KB}$ $\frac{K(A \rightarrow B)}{KA \rightarrow KB}$ KI: $\frac{\vdash A}{KA}$ KE: $\frac{KA}{A}$

KEI: $\frac{KA \quad KB}{KC}$ Where $(A\&B) \rightarrow C$ is a truth of sentential logic. KK: $\frac{KA}{KKA}$

KD entitles distribution of the knowledge operator over conjunctions and material conditionals. KI makes all logical truths known and KE represents 'Knowledge implies truth'. KEI insures that the knower knows all of the consequences of what he knows, and can be derived from the preceding rules. KK guarantees that if one knows, then one knows that one knows.

{1}	1. Ka_3S	Assumption
{2}	2. p_4	Assumption
{1}	3. $Ka_3(p_2 \vee p_4)$	1, KD, TF
{1}	4. S	1, KE
{1, 2}	5. $-Ka_3p_4 \ \& \ Ka_3-p_2$	2, 4, TF
{1, 2}	6. Ka_3p_4	3, 5, TF, KEI
{1, 2,}	7. $-Ka_3p_4$	5, TF
{1}	8. $-p_4$	2, 6, 7, Reductio
{ }	9. $Ka_1S \rightarrow -p_4$	1, 8, conditionalization
{10}	10. Ka_1S	Assumption
{10}	11. Ka_3S	10, KE, TF (temporal retention)
{10}	12. $-p_4$	9, 11, TF
{ }	13. $Ka_1S \rightarrow -p_4$	10, 12, conditionalization

{ }	14. $Ka_1(Ka_1S \rightarrow -p_4)$	13, KI
{ }	15. $Ka_1Ka_1S \rightarrow Ka_1-p_4$	14, KD
{10 }	16. $Ka_1S \rightarrow Ka_1Ka_1S$	10, KK
{10 }	17. $Ka_1S \rightarrow Ka_1-p_4$	15, 16, TF
{10 }	18. Ka_1-p_4	10, 17, TF
{10 }	19. $Ka_1(p_2 \vee p_4)$	10, KD
{10 }	20. Ka_1p_2	18, 19, KEI
{10 }	21. $p_2 \rightarrow -Ka_1p_2$	10, KE
{10 }	22. p_2	20, KE
{10 }	23. $Ka_1p_2 \ \& \ -Ka_1p_2$	21, 22, 23, TF

Whereas Binkley suggested that we reject lines 1 and 10, Harrison suggests that we reject KK. KK is the most philosophically controversial principle in the set of rules and so is prime suspect. Further, we can follow our inclination to accept the base step of the induction and yet reject the induction step since appeal to KK is only necessary at step 16, leaving lines 1 through 9 intact. In other words, the instuctor can give the test on any day except the last. One might then conclude that the philosophical significance of the prediction paradox is that it is further evidence against KK.

Although Harrison's paper is not mentioned, his main results are duplicated in J. McLelland's "Epistemic Logic and the Paradox of the Surprise Examination" in the International Logic Review. A few years later, McLelland teamed up with C. Chihara to provide a more perspicacious duplication in their "The Surprise Examination Paradox"

in the Journal of Philosophical Logic (here, Harrison's work is cited).

After a few years of silence, the self-referentialists surfaced again with A. K. Austin's "On the Unexpected Examination" in 1969. Apparently in the hope of refuting the objection that the elimination argument fails to show that the students are expecting anything since they have an argument for eliminating every day of the week, Austin argues that the students can "expect" the examination everyday of the week by a series of incomplete proofs. The students should first construct a proof which follows the elimination argument only to the elimination of every day except Monday. Thus on Sunday, the students will be expecting a test on Monday. If no test is given Monday, the students should construct a similar proof which eliminates Wednesday, Thursday, and Friday. Thus on Monday evening, the students will be expecting a test on Tuesday. In a like manner, an expectation is formed for every day of the week.

In 1972, Paul Dietl published "The Surprise Examination" in Educational Theory. After criticizing Quine for failing to realize that a last day surprise examination is logically impossible, Dietl sets out to show that there are progressively weaker empirical grounds for ruling out the preceding days. The teacher will not give an examination on the penultimate day because he would have to assume that his students are stupid enough not to expect it since it is the last possible day. The third to last day is a highly improbable examination day but not as improbable as a penultimate day examination since we must assume that the teacher has gone through the preceding

reasoning in order to rule out Friday and Thursday. The fourth to last day is a genuine possibility since another assumption about the teacher's reasoning must be made. According to Dietl, once we reach, say, the 23rd to last day, it is plain that the students have no grounds to expect it then. So like the KK rejectors, Dietl accepts the base step of the induction but rejects the induction step. However, unlike the KK rejectors, Dietl provides a rough probabilistic ranking of the examination dates.

Later in 1972, another self-referentialist appeared in order to attack the "Quine-Binkley interpretation" and to exhibit the merits of conceiving the problem in terms of "formal prediction". Jorge Bosch provides five reasons for rejecting the Quine-Binkley interpretation:

(a) Before considering from the beginning as a possibility that the announcement will not be fulfilled, it is necessary to clarify the sense of "to know in advance".

(b) In Quine's version from "K persuaded himself that the sentence could not be executed" and "the arrival of the hangman took place at 11:55 the following Thursday morning", the conclusion "K's argument was was erroneous" does not follow.

(c) If we give the announcement the more precise form [that is, the self-referential form], the paradox remains in exactly the same terms, but the Quine-Binkley interpretation does not apply.

(d) If we accept the Quine-Binkley interpretation, which leads to the form [p & -Kap], the paradox does not remain in the same terms but we are faced with <u>another</u> problem. Condition at the end of 1.4 is not fulfilled [which is:] "To explain" or "to solve" the paradox signifies to give the announcement an

interpretation such that the informal proof be still relevant, and to decide--within the framework of this interpretation--whether the informal proof is correct or not and where does "the cause" of the paradoxical effect lie.

(e) In the usual form of the paradox, the announcement seems normal and the conclusion seems paradoxical, while in the Quine-Binkley interpretation the announcement seems paradoxical and the conclusion seems normal.[40]

Bosch then goes on to argue that although Bennett's review of Shaw shows that self-reference is not essential to the paradox, there is a kind of circularity involved due to

> . . . the unusual fact that the formal unpredictability of a proposition is referred to a system which includes the formal unpredictability of the same proposition.[41]

Despite the new terminology, Bosch's conclusion resembles Fitch's. The announcement is self-contradictory and is only psychologically puzzling because we tend to make the confusion Lyon dwelt on.

A late, but distinguished newcomer to the debate was A. J. Ayer in 1973 with his "On a Supposed Antinomy". Ayer argues that the puzzle turns on the ambiguity

> . . . between being unable to predict before the sequence is run through when the event in question will occur and being unable to make this prediction in the course of the run, however long it continues. In the first case, there is uncertainty, but in the second there may not be.[42]

To see how Ayer intends this distinction to work, consider the two day version of the surprise test variation. If the announcement means

that the students do not know which day has been selected at the time of announcement, then the announcement can fulfilled. If the announcement means that there could not be a time when the students knew which day the examination will take place, then the announcement is false. We fall into puzzlement when we project the second case on the first and argue that because there could be circumstances in which all uncertainty has been removed, there is no uncertainty at the start.

Further self-referentialist contributions were made by Peter Windt in his "The Liar in the Prediction Paradox" and by Martin Edman in his "The Prediction Paradox". Edman's approach resembles Fitch's. The teacher tells his class that (a) during the coming week there will be an examination, and (b) the day of examination will be a surprise. According to Edman, when we are tempted to view (a) and (b) as incompatible, we are interpreting (b) as self-referentially saying that the date of the examination cannot be deduced from (a) and (b). When we view (a) and (b) as compatible, we are interpreting (b) as either saying that the only relevant known fact about the examination date is (a) or as saying that the examination date cannot be deduced <u>two</u> days in advance. On the latter interpretation, the examination is a surprise because "If the warning time is shorter than the necessary reaction time we tend to say that the event came as a surprise." The paradox is resolved once the necessary distinctions are drawn. Thus Edman's proposal differs from Fitch's only in the kind of alternative sense of the announcement which can be confused with the self-referential one.

In "The Examiner Examined", B. H. Slater claims that the teacher is doing something reprehensible when he makes the announcement although the teacher does not thereby contradict himself. In the single day version

> He says 'There will be an exam on Tuesday', but he also denies that the pupils can know this by saying in addition 'The exam will be unexpected'. While not making contradictory remarks, by making the first remark in his position of authority he influences the truth of the second. It is not so much that the truth of the first opposes the truth of the second, but that his asserting the first makes the second untrue.[43]

Slater compares the teacher's announcement with someone saying 'My name is Tom; but you don't know what my name is'. According to Slater, although the conjuncts are compatible, the person's

> performance in telling us his name . . . makes impossible our remaining in ignorance of it. 'We don't know what your name is' is an inference from what has been said. 'We have been told your name is Tom' is a description of what has been said. It is these two which are at odds: one can't both not know, and at the same time have learnt, what his name is. But 'We learn, from what you say, that your name is Tom' is not deduced from what is said. It is something we say on the basis of observation, not inference. We learn by listening, not by arguing.[24]

In 1976, further dissatisfaction with the propriety of uttering the announcement corresponding to the single day version was expressed in T. S. Champlin's "Quine's Judge" in Philosophical Studies:

> The function of the notorious regress argument, bete noire of most commentators on the paradox, is to show: (i) that the judge who utters the words of the traditional version . . . is committed

> to uttering the words of the shortened version . . . should the execution occur on the last day (a thesis to which we have seen that Quine subscribes): and (ii) that a sincere judge who uttered the traditional words in normal circumstances would have to agree that for him to utter the words of the shortened version . . . would indeed be self-contradictory and thus he would detect a hidden contradiction in his original seemingly harmless pronouncement. Far from being incompatible with (ii), the possibility of Quine's querulous [man] and of the judge telling him the date whilst predicting his remaining ignorant is actually required by it. By shifting from the [longer] to the [shortened] version Quine sidesteps the key question posed by (ii), viz. 'Could a sincere and sober judge deliver his sentence as in [the longer version] to a reasonable [man] without self-contradiction?[45]

This concern with the pragmatic aspect of the teacher's announcement is also reflected in "The Paradox of the Unexpected Examination" published by Crispin Wright and Aidan Sudbury in the Autralasian Journal of Philosophy the next year. After stating the paradox, Wright and Sudbury list six conditions which they claim to be jointly sufficient for an intuitively satisfying solution.

(A) The account given of the content of the announcement should make it clear that it is satisfiable, since a surprise examination is palpably, a logical possibility.

(B) The account should make it clear that the headmaster can carry out the announcement after he has announced it since, palpably, he can

The two conditions require that the paradox not be construed as straightforwardly one of impredicativity or "pragmatic self-refutation'.

(C) The account must do justice to the intuitive meaning of the announcement. An extraordinary proportion of commentators have chosen to discuss quite unnatural interpretations of it. (D) The account must do justice to the intuitive plausibility of the pupil's reasoning.

(E) The account should make it possible for the pupils to be informed by the announcement: we want the reaction of someone who notices no peculiarity but just gets on with his revision to be logically unobjectionable.

(F) The account must explain the role, in the generation of the puzzle, of the announcement's being made to the pupils; there is, intuitively, no difficult if, e.g., the headmaster tells only the second master or keeps his intentions to himself. Most of the interpretations in the literature which identify the problem as one of impredicativity fail to meet this condition.[46]

The proposal which Wright and Sudbury believe can satisfy all of these conditions can be stated simply: reject the temporal retention principle. Thus in the proof presented a few pages ago, Wright and Sudbury would reject S because the fourth conjunct is the temporal retention principle. Following Binkley, Wright and Sudbury note the resemblance between the announcement corresponding to the single day case and Moore's problem. They agree with Binkley and Quine that the students cannot reasonably believe the announcement. They disagree on whether the students can reasonably believe the announcement corresponding to the n + 1 day case. According to Wright and Sudbury, the students can reasonably believe the n +1 day announcement as long as there are n days left. After all, there is nothing wrong with believing on Sunday that on one of the next five days one it will be

the case that 'Today is the examination day but is not the case that I believed so the night before'. However, if the test is not given by Thursday, there is something wrong in believing that on Friday it will be true that 'Today is the examination day but it is not the case that I believed so the night before'. The problem is Moore's problem since one would in effect be believing that it is both the case that there will be examination tomorrow and that one does not now believe it. Thus the teacher's announcement makes a hiatus in reasonable belief possible. People who are not surprisees are not vulnerable to this hiatus since Moorean sentences implied by the fact that the examination will be a surprise, are not about them. Conditions (A), (B), (E), and (F) are met since the teacher can give an informative announcement which will surprise the students even if he give it the last day and yet such a last day examination will not surprise nonstudents.

Reason to suspect that self-referentialists persist even today can be found in A. K. Austin's second contribution to the debate, "The Unexpected Examination", published in Analysis in 1979. After reviewing his previous results, Austin notes that if the announcement means that a single date cannot be deduced, the paradox returns.

Notes

1-D..J. O'Connor "Pragmatic Paradoxes", Mind, vol. LVII no. 22, (July 1948), p. 358.

2-Ibid., p. 359.

3-Ibid.

4-L.J. Cohen "Mr. O'Connor's 'Pragmatic Paradoxes' ", Mind, vol. LVIV no. 233, (January 1950), pp. 86-87.

5-Peter Alexander "Pragmatic Paradoxes", Mind, vol. LVII no. 22, (October 1950), p. 538.

6-Michael Scriven "Paradoxical Announcements", Mind, vol. LX no. 239, (July, 1951), pp. 406-7.

7-Paul Weiss, "The Prediction Paradox", Mind, vol. LXI no. 242, (April, 1952), p. 265.

8-Ibid., pp. 265-266.

9-W.V. Quine "On a so-called Paradox", Mind, vol. LXII, no. 245, (January, 1953), pp. 66-7.

10-R. Shaw "The Paradox of the Unexpected Examination", Mind, vol. LXVII no 267, (July, 1958), p. 386.

11-Ibid.

12-Ibid., p. 383.

13-Ibid.

14-Ibid., p. 384.

15-Ardon Lyon "The Prediction Paradox", Mind, vol. LXVII no. 272, (October, 1959), pp. 512-3.

16-Ibid., p. 513.

17-David Kaplan and Richard Montague "A Paradox Regained", Notre Dame Journal of Formal Logic, vol. 1 no. 3, (July, 1960), p. 80.

18-Ibid., p. 84.

19-Ibid.

20-Ibid., p. 87.

21-Ibid.

22-G. C. Nerlich "Unexpected Examinations and Unprovable Statements", Mind, vol. LXX no. 280, p. 503.

23-Ibid., pp. 506-7.

24-Ibid., p. 507.

25-Ibid., p. 508.

26-Ibid., p. 509.

27-Brian Medlin "The Unexpected Examination", American Philosophical Quarterly, vol. 1, no. 1, (January 1964), p. 67.

28-Ibid., p. 68.

29-Ibid., p. 69.

30-Jonathan Bennett in a review appearing in the The Journal of Symbolic Logic, vol. 30 no. 2, (June 1965), p. 102.

31-James Cargile in a review appearing in the The Journal of Symbolic Logic, vol. 30 no. 2, (June 1965), p. 103.

32-R. A. Sharpe "The Unexpected Examination", Mind, vol. LXXIV no. 294, (April 1965), p. 255.

33-J. M. Chapman and R. J. Butler "On Quine's 'So-called Paradox'", Mind, vol. LXXIV no. 295, (July 1965).

34-James Kiefer and James Ellison "The Prediction Paradox Again", Mind, vol. LXXIV no. 295, (July 1965).

35-Judith Schoenberg "A Note on the Logical Fallacy in the Paradox of the Unexpected Examination", Mind, vol. no. 297, (January 1966).

36-J. A. Wright "The Surprise Exam: Prediction on the Last Day Uncertain", Mind, vol. LXXVI no. 301, (January 1967), p. 115.

37-Ibid.

38-M. J. O'Caroll "Improper Self-Reference in Classical Logic and the Prediction Paradox", Loguique et Analyse, vol. 10, (June 1967), p. 171.

39-James Cargile "The Surprise Test Paradox", The Journal of Philosophy, vol. LXIV no. 18, (September 21, 1967), pp. 557-8.

40-Jorge Bosch "The Examination Paradox and Formal Prediction", Logique et Analyse, vol. 59-60, (September-December 1972), pp. 510-11.

41-Ibid., p. 524.

42-A. J. Ayer "On a Supposed Antinomy", Mind, vol. LXXXII no. 325, January 1973.

43-B. H. Slater "The Examiner Examined", Analysis, (December 1974), p. 50.

44-Ibid.

45-T. S. Champlin "Quine's Judge", Philosophical Studies, vol. 29 no. 5, (May 1976), p. 351.

46-Crispin Wright and Aiden Sudbury "The Paradox of the Unexpected Examination", Australasian Journal of Philosophy, vol. 55 no. 1, (May, 1977), p. 42.

CHAPTER TWO

CRITICISMS OF PAST PROPOSALS

Unlike most philosophical problems, the prediction paradox did not arise as an objection to a philosophical thesis. There appears to be no special kind of philosopher for, which it puzzling; even nonphilosophers can quickly appreciate the oddity of the elimination argument. Although commentators have tried to solve the prediction paradox by assimulating it to other philosophical problems, it is at least superficially irrelevant to any philosophical problem. This perceived irrelevance is the source of most philosophers' indifference toward the problem. Most philosophers who have commented on the problem have done so because they were intrinsically but not extrinsically interested in the problem. Few of them have attempted to alter the prediction paradox's reputation for being philosophically sterile, isolated for the rest of our philosophical concerns.

On the other hand, there are important similarities between the history of the prediction paradox and the histories of other philosophical problems. As usual, many of the commentators on the prediction paradox have misunderstood and duplicated the work of other commentators. Often their contributions seem to be the work of desperate men. Commentators on the prediction paradox have sometimes even called their own proposals "bizarre" and "perverse". More commonly, they have attempted to purchase a solution at the expense

of distorting the problem. Like other philosophical problems, the prediction paradox seems to have a spawned a controversy that does not appear to narrow with increasing contributions. Positions have proliferated without supplanting their predecessors. The controversy seems endless.

As with other philosophical problems, several schools of thought have developed. O'Connor and Cohen were the first and last veridicalists: people who believe that the prediction paradox is a veridical paradox and so believe that the clever student is right.

A plurality of commentators are self-referentialists: people who believe that an element of self-reference is responsible for the prediction paradox. Self-referenttialists are vulnerable to the charge of distortion. Their bad faith begins with their equation of knowability with deducibility. First, 'a deduces that p' neither entails nor is entailed by 'a knows that p'. Likewise 'a soundly deduces that p' and 'a knows that p' are mutually independent. Second, as Bennett emphasized, Shaw has already shown that self-reference is not a necessary condition for the paradox. Third, the self-referentialist fails to meet any of the conditions Wright and Sudsbury set for an intuitively satisfying solution. I regard these conditions as necessary conditions for a solution to the traditional variations of the paradox. The writings of later self-referentialists can be viewed as unsuccessful attempts to satisfy some of these conditions. Finally, I merely wish to note that those who try to show that the prediction paradox is liar paradoxical, have not been trying to solve the problem. There is a difference between problem reduction

problem solution. If one reduces the prediction paradox to the liar paradox and one does not have a solution to the liar paradox, one has only in the words of David Lewis, reduced two mysteries to one mystery. Although there is much to be said for mystery reduction, it is always preferable to have a solution. Since the commentators who have advocated reduction to the liar paradox, have not also advocated a solution to the liar paradox, they are not attempting to solve the paradox.

Overlapping the self-referentialists are the clarifiers. Clarifiers believe that people who are puzzled by the prediction paradox are puzzled because they do not fully understand the announcement. For example, Alexander argues that we overlook the fact that every declaration of an intention has an implicit 'if possible' clause. Most clarifiers think that there is an ambiguity involved, whose exposure solves the paradox. The most influential commentator among these equivocationalists is Ardon Lyon. In an unusual display of confidence in his fellow philosophers, Lyon began his proposal with the prediction that anyone who read it would accept it. Lyon's overconfidence is symptomatic of a difficulty common to all instances of this approach. If the paradox is really due to any of the equivocations that have been proposed, why do the vast majority of people who understand the proposal continued to be puzzled?

Many criticisms of the Quineans (Quine, Binkley, Wu and Igal), have already been recorded in the previous chapter. However, in light of Binkey's insight, there really is only one basic flaw to this proposal: failure to meet the informativeness condition.

The KK-rejectors have both formal elegance and our inclinations about the base step and the induction step on their side. I have two basic criticisms of this proposal. First, although doubts can be cast on the applicability of KK to ordinary people, these doubts cannot be extended to ideal thinkers in ideal epistemological enviroments. The most persuasive objection to KK requires commitment to a weak sense of knowledge; roughly, a knows that p if, and only if, a is right about p nonaccidentally. The most influential alleged counterexample of KK along these lines is Colin Radford's case of the unconfident examinee. Jean, a French-Canadien who believes himself ignorant of English history, agrees to answer some questions about it to humor a friend. Much to Jean's suprise, he does very well. In fact, he does so well that according to Radford, we should conclude that at the time of Jean's answering, Jean really knows the answers although he does not know that he knows. Radford makes his case more plausible by adding that Jean then remembers having learned some English history years ago. However persuasive Radford's case is against the KK principle for the weak sense of knowledge, it need not persuade those who wish to restrict KK to a strong sense of knowledge: a knows that p if, and only if, a has nondefective evidence for his true belief that p. Danto has argued against the KK principle on the grounds that knowledge implies understanding and since there are many knowers who do not understand what knowledge is, there are many cases in which someone knows without knowing that he knows. But the defender of KK can either deny that knowledge implies understanding (at least in the strong sense that Danto has in mind) or argue that Danto is concerned with a sense of knowledge which is stronger than the one the KK

defender has in mind. More straightforwardly, one can point out that Danto's argument is far too strong since it precludes knowing that others sometimes know. A plausible counterexample to KK must not conflict with the fact that one sometimes knows that others know, and the fact that one sometimes knows that one knows. Danto also uses the example of the sceptic who believes he does not know that he has two hands. Here we are inclined to say that the sceptic really knows that he has two hands despite the fact that he believes otherwise. But winning our assent to the proposition that he therefore knows without knowing that he knows is still difficult. The defender of KK can merely explain that the sceptic has conflicting beliefs about whether he knows. This sample of how the KK defender can meet objections to the application of KK to ordinary knowers should indicate how hardy an opponent he can be when he only has to defend the priniciple for ideal thinkers in an ideal epistemological environment. About the only agument that Harrison, McLelland and Chihara will be able to advance against the KK defender is that acceptance of KK leads to the prediction paradox. And of course this argument only works if there are no alternative solutions to the prediction paradox.

My second criticism of the KK-rejectors also applies to the temporal retention rejectors, Wright and Sudbury: neither the KK principle nor the temporal retention principle is essential to all variations of the prediction paradox.[1] Consider the designated student paradox. Robinson Crusoe discovers four people on his island in addition to Friday and names the rest of them for each day of the work week. Crusoe decides to teach them English history. Since his

resources are limited, he can only give one student a test. Since he wants the test to be a suprise, he first lines the students up in accordance with the order of the days in the work week so that Friday can see the back of Thursday and the backs of all those in front of Thursday, and Thursday can see the backs of Wednesday, Tuesday, and Monday (but not Friday's, since Friday is behind him), and so on. Robinson Crusoe then shows the students four silver stars and a gold star. He announces that he will put a gold star on the back of the student who has to take the test and silver stars on all the rest. Further, the test will be a surprise in the sense that the designated student (the one with the gold star) will not know he is the designated student until after the students break formation (although others can know). One of the students objects that such a test is impossible. "We all know that Friday cannot be the designated student since, if he were, he would see four silver stars in front of him and deduce that he must have the gold star on his back. But then he would know that he was the designated student. The designated student cannot know that he is the designated student; contradiction. We all know that Thursday cannot be the designated student since, if he were, he would see silver stars in front of him, and since he knows by the previous deduction that Friday is not the designated student, he would be able to deduce that he is the designated student. In a similar manner, Wednesday, Tuesday, and Monday can be eliminated. Therefore, the test is impossible. The teacher smiles, has them break formation, and Wednesday is suprised to learn that he has the gold star, and so is the designated student, and so must take the test.

In this variation, knowledge is accumulated perceptually rather than by memory. So no appeal to the temporal retention principle is needed. Thus, the designated student paradox undermines the Wright/Sudsbury proposal.

The designated student paradox also undermines the proposal that KK be rejected since it does not require appeal to KK. For the sake of simplicity and to stress the resemblance between the designated student variation and the traditional variations, consider a two person variation of the designated student paradox. Here, Alvin the first and Alvin the third are the students. Let 'Ka_ip_k' read 'Alvin i knows that Alvin k-1 is the designated student'. Let 'S' be the conjunction $((p_2 \rightarrow -Ka_1p_2)$ & $(p_4 \rightarrow (-Ka_3p_4$ & $Ka_3 \rightarrow p_2))$ & $(p_2$ v $p_4))$. In order to meet the requrements of informativeness, one is tempted to prefix S with (x)Kx where x ranges over the students. However, (x)KxS corresponds to a variation in which the teacher makes a private announcement to each student (so that none know the others know the announcement). Since we are concerned with a variation in which the announcement is public, where everyone knows that everyone knows the announcement, it appears that a faithful representation demands (x)Kx(y)KyS. However, it can be demonstrated that (x)Kx(y)KyS leads to a contradiction in an epistemic version of the modal system KT. KT is the system one obtains if one deletes the rule KK from the set of rules used in the proof appearing in the last chapter. Any normal modal analysis of epistemic logic must contain KT, so acceptance of any representation of the situation described by the designated student paradox which implies Ka_1Ka_3S requires rejection of the normal modal analysis of epistemic logic.

By reinterperting lines 1 through 9 in the previous proof in accordance with the interpertation above, the first nine lines of that proof can double as the first nine lines of the proof that Ka_1Ka_3S is inconsistent in KT.

{ }	10. $Ka_1(Ka_3S \nrightarrow -p_4)$	9, KI
{ }	11. $Ka_1Ka_3S \rightarrow Ka_1 - p_4$	10, KD
{12}	12. Ka_1Ka_3S	Assumption
{12}	13. $Ka_1 - p_4$	11, 12, TF
{ }	14. $Ka_3S \rightarrow S$	1, 4, Conditionalization
{ }	15. $Ka_3(Ka\ S \rightarrow Ka\ S)$	14, KI
{ }	16. $Ka_1Ka_3S \rightarrow Ka_1S$	15, KD
{12}	17. Ka_1S	12, 16, TF
{12}	18. $Ka_1(p_2 \vee p_4)$	17, KD, TF
{12}	19. Ka_1p_2	13, 18, KEI
{12}	20. $p_2 \nrightarrow -Ka_1p_2$	17, KE, TF
{12}	21. p_2	19, KE
{12}	22. Ka_1p_2 & $-Ka_1p_2$	19, 20, 21, TF

Since Ka_1Ka_3S is inconsistant in KT, anything implying it is likewise inconsistent. Thus the unquantified Ka_1Ka_3S & Ka_3Ka_1S cannot be faithful representation of a public, informative announcement to the pair, Alvin I and Alvin III. One can infer Ka_1Ka_3S from (x)Kx(y)KyS in a quantified KT, so it is inconsistent as well.

Like the traditional variations of the prediction paradox, the designated student paradox can be generalized to the "n-day" case. Commentators have overlooked another way in which the traditional variation can be generalized. Consider the $1 \leq m \leq n$ case where n is the number of days and m is the number of surprises. It is easy enough for the students to argue that the m^{th} test cannot be a surprise, since after the m-1 surprise they can perform the standard elimination for the single surprise case. The students can then argue that the m-1 test cannot be a surprise since after the m-2 surprise, there would be only one surprise test forthcoming (since the m^{th} test has been shown to be unsurprising), thus enabling them to employ the standard elimination argument once more. Having tamed the last two tests, the students can run the rest through the routine. The upshot seems to be that it is impossible to inform someone that he will be surprised m times within a period of n occasions. A parallel argument for the designated student variation seems to show that it is impossible to publicly inform a group of n students that they are going to receive m tests Crusoe-style.

The traditional variations of the prediction paradox as well as the designated student variation involve a single, rigid order of elimination. The paradox of the undiscoverable position is intended to show that this feature is not essential to the prediction paradox. Consider the following game played in the matrix below.

1	2	3
4	5	6
7	8	9

The object of the game is to discover where you have been initially placed. The seeker may only move Up, Down, Left or Right, one box at a time. The outer edges are called walls. If the seeker bumps into a wall, say by moving left from 1, his move is recorded as $\overline{L}$, and his position is unchanged. Bumps help the seeker discover his initial position. For instance, if he is at 7 and moves U, U, $\overline{L}$, the seeker can deduce that he must have started from 7. The seeker has discovered where he started from if he obtains a completely disambiguating sequence of moves, i.e. a sequence which determines the seeker's initial position.

If the seeker is given only two moves, it is possible to put him in an undiscoverable position. For instance, if he is put in a position 4, every possible two move sequence is compatible with him having started from some other position. Now suppose the seeker is told 'You have been put in an undiscoverable position'. He disagrees and offers the following reductio ad absurdum. "Suppose I am in an undiscoverable position. It follows that I cannot be in any of the corners since each has a completely disambiguating sequence. For instance, if I am in 3, I might move $\overline{U}$, $\overline{R}$, and thereby deduce my position. Having eliminated the corners, I can also eliminate 2, 4, 6, and 8, since any bumps resulting from a first move completely disambiguates. For instance, $\overline{U}$ is sufficient to show that I am in 2.

Since only 5 remains, I have discovered my position. The absurity of the suppostion is made futher manifest by the existence of eight other arguments with eight distinct conclusions as to my initial position. For example, I could conclude that I am in 6 by first eliminating the corners, then 2, 4, 8, and then 5 (by sequence L, $\overline{L}$, leaving only 6 remaining). If one individuates arguments by distinct orders of elimination, there are indeed more than eight arguments. I could also conclude that I am in 6 by eliminating in this order: 7, 4 (by U, $\overline{L}$), 8, 1, 2, 5, 9, and 3. Thus I cannot be put in an undiscoverable position."

The sacrifical virgin paradox is intended to show that the subjects need not know how many alternatives there are. Every fifty years the inhabitants of a tropical paradise sacrifice a virgin to the local volcano in an elaborate ceremony. Virgins from all around are blindfolded and brought before the volcano. They all hold hands in a line and can only communicate one sentence: 'No one to your right is the sacrifical virgin'. This sentence can only be signalled by squeezing the hand of the virgin to one's left. The virgins are reliable and dutibound to so signal if and only if it is known to be true. Besides all this, the virgins also know that a necessary condition for being the sacrificial virgin is that one remain ignorant of the honor until one is tossed in. The chief must take the leftmost virgin up to the mouth of the volcano, and if the offering is acceptable, push her in and tell the rest of the virgins to go home. If the offering is unacceptable, he sends that virgin home and repeats the procedure with the new leftmost virgin. This procedure

continues until one virgin is sacrificed, so it is known that one will be sacrificed. After hearing the announcement that one virgin will be sacrificed, someone objects that the cermony cannot take place. "The rightmost virgin knows she is rightmost since her right hand is free. She knows that if she is offered, then none of the virgins to her left have been sacrificed. So if she is the sacrificial virgin then she will have to be offered knowing that she is the only alternative remaining, and thus would know she is the sacrificial virgin. Since the sacrificial virgin must not know, the rightmost virgin knows that she is not the sacrificial virgin. This knowledge obliges her to squeeze the hand of the virgin to her immediate left signaling the sentence 'None of the virgins to your right is the sacrificial virgin'. This virgin is either the leftmost virgin or a middle virgin (a middle virgin is any virgin between the leftmost and rightmost virgins). If she is a middle virgin, she will reason that if she is offered, she will know that none of the virgins to her left have been sacrified. By the signal she knows that none to her right are sacrificial virgins, and thus she will be able to deduce that she will be sacrificed. But since the sacrificial virgin cannot know she will be sacrificed, this middle virgin knows she will not be sacrified. Therefore, she will squeeze the hand of the virgin to her left, triggering the same deduction if this third virgin is a middle virgin. Once the leftmost virgin is reached, she will know that none of the virgins to her right is the sacrificial virgin since she is the only alternative left. However, she would then both know and not know she is the sacrificial virgin. Therefore, the ceremony is impossible."

The rightmost and leftmost virgins only know that there are at least two virgins in line. Middle virgins only know that there are at least three virgins in line. In the versions previously considered it is essential that the subjects know what the alternatives are. In the surprise examiniation version the students need to know that examination is on one of the five weekdays. In the sacrificial virgin version the subjects do not even have a rough estimate as to how many alternatives there are. Middle virgins only know that they are somewhere in the middle of an arbitrarily long line. So it is not essential that the subjects know the order in which members of the series are arranged. Unlike the designated student paradox, the middle virgins repeat the same deduction but do not replicate each other's deductions. No middle virgin knows more than any other middle virgin. Unlike the other versions, there is no characteristic deduction for each subject. In the designated student paradox, Don can only eliminate himself by replicating Eric's reasoning; Carl can only eliminate himself by replicating Don's replication of Eric's reasoning, and so on. As in the surprise examination paradox, each virgin replicates the reasoning of her "future self" but not the reasoning of others.

To summmarize, these three variations of the prediction paradox show that the temporal retention principle, the KK principle, the order of elimination, and knowledge of the number of alternatives to be eliminated, are each inessential to the prediction paradox. Since past analyses of this paradox do assume the above are essential, I conclude that we do not yet have a formulation of the structure of the prediction paradox.

Notes

1-With the exception of a few minor changes, the following portion of this chapter is taken from my "Recalcitrant Variations of the Prediction Paradox" forthcoming in the Australasian Journal of Philosophy (probably in the December, 1982 issue). I thank the editor of this journal for permission to use this material here.

CHAPTER III

PURE MOOREAN PROPOSITIONS: A SOLUTION TO THE PREDICTION PARADOX

Despite my criticism of the followers of Quine and the Wright/Sudbury proposal, I think that they are on the right track. Moore's problem is relevant to the prediction paradox. Further, I think the difficulties associated with the above proposals are symptoms of the fact that Moore's problem has not been solved. So in this chapter, I propose a solution to Moore's problem. After explaining what Moore's problem is and after considering the main approaches toward solving the problem, I provide a definition of Moorean sentences in terms of pure Moorean propositions. My solution to Moore's problem essentially involves a description of how one can contradict oneself without uttering a contradiction, and a set of definitions that exactly determine which sentences are Moorean and which are close relatives of Moorean sentences.

Moore's problem is the problem of explaining why sentences like the following are odd.

(1) It is raining but I believe it is not raining.
(2) It is raining but it is not the case that I believe it is raining.

Many people are tempted to dismiss (1) and (2) as contradictions. But since, among other reasons, these Moorean sentences could be true, they cannot be contradictions. After all, (1) merely describes a

commissive error and (2) merely describes an omissive error. We often have a practical interest in conditionals whose antecedents are Moorean sentences. A person who believes 'If I eat this mushroom and it is, contrary to my belief, poisonous, than I will die' will probably not eat the mushroom unless his confidence in his belief is very high. But if the antecedent were a contradiction, the conditional would have no more practical interest then "If some mushrooms are not mushrooms, then I will die'.

Commentators on Moore's problem have also noted that the oddity of (1) and (2) is not present in their future and past tense counterparts.

(3) It is raining but I will believe otherwise in the future.

(4) It is raining but it is not the case that I believed so in the past.

Sentence (3) is a clumsy way of saying that one anticipates that one will change beliefs. (4) is an equally clumsy way of saying that the rain was not expected. Nor is the oddity present in the third person counterparts of (1) and (2).

(5) It is raining but he believes that it is not raining.

(6) It is raining but it is not the case that he believes that it is raining.

However, the first person plural and second person counterparts of (1) and (2) are odd.

(7) It is raining but we believe that it is not raining.

(8) It is raining but it is not the case that we believe it is raining.

(9) It is raining but you do not believe that it is raining.

(10) It is raining but it is not the case that you believe it is raining.

Although the oddity of (7) and (8) is the same as the oddity of (1) and (2), (9) and (10) seem to be odd in a different way. In the first person examples, the speaker appears to be contradicting himself. In the second person examples, the speaker seems to be saying something which is self-defeating.

A complete solution to Moore's problem should show how one, in some sense, contradicts oneself when one utters a Moorean sentence and should show what a Moorean sentence is. Commentators have concentrated on constructing proposals that satisfy the first condition. Their interest in the second condition has been slight. Although they are content to characterize Moorean sentences with the open sentence 'p but I do not believe it', there are Moorean sentences that do not conform to this characterization.

(11) God knows that we are atheists.

(12) Although you do not agree with me about anything, you are always right.

Neither (11) nor (12) bears any significant grammatical resemblance to the open sentence. Nevertheless, whatever tempts us to call (1) and (2) contradictions also tempts us to call (11) and (12) contradictions. So the second condition of providing criteria for what counts as a Moorean sentence cannot be satisfied with the open sentence characterization. Any proposal that does not satisfy the second condition is incomplete since one is left without a way of determining which sentences are "like" sentences (1) and (2).

The first adequacy condition is not entirely independent of the second. So although little attention has been given to the second condition, proposals that satisfy the first condition often have implications concerning the second. So it is not the case that all previous proposals are inadequate simply because they fail to supply criteria for being a Moorean sentence. Nevertheless, I think that those proposals which do not contain omissive errors concerning the second condition, contain commissive errors. As we shall soon see, past proposals satisfying the first condition usually imply definitions of 'Moorean sentence' that are either too narrow or too broad.

In "Mr. O'Connor's 'Pragmatic Paradoxes'", L. Jonathan Cohen argues that pragmatic paradoxes are consistent propositions which are falsified by their own utterance.[1] This claim was made with three of O'Connor's examples in mind:

(13) I remember nothing at all,

(14) I am not speaking now,

(15) I believe there are tigers in Mexico but there aren't any there at all.

However, none of (13)-(15) are clear illustrations of Cohen's definition. A good example for Cohen's definition is

(16) There are no sentence tokens.

The existence of any token of (16) is a sufficient condition for the falsity of the proposition it expresses. A token of (13) only falsifies the proposition it expresses if 'remember' is interpreted as 'habit memory'. A token of (14) only falsifies the proposition it

expresses if it is a spoken token. But even charitable interpreted, (15) does not conform to Cohen's definition. Suppose there are no tigers in Mexico but Jane believes that there are some there. Further suppose that she utters a token of (15) to make a point about pragmatic paradoxes. Her token would then express a true proposition.

Many commentators on Moore's problem have thought that the fact that tokeners of (1) and (2) cannot, in some sense, believe the propositions which their tokens express, is largely responsible for the oddity of (1) and (2). Overlapping this group are those who believe that the fact that (1) and (2) cannot, in some sense, be asserted is largely responsible. Commentators who adopt weak readings of 'cannot believe' are often vulnerable to the charge of triviality. After all, 'Pigs can fly' is consistent but incredible in a weak sense. Stronger readings, on the other hand, are often vulnerable to the charge of being false.

In Knowledge and Belief, Jaakko Hintikka explains the oddity (1) and (2) in terms of doxastic indefensibility. Roughly, a statement is doxastically indefensible if and only if the speaker cannot consistently believe it. Uttering such a statement is self-defeating since it gives its hearers all they need to overthrow the statement (for the exact definition of doxastic indefensibility, see Knowledge and Belief p. 71). Hintikka explains that when someone makes a statement there is a presumption that it is at least possible for the speaker to believe what he is saying. This presumption is violated in the case of Moorean sentences because it is obvious to the speaker and his audience that he cannot consistently believe what he is saying.

To demonstrate the doxastic indefensibility of (1) and (2), Hintikka uses a logic of belief that is a doxastic interpretation of the modal system deontic S4. Many philosophers are reluctant to accept Hintikka's analysis because they believe his logic of belief is too strong, even for ideal thinkers. In addition to committing us to an implausible view of belief, Hintikka's analysis makes a problematic appeal to what is obvious to speakers and their audiences. Thus his analysis is not generally accepted.

Some philosophers are reluctant to accept any logic of belief because they are sympathetic to the view that belief is "crazy as hell" and would endorse strong gullibilism.

(SG) Strong gullibilism: $(p)\ (x)\Diamond Bxp$.

(WG) Weak gullibilism: $(p)\ (\exists x)\Diamond Bxp$.

Hintikka's claim that the tokener of (1) and (2) could not possibly believe the proposition his token expresses is incompatible with (SG). Presumably, Hintikka would claim that no one could ever believe (18) It is raining but at no time does anyone believe it. Thus Hintikka's position is also incompatible with weak gullibilism.

Although (1) and (2) do appear to be in some sense unbelievable, (19) and (20) do not.

(19) I believe at least one false proposition.

(20) I fail to believe at least one true proposition.

Most people would assent to (19) and (20). So it appears that people only have difficulty in believing specific descriptions about their current errors. Also, tokens of (19) always express propositions that are incorrigible for their tokeners. For suppose that I falsely

believe (19). It would then follow that I have at least one false belief, namely (19). It would then follow that I have at least one false belief, namely (19) itself. Thus I can only falsely believe (19) if I truly believe it. So necessarily, if I believe (19), (19) is true. Further, anyone who believes (19) has inconsistent beliefs in the sense that it is logically impossible for all of them to be true. Max Deutshcer has tried to solve Moore's problem by pointing out that (1) and (2) resemble contradictions because their tokeners can believe them only at the expense of having inconsistent beliefs.

> To assert that p is to present 'p' to the audience as a view to which the speaker subscribes. Since the speaker presents 'p but I don't believe that p' to the audience as a view to which he subscribes, he presents to the audience a set of views such that it is logically necessary that one of them is false. This is what is wrong with asserting 'p but I don't believe that p'.[2]

Deutscher might concede that (19) could be reasonably asserted and might even concede that

(21) There is a proposition which I both believe and disbelieve,

could be reasonably asserted and yet dismiss them as counterexamples to his analysis on the grounds that the inconsistent set of propositions must be specified. (19) and (21) are believable because the speaker does not know which propositions are members of the inconsistent set. However, the Preface and Lottery paradoxes can be used in response to this move. In the (simple Preface paradox, we are told of an author who writes in the preface of his book that there are

sure to be some mistakes in his book. He has written similar books which were found to contain errors, knows that the subject-matter is tricky, etc. and in short, has excellent evidence that a few of his claims are false. If presented with any particular claim, however, he will affirm it. Though his remark in the preface is eminently rational, he appears to be committed to inconsisitent beliefs. Let $p_1, p_2, \ldots, p_n$ (where 'Bap_i' reads 'The author believes p_i). Assuming tht belief collects over conjunction, $Ba(p_1 \& p_2 \& \ldots \& p_n)$ holds. But for the reasons cited, the author of the book denies the conjunction of his claims, so $Ba\text{-}(p_1 \& p_2 \& \ldots \& p_n) \& Ba(p_1 \& p_2 \& \ldots \& p_n)$. Denying that belief collects over conjunction permits one to sharply distinguish between affirming a set of propositions distributively and affirming a set of propositions collectively. The Lottery paradox is a neater version of the Preface paradox. Now let 'p_i' stand for 'Ticket number i is the winning ticket in the lottery'. Given that the lottery is fair and that each ticket has an equal chance of being the winning ticket, one would naturally answer negatively to the question 'Do you believe p_i?' because it is wildly improbable (say n = 1,000,000). But after listing all of your claims about the tickets, you find that no ticket will win. An advantage of the Lottery paradox over the Preface paradox is that one can plausibly assume each p_i is equiprobable and that there is exactly one p_i which is true. The difference in convenience still leaves the two as essentially the same paradox, so one should expect that a correct solution to either disposes of both. I favor rejecting the principle that belief collects over conjunction.

Thus I accept the consequenes that a rational agent can knowingly have inconsistent beliefs in order that I avoid the consequence that he can knowingly be directly inconsistent with respect to a specific proposition q (agent a is directly inconsistent with respect to q just in case Baq & Ba-q). However, Deutscher cannot make the same move since he would then be admitting that one can reasonably set forth a specific, yet inconsistent, set of views. The sentence

(22) I believe p_1 and I believe p_2 and . . . and I believe p_n,

might be odd because of its length, but it not odd because of the utterer's unreasonable subscription to an inconsistent set of views.

Deutscher's analysis also fails to distinguish Moorean sentences from direct inconsistencies and blatant inconsistencies (a is blatantly inconsistent with respect to q just incase Ba(q & -q)).

(23) I believe it is both raining and not raining.

(24) I believe it is raini9ng and I believe it is not raining.

Although (23) and (24) are consistent propositions which cannot be uttered without, in some sense, contradicting oneself, they are not thereby Moorean. Unlike (1) and (2), (23) and (24) are not teasing. Since (23) implies that the tokener has conflicting beliefs about whether it is raining, we can understand how the tokener is contradicting himself without uttering a contradiction. But (1) and (2) do not even imply that their tokeners are inconsistent.

Commentators on Moore's problem tend to dwell on the obvious examples, (1) and (2). However, there are nonobvious Moorean sentences.

(25) The atheism of my mother's nieceless brother's only nephew angers God.

This implies "My atheism angers God', which in turn implies 'God exists but I believe that God does not exist'. Suppose Steve has little analytical talent but has a good memory and a healthy respect for authority. He overhears an authority assert that the atheism of his mother's nieceless brother's only nephew angers God. Since, contrary to the authority, Steve is a God-fearing but gossipy Christian, he believes the authority and tells his friend (25). One might object that Steve does not really understand (25) and so does not believe it. But this objection rests on two dubious assumptions. First, it assumes that believing that p implies understanding p. Yet people certainly seem to believe many things that they do not understand. For example, most people with casual contact with physics believe $E = mc_2$ and believe that space is curved. Second, even if believing implies understanding, one cannot require that one believes only if one believes all of the logical consequences of what one believes. Since the truths of logic are logical consequences of any belief, and no one believes all of these, it would follow that no one believes anything. Yet, nothing short of this condition for understanding will guarantee that no one believes the proposition expressed by a nonobviously Moorean sentence. Once one grants that some Moorean sentences are believable and assertible, one must reject the commonly held view that incredibility and unassertibility are essential features of all Moorean sentences.

Commentators also tend to use examples describing error about contingencies. I think commentators shy away from examples like

(26) Although I am an unmarried bachelor, I believe that all bachelors are married.

(27) All bachelors are males but I believe some bachelors are not males.

because analytical errors can be met with familiar kinds of criticisms. In (26) and (27), the belief claim alone provides a sufficient basis for criticizing the speaker. We know how to criticize someone for making a logical error and are familiar with the sense in which such a person is contradicting himself. So although (26) and (27) are Moorean sentences, they are impure cases. In addition to being criticizable in the same way a tokener of (1) or (2) is, a tokener of (26) or (27) can be criticized in other ways.

Although sentences (23) and (24) are not Moorean, they help one to understand Moorean sentences because they show that it is possible for someone to contradict himself without uttering a contradiction. In order to show how an utterer of (1) or (2) contradicts himself, I shall distinguish between two kinds of logical criticism. Let 'Batp' read 'a believes at time t that p'. A direct doxastic criticism of agent a at time t with respect to q on the basis of p is a criticism which attempts to show that p is a consistent proposition that implies that if a is absolutely thorough at t, then he is directly inconsistent with respect to q at t. Agent a is directly inconsistent at t, if and only if he both believes and d is believed, i.e., Batq & Bat-q. Agent a is absolutely thorough at t just in case his beliefs are deductively closed and distribute over material conditionals at t. His beliefs are deductively closed at t if and only if he believes all of the logical consequences of what he believes at time t; (q) (r)

(Batq & $\Box$(q→ r))→ Batr). It is empirically obvious that ordinary people are never absolutely thorough. Some may be more thorough than others, and some may increase their thoroughness, but none fully work out the consequences of their beliefs. Logical truths are consequences of any proposition, so anyone who has some beliefs and is absolutely thorough is logically omniscient. For us imperfect logicians, thoroughness and consistency are important desiderata which, nonetheless, cannot be completely satisfied in practice. Lastly, a's beliefs distribute over material conditionals at t just in case (q)(r)(Bat(q →r) →(Batq →Batr)). If a uttered a token of

(28) I believe I am taller than myself,

or any of (23), (24), (26), (27), he would be susceptible to a direct doxastic criticism. If he uttered (1), (2), or (25), he would be immune to this type of criticism. However, he would be susceptible to a belief-based criticism. A belief-based criticism of agent a at time t with respect to q on the basis of p is a criticism which attempts to show that p & Batp is a consistent proposition that implies that if a is absolutely thorough at t, then he is directly inconsistent with respect to q at t. Let the proposition expressed by (1) be q & Bat-q and the one expressed by (2) be q & Batq. One is now in a position to prove that anyone who utters (1) is susceptible to a belief-based criticism. This can be done by showing that the supposition that a has a true belief that (1) which does not lead to a direct inconsistency and does not prevent him from being absolutely thorough (see line 1 below), is inconsistent. So anyone who utters 'It is raining but I believe it is not raining' either does not correctly

believe the proposition expressed by his utterance, is directly inconsistent about whether it is raining, or is not absolutely thorough. So he is susceptible to a certain kind of criticism even if he is immune to direct doxastic criticism. In the proof below, notice that line 2 is justified by the fact that a's absolute thoroughness at t, abbreviated Tat, implies that his beliefs distribute over conjunction; (q)(r)(Bat(q & r) $\rightarrow$ (Batq & Batr).

1. ((q & Bat-q) & Bat(q & Bat-q)) & (-(Batq & Bat-q) & Tat)
2. Batq & BatBat-q 1, TF, UI
3. Batq & Bat-q 1, 2, TF
4. (Batq & Bat-q) & -(Batq & Bat-q) 1, 3, TF

This demonstrates that a is susceptible to belief-based criticism concerning q at t if he utters (1). Notice that Batp played an essential role in the criticism. The same holds for the criticism of (q & -Batq).

1. (q & -Batq) & Bat(q & -Batq) & (-Batq & Bat-q) & Tat)
2. Batq & Bat-Batq 1, TF, UI
3. Batq & -Batq 1, 2, TF

A pure Moorean proposition for a at t with respect to q is a proposition which cannot serve as a basis for a direct doxastic criticism of a at t with respect to q but which can serve as a basis for a belief-based criticism of a at t with respect to q. One can be inconsistent without being susceptible to either kind of criticism. Thus (19), (21), and (22) do not express propositions which are pure Moorean propositions. Steve's utterance of (25) does express a pure Moorean proposition. His friend could show him that he has contradicted himself concerning the existence of God by first assuming

Steve's belief is a true belief. Second, his friend could show Steve that a consequence of his belief is that Steve is an atheist who angers God. Third, he could show that this in turn implies that God exists but Steve believes that God does not exist. So Steve is committed to believing that God exists and to believing that God does not exist, and so he is contradicting himself. Of course, Steve does not actually both believe and disbelieve that God exists. Steve has merely failed to be absolutely thorough. His friend's criticism forces Steve to give up his belief (given that Steve wants to avoid contradicting himself).

Those who believe that (1) and (2) could not possibly be believed might object that my account does not explain this unbelievability. Given that they also believe that people cannot believe obvious contradictions, one could answer that there are different ways of contradicting oneself and so different ways in which one can obviously contradict oneself. Thus the unbelievability of (1) and (2) would be explained as on par with, but not reducible to, the unbelievability of obvious contradictions like (28).

As J. N. Williams has emphasized in "Moore's paradox: one or two?", there is an important difference between (1) and (2).[3] I think this difference is best brought out by distinguishing doxastic blindspots from other pure Moorean propositions. A <u>doxastic blindspot</u> is a pure Moorean proposition, p, satisfying the following condition: $-\Diamond$(p & Batp & Tat). Unlike the reductio ad absurdum showing that <u>q</u> & Bat-<u>q</u> is a pure Moorean proposition the reductio showing that q & Batq is a pure Moorean proposition did not require the conjunct -(Batq & Bat-q): the contradiction can be derived from just p & Batp & Tat.

So unlike q & Bat-q, q & -Batq is a doxastic blindspot.

Pure Moorean sentences can be defined in terms of pure Moorean propositions. An omnitemporal, universal, pure Moorean sentence is a sentence-type such that all of its tokens express propositions which are purely Moorean for everyone at every time. Consider

(29) No one has believed, believes, or will believe anything.

If an agent, a, believes this, he is committed to the direct inconsistency of both believing and disbelieving (29). Further, the proposition in question is a doxastic blindspot. The sentence

(30) No one believes anything now,

expresses propositions which are purely Moorean for everyone at the time of tokening but not before or after, and thus is universal but not omnitemporal.

A user, pure Moorean sentence is a sentence-type such that each of its tokens expresses a proposition which is a pure Moorean proposition for the user of that token at the time of use. For example,

(31) I have no beliefs now,

along with (1), (2), (7), (8), and (25), are user, pure Moorean sentences. Sentences like

(32) Ronald Reagan has no beliefs now,

are not user, pure Moorean sentences since people other than Ronald Reagan can use tokens of them without thereby expressing propositions which are pure Moorean propositions for them.

One might object that it is possible for a tokener of (31) to have expressed a proposition which was not Moorean for him.

Various kinds of cases can be imagined.

(i) Smith dies on January 1, 1982. The next day, Jones writes (31) on Smith's forehead thereby expressing the proposition that Smith has no beliefs on January 1, 1982.

(ii) Smith writes (31) into his will, to be read after he dies.

(iii) In order to have typewritten copy of Smith's will, his secretary types (31) along with the rest of the will.

(iv) Having only a French copy of Smith's will, Smith's lawyer translates the French counterpart of (31) into the English sentence (31).

Case (i) is constructed by exploiting the fact that first person pronouns can be used in an extended sense to merely denote the bearer of the sentence token containing the pronoun. In this sense, personal pronouns can denote inanimate objects, as when someone puts the label 'Please eat me' on a piece of cake. Case (ii) exploits the fact that temporal indexicals can be indexed to the time the token is produced or the time it is received.

For example, people with dictaphones commonly begin their recorded messages with 'I am not here now'. When, as usual, times of production and reception are the same, the two ways of indexing yield the same result. Case (iii) takes advantage of the fact that sentence tokens which are quotations need not have their indexical elements indexed to the quotation token; usually they are indexed to the original token. Case (iv) takes advantage of the parallel lack of dependency for tokens which are translations.

The use/mention distinction can be drawn to handle cases like (iii) and iv). Cases like (i) are avoided by the understanding that I am only concerned with tokens which are read in the unextended sense. Finally, cases like (ii) are avoided by limiting my claims to tokens in which there is no divergence between the results of production and reception indexing.

An addressee, pure Moorean sentence is a sentence-type such that each of its tokens expresses a pure Moorean proposition for the addressee of the token at the time that token is used. For example,

(33) Christmas is closer than you believe,

along with (9) and (10), are addressee, pure Moorean sentences. In Meaning, Stephen Schiffer claims that a speaker tells his audience that p only if he intends to produce in the audience the activated belief that p. Although there are counterexamples to this principle, I think this intention is common enough to explain why addressee, pure Moorean sentences seem self-defeating. The user of such a sentence can fulfill his intention only if the addressee contradicts himself.

A user-specific, pure Moorean sentence is a user, pure Moorean sentence which is neither universally nor addressee Moorean. An addressee-specific, pure Moorean sentence is an addressee, pure Moorean sentence which is neither universally nor user Moorean. A user-addressee-specific, pure Moorean sentence is a sentence which is both a user and an addressee pure Moorean sentence but which is not universally Moorean. For example,

(33) You and I are solipsists,

along with (9) and (10) (given the inclusive sense of 'we'), are

user-addressee-specific, pure Moorean sentences.

Interestingly enough,

(34) I do not exist,

can be shown to be a user doxastic blindspot. If a uses (34), he expressed the proposition that it is not the case that a exists. However, if a believes something, then a exists. Since the absolute thoroughness condition can be satisfied vacuously, uttering (34) does not make one susceptible to direct doxastic criticism. Susceptibility to belief-based criticism can be shown by merely supposing that a has a true belief that (34). Thus (34) is a sort of doxastic super blindspot, satisfying the condition: $-\Diamond(p \ \& \ Batp)$.

Given certain assumptions concerning the analytical connections between belief and various other propositional attitudes, sentences like the following can be shown to be pure Moorean sentences.

(35) It is raining but I doubt whether it is raining,

(36) It will rain but I expect that it will not rain,

(37) You know that it is raining but I do not agree with you.

For example, to doubt that p, is to disbelieve p. To doubt whether p, is equivalent to doubting whether -p; one is in a state of suspended judgment. So doubting whether p implies neither believing nor disbelieving p. But then (35) implies

(2) It is raining but it is not the case that I believe it is raining.

Since (35) cannot serve as a basis for a direct doxastic criticism, and (2) is a doxastic blindspot, it follows that (35) is a doxastic blindspot.

None of the following sentences count as Moorean under my present definition of 'pure Moorean proposition':

(38) It is raining but I do not know that it is raining,

(39) It is raining but you do not know that it is raining,

(40) No one knows anything.

Nevertheless, they have much in common with the previous examples of pure Moorean sentences. The oddity of (38) and (39) is not displayed by their past tense, future tense, and third person counterparts. Although the temptation is not quite as strong, one is still inclined to dismiss (38) and (40) as contradictions. Like (9) and (10), (39) seems self-defeating.

In Knowledge and Belief, Hintikka notes that the primary purpose of addressing a statement to a is to inform a of something. So it seems to follow that if one addresses

(41) p but you do not know that p

to a, it must be possible for a to know that what (41) expresses is true. Assuming (41) is not intended to convey something like 'p but you did not know that p', (41) appears to be equivalent to

(42) You know that the case is as follows: p but you do not know that p.

> Of course, what [(41)] expresses may very well be true. It may even be known to be true. But it can remain true only as long as it remains sotto voce. If you know that I am well informed and if I address the words [(41)] to you, these words have a curious effect which may perhaps be called antiperformatory. You may come to know that what I say was true, but saying it in so many words has the effect of making

> what is being said false. In a way, this is exactly the opposite of what happens with some typical utterances called performatory. In appropriate circumstances, uttering the words "I promise" is to make a promise, that is, to bring about a state of affairs in which it is true to say that I promised. In contrast, uttering [(41)] in circumstances where the speaker is known to be well informed has the opposite effect of making what is being said false.[4]

So according to Hintikka, (39) is an antiperformatory sentence. These kind of sentences have aroused the interest of some philosophers. For example, in "Meaning and Knowledge"[5], David Cole lists the following sentences in order to show that there is no logical connection between knowing the meaning of a proposition and knowing how one would in principle determine its truth value:

(43) No one has any self-knowledge,

(44) Everyone is unconscious,

(45) No one knows that p, yet p.

Sentence (40) fits in well with Cole's examples. Cole considers them counterexamples to the view that one can only know the meaning of a proposition if one knows how one would in principle determine its truth. We know what it would be like for (43)-(45) to be true enough though they could not possibly be known to be true. This is because (43)-(45) imply that they are not known to be true.

I think the similarity between pure Moorean sentences and (38)-(40) can be brought out by parallel definition. Agent <u>a</u> is epistemically, absolutely thorough at t, abbreviated $T^{e}at$, just in

case a knows all of the logical consequences of what he knows and his knowing distributes over material conditionals. Proposition p is an epistemic blindspot for a at t if, and only if p is consistent and the following condition is met: $-\Diamond(p \,\&\, Katp \,\&\, T^{e}at)$. In this case a is susceptible to knowledge-based criticism. Definitions for the various kinds of epistemic blindspot sentences are obtained by substituting 'epistemic blindspot' for 'pure Moorean proposition' in the definitions of the various kinds of pure Moorean sentences.

One might object that the definition of epistemic blindspot containts a redundancy since knowledge implies truth. My reply is that I wish to define these propositions without committing myself to an epistemic logic and presupposing this implication would commit me to an epistemic logic (albeit a very plausible one). The spirit of my definitions and solution to Moore's problem is Hintikka's despite the changes made in order to avoid commitment to a doxastic or epistemic logic.

In addition to this neutrality, my solution to Moore's problem does not appeal to the obviousness of certain inferences and it is not limited to explaining the oddity only for perfect logicians. My solution does not require idealization; it is completely at home in the ordinary world. Direct doxastic and belief-based criticisms are just two ways people criticize one another. The two kinds of criticisms are quite similar, in fact pure Moorean propositions are the only counterexamples to the thesis that they are the same. Crucial to my solution is the reconciliation of the fact that neither (1) nor (2) is a contradiction with the intuition that anyone who said

(1) or (2) would be contradicting himself. The reconciliation is brought about by denying the conditional 'If a contradicts himself, he must believe a contradiction'. The sense in which the sayer of (1) or (2) is contradicting himself is specified by my definitions of direct doxastic and belief-based criticisms. By defining 'pure Moorean proposition' in terms of immunity to direct doxastic criticism and susceptibility to belief-based criticism, the similarities between pure Moorean propositions and their cousins are illuminated. By distinguishing between user, addressee, and universal pure Moorean propositions, the Moorean sentences I first had to classify according to their grammatical features of person and number can now be distinguished without appeal to their grammatical features. Thus a sentence like

(46) No one, except for me, has any true beliefs

can be put into the same class as second person Moorean sentences like (9) and (10) even though it is not in the second person. Further generality is obtained by means of the parallel definition of epistemic blindspots. Thus, in addition to explaining how utterers of (1) and (2) contradict themselves, my solution satisfies the second condition of explaining what a (pure) Moorean sentence is. Since my proposal satisfies the previously stated adequacy conditions, my proposal is indeed a solution to Moore's problem.

Notes

1-Cohen's article appeared in Mind, vol. LVIV, no. 213 (January 1950).

2-Max Deutscher, "Bonney on Saying and Disbelieving", Analysis, vol. 27, no. 6 (June 1967).

3-Williams's article appeared in Analysis, vol. 39, no. 3 (June 1979).

4-Jaakko Hintikka, Knowledge and Belief, (Ithaca: Cornell University Press, 1962) pp. 90-91.

5-Cole's article appeared in Philosophical Studies, vol. 36, no. 3 (October 1979).

CHAPTER IV

NEW LIMITS FOR EPISTEMOLOGY

In this chapter I develop some of the implications epistemic blindspots have for epistemology. According to introductory philosophy books, epistemology is the study of the scope and limits of knowledge; the epistemologist is interested in what we can know and what we cannot know. I shall motivate the study of blindspots by showing how they can be used to establish new limits on knowledge in addition to showing how they reinforce familiar epistemological limits. Finally, I show that the manner in which blindspots reinforce one traditional epistemological limit can be used to criticize a recent psychologism which epitomizes the kind of error I hold responsible for the prediction paradox; confusing indefensibility with inconsistency.

Epistemic blindspots are important because they imply the existence of strange, new limits for epistemology. Let 'Kxtwp' read 'x knows at t in way w that p.' The principle that whatever can be true can be known can be given a strong and a weak formulation.
(SAA) Strong absolute access principle: $(p)\ (\Diamond p \rightarrow (x)(t)(w)\Diamond Kxtwp)$
(WAA) Weak absolute access principle: $(p)\ (\Diamond p \rightarrow (\exists x)(\exists t)(\exists w)\Diamond Kxtwp)$
Although these principles are tempting, they conflict with a major part of the epistemological enterprise: describing the limits of what we can know. For example, Kant held that propositions concerning the

nature of things in themselves cannot be known. Moses Maimonides argued that we can only know negative propositions about God. The most famous limit to scientific knowledge is Heisenberg's principle of uncertainty: one cannot know both the position and velocity of an electron. Even philosophically unsophisticated people express scepticism about the possibility of knowledge in politics, aesthetics, and morality. In any case, almost all epistemologists are committed to positions which imply a proposition of the form 'Only propositions about x can be known'. They thereby set a limit on the propositions that can be known. Thus almost all epistemologists are committed to the negation of (WAA) and therefore to the negation of (SAA). Of course, epistemologists have other interests. They often address the question of whether there is a basic way of knowing or whether a given way of knowing is reducible to another way of knowing. For example, Hume maintained that anything that can be known by testimony can be known by observation.

Even if one abandons the absolute access principles, one might still wish to claim that we can know the same things in the same ways.(IW) Interway access principle:

$(p)(x_1)(x_2)(w)(\Diamond(\exists t)Kx_1twp \rightarrow \Diamond(\exists t)Kx_2twp)$

However, propositions like 'Dan is blind' are counterexamples to (IW) since Dan cannot know he is blind by looking at his eyes but his optometrist can. Many philosophers reject (IW) by insisting that we have private access to some propositions. For example, I can know that I have a toothache by merely feeling it. However, no one else can feel my toothache, so no one else can ever know that I have a

toothache merely by feeling it. They must learn about my toothache in other ways (by being told, by observing me hold my jaw, etc.). This kind of limit engenders philosophical problems if conjoined with the thesis that ways other than the private way either are significantly less reliable or nonexistent. The problem of other minds is the problem of justifying one's psychological descriptions of other people. At least my simple psychological descriptions of myself ('I feel dizzy', 'I am in pain', 'I am thinking about red roses'), appear to be justified by introspection. Further, introspection seems to be the surest way of knowing our own psychological states. Since I cannot use this method to learn about other people's psychological states, it appears that I must settle for a second-best way of knowing their psychological states: observing their behavior. But why bother attributing any psychological states to them? (Or 'Why suppose that they too have minds?') The traditional answer was the argument from analogy: my psychological states correlate with kinds of behavior, so it is reasonable to assume that when others manifest those kinds of behavior, they have psycholgoical states similar to my own. But as Wittgenstein and Austin pointed out, this induction is too weak to justify the high degree of confidence we have in our psychological descriptions of others. So we are left feeling as though we are jumping to conclusions when we make apparently straightforward psychological judgments.

Despite their rejection of the principle that everything that is true is knowable and their rejection of the principle that what can be known can be known by all in the same way, few epistemologists have

rejected the principle that anyone who can know a proposition at one time, can know it at any other time.

(IT) Intertemporal access principle:

$$(p)(x)(t_1(t_2)\ (\Diamond (\exists w)kxt_1wp \rightarrow \Diamond(\exists w)kxt_2wp)$$

In addition, epistemologists have generally accepted the principle that whatever one can know, can be known by any other.

(IA) Interagent access principle:

$$(p)(x_1)(\Diamond (\exists t)(\exists w)Kx_1twp \rightarrow \Diamond(\exists t)(\exists w)Kx_2twp)$$

However, there are blindspots which are counterexamples to (IT) and (IA).

(17) John will first know that he was adopted on his eighteenth birthday.

(18) John never did, does not, and never will know that he was adopted.

Contrary to (IT), John can know (17) when he is nineteen but cannot know it when he is sixteen. Contrary to (IA), John's mother can know (18) but John cannot. On the other hand, there are some propositions which only John can know, like

(19) No one other than John knows that he keeps a diary.

Thus private knowledge is possible. Indeed, there are some propositions which can only be known by one person at one time, like

(20) The only one who ever knew that Mrs. Lincoln had gangrene was her doctor, and he only knew it a moment before he died of a stroke.

Epistemic blindspots can also be used against (SAA), (WAA), and (IW). Omnitemporal, universal, epistemic blind-spots are counter-examples to (WAA) and thus (SAA). A counterexample to (IW) is

(21) Before now, Dan did not know anything.

Although Dan cannot know (21) by memory, someone else can.

The assumption that logical and epistemic space are coextensive is a natural one. Further, it is an assumption that some philosophers now accept, and indeed, has served as the basis for a recent challenge to the classical definition of validity.

In "Epistemic Foundations of Logic" and more recently in his book Rational Belief Systems, Brian Ellis argues in favor of an epistemic definition of validity.

> I consider an argument in a given language to be valid if there is no rational belief system on that language in which its premises are accepted and its conclusion rejected. Thus, for me, validity is an epistemic notion, and my problem is to define a rational belief system on a language, and to specify acceptance and rejection conditions for the sentences of that language.[1]

Ellis expresses dissatisfaction with the classical definition of validity which states that an argument is valid if and only if there is no interpretation of its nonlogical terms in which its premises are true and its conclusion is false. Ellis develops his epistemic conception of validity in order to avoid problems which seem to inevitably follow from the classical definition of validity and to explain why tautologies cannot be rationally rejected. After developing some languages he hopes will supplant our familiar first order, second order, and modal languages, Ellis summarizes his approach.

> The concept of a rational belief system on a language is that of a belief system on a language which is ultimately defensible before an audience of competent speakers. I offer completability through every extension of the vocabulary of the language as my criterion for ultimate defensibility. That is, if a belief system is rational, then it is possible to decide every undecided sentence of the language, and of every extension of the language, without violating any requirements of linguistics competence.
>
> Given this theory of the rationality of a belief system, a tautology of a language may be defined as any sentence of the language which is not rejected in any rational belief system on the language. Thus, we can explain why the tautologies of a given language may not be rejected by any rational speaker.[2]

Claiming that one has a foundation for logic commits one to the completeness claim of being able, in principle, to provide a correct analysis of all sentences in a language. However, there is reason to doubt that Ellis' epistemic foundation for logic can handle belief sentences themselves. Ellis does not provide a language to reflect the logic of belief and it seems that his definition of validity prevents him from doing so. Consider the following sentences.

(1) Someone has a belief.

(2) No one has a belief.

(3) Gold is expensive.

Since (1) cannot be rejected by anyone who is consistent, and consistency is a necessary condition for a rational belief system,

(1) is not rejected in any rational belief system. In view of Ellis' definition of validity, it seems that any language for belief sentences must treat (1) as a tautology even though it is not a classical tautology. Then, the argument which has (3) as its sole premise and (1) as its conclusion is valid. Another example is the argument whose sole premise is (2) and whose conclusion is (3). Since (2) is not accepted in any rational belief system, there is no rational belief system in which (2) is accepted and (3) is rejected. So it appears that any language for belief sentences conforming to Ellis' definition would make epistemic validity broader than classical validity.

In effect, Ellis' new foundation for logic abrogates the distinction drawn by Hintikka in Knowledge and Belief between doxastic indefensibility and inconsistency. One's reluctance to accept these new foundations should therefore be at least equal the degree to which one feels this distinction is worth preserving.

If Ellis' logic must treat belief sentences as I have argued, then his logic provides a harsher environment for the sceptic than classical logic since there would be no difference between a universally indefensible position and an inconsistent position. Then once one has shown that a position is universally indefensible, one has shown that the position is false. In classical logic, the distinction between universal indefensibility and inconsistency is preserved, so the possibility of true yet universally indefensible positions is preserved. Thus in classical logic one cannot prove that a sceptical position is false by proving that anyone who believes it

is in an indefensible postion. In Ellis' logic, it seems one can. In that case there is an important philosophical difference between the two definitions of validity. The question of whether logical space coincides with epistemic space is not philosophical difference between the two definitions of validity. The question of whether logical space coincides with epistemic space is not philosophically sterile.

Notes

1-Brian Ellis, "Epistemic Foundations of Logic", Journal of Philosophical Logic vol. 5 no. 2, (May, 1976), p. 187.

2-Ibid., p. 201.

CHAPTER V

THE BLINDSPOT FALLACY:

A SOLUTION TO THE PREDICTION PARADOX

In this chapter I propose a solution to the prediction paradox. In the first section I consider analogues to the single day variation of the surprise test variation. These analogues are epistemic blindspots from credible sources. I argue that they are puzzling because they invite us to mistake certain inductive arguments for deductive arguments. In order to avoid this mistake, it is necessary to become familiar with the limits for epistemology examined in the previous chapter. If one assumes the interpersonal access principle, one is naturally drawn into the blindspot fallacy of confusing blindspots with contradictions. In section 2 the significance of the single day variation, and thus its analogues, for the debate within the Quinean tradition it shown. Quine's rejection of the base step of the induction is vindicated with the help of a comparison between the analogues and this step. In the final section Binkley's error concerning the induction step of the prediction paradox is corrected by pointing out that it rests on the intertemporal access principle. I then show how the designated student paradox rests on the fallacy of division. My treatment of the traditional variations of the prediction paradox resembles the Wright/Sudbury proposal. But there are two important differences between our proposals. First, I accept

and develop a couple of counterintuitive consequences of their proposal and my proposal. Second, I deny that one solves the prediction paradox in a purely negative fashion by rejecting the termporal retention principle. I argue that the prediction paradox is a symptom of our unfamiliarity with the limits blindspots place on knowledge. Unfamiliarity with how the interpersonal access principle fails is the principal cause of the error concerning the base step and unfamiliarity with how the intertemporal access principle fails is the principal cause of the error concerning the induction step. Our unfamiliarity with the limits established by epistemic blindspots leads us to either assume that we can know more than we can or less than we can. The former error dominated the literature on the prediction paradox before Quine, and the latter error dominated the literature after Quine. Solving the prediction paradox is a matter of avoiding these two errors and correctly drawing the line between the knowable and the unknowable.

It should be noted that I will sometimes claim that someone cannot know p where p is an epistemic blindspot. In such cases, the 'cannot' should be read as 'cannot without susceptibility to knowledge-based criticism'. My appeals to the epistemic logic KT are only used to establish that p is an epistemic blindspot. KT can be used in this way because its inference rules are embodied in the definition of 'epistemic blindspot'. I remain officially neutral on the question of which, if any, epistemic logic is correct.

It should also be noted that I make use of the distinction between blindspots and conditional blindspots. A conditional

blindspot is a conditional which is equivalent to a conditional whose consequent is an epistemic blindspot. For example, 'If Ralph survived, he is the only one who knows it'. Although it is possible to know a conditional blindspot, and it is possible to know its antecedent, it is impossible to know both at the same time.

Suppose you turn on your radio and hear the announcer say:

(1) Today's anonymous sponsor is Alvin White.

You might be tempted to say that the announcer has contradicted himself, reasoning as follows. The announcer's audience knows that

(2) Today's sponsor is Alvin White,

since the announcer has told them (1). But by (1), it also follows that

(3) The audience does not know that today's sponsor is Alvin White.

So the announcement implies that the audience knows and does not know (2). Therefore, the announcer has contradicted himself.

However, (1) is not a contradiction. Given that 'anonymous' in (1) does not exclude radio station personnel from knowing, the announcer could privately tell (1) to the station manager without absurdity. The absurdity seems to arise only if (1) is announced publicly. A clearer example is

(4) I am keeping the following a secret from you: John intends to give Bob a puppy for Chirstmas.

Someone who asserts (4) seems to be falsifying what he says merely by saying it. On the other hand, someone who asserts

(5) I am keeping the following a secret from Mary: John intends to

give Bobby a puppy for Christmas,

does not appear to be falsifying the proposition expressed by (5) (assuming he is not addressing (5) to Mary).

Someone might object that, precisely interpreted, (1) is a contradiction. What (1) really says, according to this objection is that members of the audience cannot deduce the identity of today's sponsor from (1). So despite appearances, (1) is self-referential. Besides being able to deduce (2) from (1), the audience can deduce (3), which really says that members of the audience cannot deduce (2) from (1). Since (1) implies the contradiction that one can and cannot deduce (2) from (1), the proper conclusion is that (1) is a contradiction.

Suppose Brother Jay, a religious yet logical man, stops you and asserts

(6) You agnostics anger God.

If you were impressed by the preceding argument that (1) is a self-referential contradiction, you might try the same argument against (6). "Assuming agnostics neither believe nor disbelieve that God exists, and assuming that knowledge implies belief, (6) really says that you cannot deduce

(7) God exists

from (6) since (6) implies

(8) It is not the case that you know that God exists.

However, I can deduce (7) from (6), contrary to the real meaning of (8), so (6) is a contradiction." Brother Jay concedes that the deductions from (6) to (7) and from (6) to (8) are valid. He even

claims that they are sound. Brother Jay can do this because deducing p does not imply knowing p, or vice versa. For example, Brother Jay can perform the following sound deduction even though he does not believe in evolution.

(9) All vertebrates evolved from primitive life forms.

(10) Human beings are vertebrates.

(11) Human beings evolved from more primitive life forms.

Brother Jay does not know (11) because he does not believe (9). Brother Jay denies that (6) says anything about deducibility. His denial seems more plausible than the denial that (1) makes no claim about deducibility because we are not tempted to say that Brother Jay divulges the existence of God in asserting (6) to you.

Despite this difference between (1) and (6), they share the unusual property of not being knowable to certain people. Since Brother Jay is logical, he realizes that you cannot know (6) because it could then be true that you both know and do not know that God exists. He told you (6) in order to bring it to your attention, not to let you know that (6) (although perhaps he did want to let you know that he believed (6)). To see that (1) cannot be known to the announcer's audience, let 'Kap' read 'The audience knows that today's sponsor is Alvin White'. Then the announcement, (1), implies (p & -Kap). The proof that the audience cannot know (1) runs as follows.

1.	Ka(p & -Kap)	Assumption
2.	Kap & Ka-Kap	1, KD
3.	Kap & -Kap	2, KE, TF

Notice that the station manager, b, can know (1) since no contradiction can be obtained from Kb(p & -Kap).

When we infer that the audience knows that Alvin White is today's sponsor because the announcer told them, we tend to think that the inference is the straighforward one from Ka(p & -Kap) to Kap, just like the inference that the station manager makes from Kb(p & -Kap) to Kbp. Although the former inference is valid, the above proof demonstrates that it is unsound. Unlike the station manager, the audience cannot know p merely by taking the announcer's word for it. The audience needs a more complicated argument.

(12) The announcer, who is generally reliable, said that today's anonymous sponsor is Alvin White.

(13) The announcer is much more likely to have inadvertently divulged the sponsor's identity than to have been lying, joking, of misspeaking.

---

(2) Today's sponsor is Alvin White.

If members of the audience use an inductive argument like this one, they can know (2). They would naturally be described as knowing (2) because they were told (1). At this point, the subtle difference between knowing because one was told so and knowing by authority is overlooked. The station manager knows (2) because he knows (1) by the announcer's authority, so we have a simple, salient deduction from Kb(1) to Kb(2) explaining why the station manager knows (2). The simplicity and saliency of this deduction leads us to erroneously assume that the audience's knowledge of (2) is explained in the same way as in the case of those people not in the extension of 'anonymous'; by their knowledge of (1). We then validly infer from Ka(1) that the audience both knows and does not know (2). We are left

feeling that the sponsor's anonymity is preserved in a perverse way because the audience's only apparent way of knowing the sponsor's identity (knowing (2)) also guarantees that the audience does not know the sponsor's identity. However, as explained by the above inductive argument, the audience can know (2) without knowing (1). Once we realize that '_a_ knows that p because _a_ was told that p & -Kap' is part of an _inductive_ explanation of Kap and cannot be a _deductive_ explanation of Kap, we begin to realize that part of our puzzlement is the result of our attempt to fit a piece of inductive reasoning into the mold of a simple, salient deductive inference. As shown in my discussion of the falsity of the interway access principle, the fact that a person can know a proposition in a particular way does not imply that it can be known in the same way by anyone else.

When Brother Jay tells you that (6) is true, you are not led into a similar puzzle because asserting (6) does not provide the basis for a strong inductive arguement for God's existence that can be confused with the simple deductive argument for God's existence from your knowing (6). Suppose a wild-eyed drunk suddenly confronted you with a box of crackerjacks and while pointing to the dog pictured on the box told you

(14) You don't know it, but this dog's name is Bingo.

Probably, you would not believe that the dog's name is Bingo because the drunk is not a credible source. If so, the drunk would have been correct since the dog's name really is Bingo. Some people have little difficulty in keeping a secret because nobody listens to them. The kind of puzzle (1) can lead us into are ones involving sources

credible enough to provide the bases for inductive arguments that can pass as deductive ones. We only try to give explanations of how someone knows something if we believe he really knows it. In the absence of a reason to suppose the fact that the drunk told you (14) gives you sufficient evidence for you to know (14), you are not tempted to offer an erroneous explanation. A man tries to hammer a nail with a glass bottle only if he believes there is a nail to hammer. You are not led to misexplanation because you do not believe that there is any knowledge to be explained.

In the surprise test variation of the prediction paradox, the elimination of the last day is quite persuasive. The announcement conjoined with all the information the students have at their disposal implies

(15) There will be a test tomorrow but you (the students) do not know that there will be a test tomorrow.

One then argues that since the students know (15), they know that there will be a test tomorrow. But for the reasons given concerning (1), any argument from the students knowledge of (15) to their knowledge of a test on Friday must be unsound. As described in the chapter on the history of the prediction paradox, commentators have attempted to show that (15) is a self-referential contradiction. In addition to the criticisms recorded in that chapter and the one following it, there is the new problem that this approach would also make unpuzzling statements like (6) and (14) self-referential contradictions. Other variations of the self-referential approach are also affected by (6) and (14) since these statements should be as

puzzling as the puzzling cases if this approach is correct.

Quine attempts to solve the prediction paradox by denying that the students can eliminate the last day since they do not know that the teacher's announcement is true. Quine's proposal has been criticized on the grounds that he does not have a good reason for claiming that the students do not know (15). After all, the students do appear to know on the basis of the teacher's authority. We do not want to deny that one can know by authority since so much of our knowledge is based on authority.

Binkley was the first to attempt a defense of Quine's proposal by supplementing Quine's argument with a plausible reason why the students cannot know (15). Binkley points out the resemblance between (15) and the sentences whose oddity constitutes Moore's problem. Since it is widely agreed that one cannot consistently believe a Moorean proposition and (15) seems like a Moorean proposition, Binkley had produced a good reason why the students cannot know (15). Binkley's attachment to the temporal retention principle then forced him to conclude that the original announcement,

(16) There will be a surprise test next week,

was just as unknowable to the students as (15).

Champlin, Wright, and Sudbury rejected this supplement to Quine's proposal because the teacher's announcement of (16) informs the students that (16). If the original announcement is informative, then the students must know (16). Wright and Sudbury then tried to show that (16) can be knowable while (15) is not, by rejecting the temporal retention principle. Although this proposal seems acceptable when one

considers the traditional variations of the prediction paradox, it fails in the face of the designated student variation.

The progress made in the Quinean tradition can be briefly summarized. Quine's crucial contribution to the discussion on the prediction paradox is his avoidance of the blindspot fallacy, the fallacy of mistaking blindspots for contradictions. Quine's analysis would have been more persuasive if it contained an adumbration of the nature of blindspots. Binkley et. al. made progress by exploiting the resemblance between Moorean sentences and the announcement. Unfortunately, their attachment to the temporal retention principle led them to view conditional blindspots as actual ones. Here, the mere possibility that the students are put in a Moorean bind by the teacher failing to give the test by the penultimate day, is considered sufficient to infect the original announcement with incredibility. Wright and Sudbury make a major contribution by showing how Binkley's initial insight that the prediction paradox has a Moorean element undermines the temporal retention principle. Given the effectiveness of their proposal in handling the traditional variations of the prediction paradox, they quite understandably err by considering the rejection of this principle as a sufficient rather than a necessary condition for the solution of the prediction paradox.

Wright and Sudbury emphasize that their proposal has the advantage of allowing them to maintain that the students are informed by the teacher's announcement and yet the students can no longer know the announcement on Thursday if the test has not yet been given. Further, their proposal enables them to explain why outsiders can know

on the basis of the teacher's announcement that a test will be given Friday if no test is given by Thursday, despite the necessary ignorance of the students. However, they do not call attention to the fact that this latter consequence implies that ideal thinkers can disagree. I do not consider this implication to be a sound objection to the proposal put forth by Wright and Sudbury, since, as I shall later show, the possibility of disagreement between ideal thinkers can be established independently of the prediction paradox. I consider the possibility of such a disagreement to be a veridical paradox, and so consider the fact that it is implied by this proposal to be a virtue rather than a vice of the proposal. One might object that real-life students are much more likely to infer that a test will be given Friday if no test has been given by Thursday, than they are likely to be thrown into doubt. In defense of Wright and Sudbury, one can use the example of the radio announcer who says:

(1) Today's anonymous sponsor is Alvin White.

There is nothing wrong with saying that the audience knows that Alvin White is today's sponsor because the announcer told them (1). However, it is fallacious to argue that the audience must therefore have deduced the identity of today's sponsor from their knowledge of (1). Unlike radio station personnel, the audience cannot know the sponsor's identity merely by knowing (1) since it is impossible for the audience to know (1). The audience needs to know some additional facts about the announcer. The same holds for the real-life students the objector has in mind. They employ the extra premise that the teacher prefers to give a test (albeit unsurprising) than no test at

all. Given that they know this premise is true, the students do justifiably believe that a test will be given Friday. It should be noted, however, that as common as this preference is, it does not hold universally in practice. There are some teachers who would prefer to give no test at all. In any case, the students described in the paradox do not have this extra information about their teacher's preferences. And if we redescribe the paradox so that they do have the necessary psychological premises, the clever student's elimination argument is undermined since he can no longer eliminate the last day.

One might object that Wright and SUdbury cannot explain how it is possible for the students to know the announcement at the begining of the week on the basis of the teacher's authority and yet have the epistemic force of this authority evaporate by Thursday. Although Wright and Sudbury reply that the students' evidence has changed by Thursday, one can still object that this evidence is irrelevant to the question of how reliable the teacher is. After all, nonstudents still know the announcement on Thursday on the teacher's authority.

My reply is that the students no longer know on Thursday by the teacher's authority because there is no way to know something which is unknowable. No one, no matter how authoritative he was, could get the students to know by authority that

(15) There will be a test tomorrow but you (the students) do not know that there will be a test tomorrow.

It is not the case that the student's have acquired new information which undermines the teacher's authority. Their ignorance is explained by the trivial principle that one can only know p by authority

if one can know p. Their ignorance is an example of the breakdown of the intertemporal access principle. The students are no longer in a position to know.

One might object that the students can anticipate that the teacher might maneuver them into the blindspot just described and so should have little confidence in the teacher's announcement when they first hear it. As Binkley says "You cannot trust a man who tells you things you cannot believe". Binkley's claim seems plausible because he does not relativize belief and trust to times. Although I cannot trust at t_1 what you say if I do not believe at t_1 what you say, nothing prevents me from trusting what you say at t_1 if I do not believe you at t_2. If anything, the possibility of the teacher maneuvering the students into a blindspot should increase their confidence in his announcement at the beginning of the week since this possibility ensures that the teacher can give the surprise test on any day of the week, making the announcement more probable.

In the case of the designated student variation of the prediction paradox, it is important to bear in mind that the announcement is addressed to a group of students. This feature is relevant since informing a collective differs from informing an individual. To say that a group knows something is not to say that all of its members know it. To infer $(x)\ (x \in g \rightarrow Kxp)$ from Kgp is an instance of the fallacy of division. Groups know that member-wise blindspots are true of some of their members. Suppose the students are members of the student council. After hearing that Eric has complained that the cafeteria's food is not nutritious, the council may state that Eric

is mistaken. If one employs the pattern of inference 'Person e is a member of group g', 'g knows p', therefore 'e knows p', one is committed to 'Eric knows that he mistakenly believes that the cafeteria's food is not nutritious'.

'The designated student does not know that he is the designated student' can be used to make an informative announcement to the class. The students standing in front of Eric cannot know that Eric is not the designated student since the supposition that he is the designated student only leads to the conclusion that the announcement is a blindspot for Eric. Unless Eric has outside evidence, he cannot know that he is the designated student. In the situation described in the paradox, no such extra information is available, so Eric cannot know. Thus the base step of the induction is fallacious. Rather than representing the informativeness of a public announcement to the class with the prefix 'Everyone knows that everyone knows that', the prefix "The class knows that' suffices. The designated student variation is solved by moving from the latter prefix with greater caution.

The undiscoverable position and the sacrificial virgin variations present no difficulty once the solutions to the variations just considered are understood. These variations depend on the temporal retention principle. The crucial point is that both variations are avoided by avoiding the blindspot fallacy despite their different structures. Solving the prediction paradox does not require an investigation into the various formal structures its variations can take on. Rather, it is a matter of familiarizing oneself with the peculiar borders these variations create between the knowable and the unknowable.

CHAPTER VI

PRE-DECISIONAL BLINDSPOTS AND PREDICTIVE DETERMINISM

In order to support my contention that the prediction paradox is a symptom of our unfamiliarity with the epistemic limits marked by Moore's problem, I have devoted this chapter to a more traditional philosophical problem which is also symptomatic of our unfamiliarity with these limits: predictive determinism. In Section I, two recent attempts to prove that decisions are not caused are examined and rejected. The next section covers two recent attempts to reconcile determinism and decision-making through an infinite regress arguement. Section III is devoted to a refined version of this infinite regress argument developed by Michael Scriven. My promissory note for an independent argument for the possibility of disagreement amongst ideal thinkers is fulfilled in Section IV as a criticism of Scriven's argument against predictive determinism. Nevertheless, in the final section I conclude that predictive determinism is false.

The problem of determinism is usually presented in terms of a conflict between one's inner conviction that one is free and one's scientific commitment to causal determinism. People who argue that there is a conflict, incompatibilist, argue that if one is causally determined, then one is forced to fulfill one's causal fate. Although we may feel like we could do otherwise, this feeling must be illusory given causal determinism. Incompatibilists who reject causal

determinism are called libertarians. Incompatibilists who accept causal determinism and reject freewill are called hard determinists. Although causal determinism is the most discussed form of determinism, it should be noted that there are as many kinds of determinisms as there are kinds of laws (logical, physical, psychological, etc.).

Philosophers have made their talk about "feeling free" more exact by concentrating on decision-making. In "Deliberation and Foreknowledge", Richard Taylor takes great care in distinguishing between deciding and predicting. Taylor points out that we only deliberate about that which we believe is not inevitable. If one knows that p, then one cannot deliberate about whether one should try to make p the case. Taylor then argues that if one knows that either p is inevitable or -p is inevitable, then one cannot deliberate about p. Resignation is the only option in such cases. So if one knows that someone else either knows that p or knows that -p, then one cannot deliberate about whether to try to make it the case that p. Taylor then argues that no one can know what another person will do as a result of that person's deliberation.

> If someone knew what another was going to do as a result of forthcoming deliberation, then he would know on the basis of some kind of evidence; that is, on the basis of knowledge of certain conditions that were sufficient for the agent's doing the thing in question, and from which it could be inferred that he would do that. But if there were such conditions they could also be known by, or made known to, the agent himself, such that he too could infer what he was going

> to do. Indeed, the agent cannot even believe that any such conditions, known or unknown, exist, and at the same time believe that it is within his power both to do, and to forego doing, the thing in question. This, as we have seen, appears to be a necessary condition of deliberation.[1]

Taylor's rejection of the predictability of decisions is unqualified.

> If God had foreknowledge of the deliberate act of some man, then that knowledge could be shared with that man himself. At least, there is no reason why it could not. But that is impossible, for no man can continue to deliberate about whether to do something, if he already knows or can know what he is going to do.[2]

Carl Ginet presents a more concise version of this argument in "Can the Will be Caused?". According to Ginet, one can soundly conclude that the will cannot be caused from the following two premises.

> (i) It is conceptually impossible for a person to know what a decision of his is going to be before he makes it.
> (ii) If it were conceptually possible for a decision to be caused, then it would be conceptually possible for a person to know what a decision of his was going to be before making it.[3]

Premises (i) and (ii) have both been challenged. As one might surmise from the preceding chapter, I reject (ii). Taylor defends (ii) with tacit appeals to the unlimited access principle and interagent access principle. Ginet is more sophisticated.

> One can, of course, describe a set of circumstances that it would be logically impossible for the decider to know in advance of his decision. (One need only include in the set of circumstances that the decider remains ignorant of certain other circumstances in the set at least until the time of the decision. It might be imagined, for example, that an agent's having a certain set of desires, beliefs, perceptions, and attitudes was always sufficient to produce a certain decision provided also that the agent was not aware at the time of some of those attitudes.) And a set of circumstances would not be a less plausible candidate for the cause of a decision merely because it has this feature. But neither could a set of circumstances be ruled out as a candidate for the cause merely because it lacked this feature.[4] I reject Ginet's last sentence. My basis is

my acceptance of a weaker version of (i):

(i') It is conceptually impossible for a person to know what a decision of his is going to be immediately before he makes it.

Unlike (i), (i') is compatible with someone predicting his own decision as long as there is a period of ignorance immediately preceding his decision. All decisions are epistemic blindspots for their deciders immediately preceding the time of decision. Since this is a consequence of (i'), and (i') is a consequence of (i), Ginet must also accept this blindspot thesis. But then he has a perfectly good reason for ruling out any set of circumstances which enabled the

decider to predict his own decision. Accepting such a set would conflict with (i). Thus Ginet's argument fails to show that our decisions are not caused.

The only other way (ii) has been challenged is by an infinite regress argument. In "Causes, Predictions and Decisions", Andrew Oldenquist rearranges Ginet's argument into three parts.

> (1) It is conceptually impossible that A, B, C are causally sufficient for decision D.
> (2) It is conceptually possible that I can know that A, B, C are causally sufficient for D.
> (3) It is conceptually possible that I can know my own decision in advance.[5]

(1) is the antecedent of (ii) and (3) is the consequent of (ii). Oldenquist concedes that (3) is false but points out that Ginet appeals to (2) when inferring (3) from (1). Oldenquist's goal is to show that, contrary to Ginet, (2) does not follow from (1), and moreover, (2) is false. However, Oldenquist feels that some preliminaries must be dealt with first.

> I will take 'know' in 'causal knowledge of the future' to be much like the ordinary strong sense of 'know': If it is true that someone has causal knowledge that the sun will rise tomorrow, then it must be the case that (1) he predicts this event (3) his prediction comes true. We must distinguish between predictions based on only some of the relevant causes, and which luckily may come true, and causal <u>knowledge</u> of the future, which is based on what one knows to be all the relevant causes.

> We often predict the likelihood or probability of our future decisions, but this is vastly different from knowing our own decisions in advance.[6]

Oldenquist then supposes that he has unlimited knowledge of past circumstances and relevant causal laws. Assume that he makes a prediction, P, and A, B, C are sufficient for D. Since P is a prima facie probable cause of his decision, Oldenquist would need to know that it is false that

(4) A, B, C, plus P are sufficient for not-D.

And indeed Oldenquist could know that (4) is false if he enjoyed the sort of epistemic bliss we are supposing. But his coming to know this generates another prima facie probable cause of his decision. Calling this new knowledge P_1, Oldenquist must know it is false that

(4a) A, B, C plus P, plus P_1 are sufficient for not-D.

Although Oldenquist can know this too, his coming to know it generates P_2, which in turn must be checked. So it appears that the process of eliminating P, P_1, P_2, etc. as countervening causes of his decision leads to an infinite regress. Therefore, Oldenquist cannot know that all circumstances besides A, B, C are causally irrelevant to his decision. The prediction cannot be completed. So it is possible for (1) to be true and (2) to be false, so (ii) is false. In "How Decisions are Caused", David Gauthier uses pretty much the same argument. Unlike Oldenquist, Gauthier emphasizes that this criticism of Ginet's argument allows that any causal principle and any circumstance is knowable. The only limit it imposes is that the complete conjunction of propositions about these principles and

circumstances is unknowable. We can know any of the conjuncts but not the conjunction taken as a whole. Gauthier also points out that the infinite regress criticism is also compatible with someone other than the decider predicting the decider's decision since such a predictor can know the relevant causal laws and circumstances and know that the decider is unaware of them.

The strangest use of the infinite regress argument occurs in D. M. MacKay's "On the Logical Indeterminancy of a Free Choice". MacKay first argues that predicting someone's decision requires that the predictor keep his prediction secret from the decider. If the decider were to learn what the prediction is, his decision might be affected, and so a new prediction would have to be made incorporating the decider's exposure to the first prediction. Without secrecy the predictors fall into an infinite regress. Whereas Oldenquist and Gauthier argue that an infinite regress arises for self-prediction, MacKay is arguing that an infinite regress arises for other-prediction if there is no secrecy. His next move is to argue that if there is such secrecy, a sort of perspectivalism arises. In the situation where a group of predictors keep their prediction about A's decision a secret, the predictors are right in believing that the decider's decision is determined and the decider is right in believing that his decision is not determined.

> There is no dispute that <u>they</u> are right to believe what they do about A's brain-processes [MacKay is interested in brain-process determinism]. But even they would insist that A would not be right to believe the same, since a precondition of

> [the prediction's] validity is that A must not be influenced by it. Clearly then the onlookers' view represents a true description of the state of affairs only for the onlookers, since if it were universally true, A would be wrong not to believe it . . . . Thus on the one hand, the idea that either party can give a universally-valid description of the 'true state of affairs' in this case is false; on the other hand, any idea that this proves that there is no 'true state of affairs' is invalidated on the assumption that the two descriptions stand in a rigid relationship. We might call them two different but related 'linquistic projections' of one and the same state of affairs. It is perhaps not surprising, if tantalising, that no single standpoint, whether of onlooker or agent, appears to allow us to put into words the whole truth about ourselves.[7]

MacKay uses this predictor/decider perspectivalism to attack the inference from 'A does not know what B knows' to 'A is ignorant of a fact known to B'.

> The interesting point which emerges is that what we are tempted to call A's 'ignorance' would not be remedied by supplying him with the proposition P describing the state of affairs to which we are trying to say he is ignorant, since P would lose its factual status if A were to entertain it. In short, the onlookers have no predictive information to give A, even if they would. A may not realise this, and may even 'wish he knew' what they know; but in respect of predictive information his wish is based on a fallacy--the fallacy of supposing that what he wants to know is a universal fact. The truth would seem to be that at this point there is no gap in his knowledge; the place of the onlookers' knowledge is already preoccupied for him by the knowledge that the choice awaits his decision. To make room for it, he would have to resign from his role as agent: but then the choice would not be

made.[8]

According to MacKay, one can only be ignorant of that which is logically determinate. In the case of logical indeterminacy, the question of ignorance does not arise. All ignorance is remedial. A proposition is logically indeterminate just in case its being believed is one of the factors determining its truth-value. Our subjective conviction of freedom does not rest on our unpredictability but rather on the fact that

> For us as agents, any purported prediction of our normal choices as 'certain' is strictly incredible, and the key evidence for it unformulable. It is not that the evidence is unknown to us; in the nature of the case, no evidence-for-us at that point exists. To us, our choice is logically indeterminate, until we make it. For us, choosing is not something to be observed or predicted, but to be done.[9]

MacKay disagrees with those who maintain that 'One cannot predict one's own decision because one must overlook at least one relevant bit of information about one's decision (due to the infinite regress)'. This disagreement does not mean that he believes one can predict one's own decision. MacKay disagrees because the situation is logically indeterminate; there is nothing to overlook because one can only overlook something one is ignorant of.

MacKay explicitly rejects what he calls the presupposition of transferability:

> . . . if A agrees that B is right in believing P, A logically commits himself to P and to all consequences deducible from it. Despite its obvious validity in most contexts, we have seen that it can break down where P is an assertion about an agent viewed by an observer. . . . This denial of simple transferability

> constitutes a kind of philosophical Principle of Relativity, very different from that exaltation of the arbitrary which goes by the name of 'moral relativism'. It resembles rather Einstein's physical principle in its insistence (i) that only one rigorously prescribable belief is valid for A if B's belief is also valid, but (ii) that the validity and meaningfulness f a belief may depend in a definite and rigorous way upon who entertains it. It differs, however, in giving no guarantee that A can even formulate from his standpoint the belief that would be valid for B (until it is out of date) and in making no assumption that their situations must be symmetrical.[10]

MacKay denies that he is committed to twe kinds of truth. However, he adds that the Principle of Relativity

> . . . does suggest that—and why—the traditional method of <u>comparing notes</u> in order to 'arrive at the truth' must break down in certain special cases, leaving the truth in such cases incapable of unique and universally valid expression.[11]

Thus MacKay endorses a form of ineffabilism.

I think MacKay's position can be best diagnosed in terms of the new limits for epistemology presented in the previous chapter. Given the unlimited access principle and KE, it follow that

(5) If A is ignorant of the fact that p, then it is logically possible that A knows that p.

MacKay's case of the predictor and the decider is a case where it is impossible for B to know the prediction. MacKay's commitment to (5)

thus forces him to deny that 'A does not know what B knows' implies 'A is ignorant of a fact known to B'. MacKay's attack on the presupposition of transferability springs from his acceptance of

(6) If B believes A is wrong to hold a certain position, then B believes A can adopt B's position.
A can adopt B's position.

Since A, the predictee, cannot adopt the position of B, the predictor, then by (6), B should find nothing wrong with A holding a different position. MacKay's "philosophical Principle of Relativity" is the negation of the interagent access principle with a dash of perspectivalism and ineffabilism.

Although I endorse MacKay's rejection of the interagent principle, I think he should also reject the unlimited access principle and thus (5). This would undermine his ineffabilism since it undermines

(7) If there is a universally correct description of a situation, then it is possible for anyone to know this description is accurate.

I think MacKay's rejection of transferability is due to the ambiguity of 'B is right to believe P'. This can mean 'B's belief is true' and it can mean 'B's belief is justified'. Philosophers who hold the presupposition of transferability read the principle in the first sense, not the second. MacKay attacks the second reading of the priniciple. His point is that people can reasonably disagree by virtue of their roles (predictor vs. decider).

In "An Essential Unpredictability of Human Behavior", Micheal Scriven gives a new twist to the infinite regress argument. Scriven's goal is to drive a wedge between causal determinism and predictive determinism by showing that there is a certain kind of situation in which it is impossible to predict what person will do. Suppose there is person whose dominant motivation is to avoid having his actions predicted. So if the predictor informs that avoider of his prediction, the avoider will act contrary to the prediction. Scriven then points out that there is a difficulty with predictor's attempt to prevent the avoider from knowing the prediction by keeping it a secret. According to Scriven, the avoider might have enough data, laws, and calculating capacity to duplicate the predictor's calculation to find out what result it gave. If predictive determinism is true, then even this avoider's acts can be predicted. But since this avoider could falsify any prediction, predictive determinism is false.

In "Scriven on Human Unpredictability", David Lewis and Jane Richardson object that Scriven falsely assumes that the predictor and the avoider can simultaneously have all the requisite data, laws, and calculating capacity. Predictive determinism is expressed by the conditional that anyone's act can be predicted if the predictor has all the requisite data, laws, and calculating capacity. If it should turn out that Scriven's compatibility assumption is false, he will have failed to shown the falsity of predictive determinism. Lewis and Richardson attack the compatibility assumption on the grounds that

> . . . the amount of calculation required to let the predictor finish his prediction depends on the amount of calculation done by the avoider, and the amount required to let the avoider finish duplicating the predictor's calculation depends on the amount done by the predictor. Scriven takes for granted that the two requirement-functions are <u>compatible</u>: i.e., that there is some pair of amounts of calculation avaiable to the predictor and the avoider such that each has enough to finish, given the amount the other has.[12]

Lewis and Richardson argue that the compatibility premise should be rejected because it leads to Scriven's thesis of essential unpredictability. According to them, we are tempted to accept the assumption of compatibility because of an ambiguity.

> It is true that against any <u>given</u> avoider the predictor can in principle do enough calculation to finish; it follows (unless the Compatibility Premise is true) that any possible avoider is in principle predictable. It is likewise true that against any <u>given</u> predictor the avoider can in principle do enough to finish: it follows . . . that any predictor is in principle avoidable. But to say that <u>both</u> can in principle do enough to finish is ambiguous. It may be read as the Compatibility Premise, i.e., as stating that against <u>each other</u> both can do enough to finish. We must see that we do not accept the Compatibility Premise inadvertenly by slipping from the first to the second.[13]

The Lewis and Richardson criticism can be neatly summarized by first letting 'Pyx' read 'y predicts x's behavior' and 'Fyx' read 'y finishes his calculations against x'. Predictive psychological determinism is

(8) $(x)(\exists y)\Diamond Pyx$

Scriven uses

(9) $(x)(\exists y)\Diamond Fyx$

to show that

(10) $(x)(\exists y)-\Diamond Pyx$

but he also needs to show

(11) $(\exists x)(y)-\Diamond Pyx$

But (11) can only be proved with the help of the compatibility assumption

(12) $(x)(\exists y)\Diamond(Fyx \ \& \ Fxy)$.

Lewis and Richardson concede (9) and (10), but accuse Scriven of fallaciously inferring (12) from (9) to prove (11) and thus to refute predictive psychological determinism, (8).

In addition to my sympathy to the above objection to Scriven's argument for human unpredictability, I have misgivings about his duplication claim. For if anyone can duplicate the reasoning of anyone else, then it appears that anyone can know what anyone else can know. So the duplication claim seems to imply the already rejected interagent access principle. Rather than resign himself to weakening his position by saying that sometimes people can duplicate the reasoning of other people, Scriven might be tempted to merely limit his duplication claim to ideal thinkers. After all, his example involves someone with knowledge of all the relevant data and laws, and with a perfectly reliable calculating capacity. Thus Scriven's predictor and avoider are perfect inductive and deductive logicians,

quite unlike ordinary people. We might then interpret the duplication claim as saying that thinkers, idealized in the manner prescribed by Scriven, can always duplicate each others' reasoning since, given that they have the same information, they must reach the same conclusion. Given that they have the same information, ideal thinkers must agree. Scriven's argument against predictive psychological determinism can be recast into a form resembling Cargile's treatment of the prediction paradox. Recall the way decision-making places us into blindspots.

(i') It is impossible for a person to know what a decision of his is going to be immediately before he makes it.

To say that I have predicted how you will decide to act, is to say that I have known something which you will not know immediately before you decide. But if we are ideal thinkers with the same information, then how is your future ignorance to be explained? (i') together with the following constitute an inconsistent triad.

(13) Given the same information, two ideal thinkers know exactly the same propositions.

(14) Given the same informatin, it is possible that one ideal thinker predicts the decision of another ideal thinker immediately before the decider makes his decision.

Scriven would have us reject (14) and conclude that the decision of an ideal thinker cannot possibly be predicted by another ideal thinker if the two have the same information and the prediction is made immediately before the decision.

Rather than (14), I find (13) the most suspicious proposition of the three. As mentioned previously, the possibility of disagreement

amongst ideal thinkers is implied by the Wright/Sudbury proposal concerning the prediction paradox. If no test has been given after the penultimate day, nonsurprisees cannot know. I accepted this consequence as holding even for ideal thinkers and claimed that the possibility of disagreement between ideal thinkers can be independently supported. In the next paragraphs I will honor this promissory note.

It should first be noted that a rather undisturbing form of disagreement can arise between ideal thinkers because they can make divergent arbitrary guesses. If two ideal thinkers are taking a test, and are able to eliminate the first two alternatives for the first question but consider the remaining three alternatives equiprobable, then they will arbitrarily choose one alternative from amongst the three. Since the choice is arbitrary, the two ideal thinkers might choose different answers, and thus disagree.

Even if one is willing to consider the above case as a disagreement, one is likely to want a more substantial form of disagreement; a case where one ideal thinker believes p and the other believes $\neg p$.[14]

Suppose Art and Bob are twin ideal thinkers as much alike as two persons can be. Both know the other is an ideal thinker, know they both know, and so on. They thus know each other to be perfect logicians, who never forget, who evaluate evidence in the same correct way, and in general, conduct themselves just as ideal thinkers should. Further, they know and agree with one another "attitudinally". Each

deems Harry Higher and Larry Lower to be authorities on matters concerning the national lottery. However, Higher is believed by Art and Bob to be more reliable than Lower in the sense that whenever Higher and Lower make conflicting claims, Higher is more likely to be correct than Lower. Monday, Higher tells the twins:

(15) Winners of the last lottery will not believe they won until Thursday.

When Art and Bob are asked whether they believe (15), they respond affirmatively since Higher is an authority. To keep clear of ambiguity and concentrate on the relevant logical structure of their beliefs, I shall employ symbolic paraphrases. Where 'a' denotes Art and 'b' denotes Bob, one so far has Ba(x)(Wx Bx-Wx) & Bb(x)(Wx Bx-Wx). Actually, they believe a bit more than this since they are asked together, hear each other's answers, and so are aware what the other believes. Tuesday, Lower tells the twins:

(16) Art is a winner of the last lottery.

Since Art and Bob are perfect logicians, they realize that if (15) and (16) are both true, then

(17) Art is a winner of the lottery but will not believe so until Thursday.

When Art is asked whether he believes (16), he answers that he disbelieves (16) and still believes (15) since Higher is reliable. Given that Art believes (15), one can understand why he disbelieves (16) by considering his other options. If he answered that he believed (16), Art would have an indefensible belief: Ba(Ba(x)(Wx

disbelieved (16), indefensibility would again arise: Ba(Ba(x)(Wx Bx-Wx) & -Ba-Wa & BaWa). Since perfect logicians do not have indefensible beliefs and disbelief is the only defensible doxastic attitude, Art must answer that he disbelieves (16).

Since Bb(Wa & Ba-Wa) is defensible, Bob can believe (17), so he is not forced to reject either (15) or (16). When asked, he asserts that he believes (15), (16), and (17). Thus, given the same evidence, twin ideal thinkers have formed opposing beliefs, Ba-Wa & BbWa. What error could Art accuse Bob of making and vice versa? The fact that Art disagrees does not weaken Bob's confidence. If anything, Art's inability to believe (16) enhances the probability of (15)-(17), for Art's disbelief insures that the necessary condition of winning obtains. Although Art is well-aware of this reasoning, he cannot be persuaded by it. Nevertheless, he cannot find any flaw in it, so he does not disparage Bob's opinion. Bob, like Art, is familiar with Moore's problem and so views Art's opinion as irreproachable as well. They know they disagree but can find nothing to argue about. Each has conducted himself as an ideal thinker should. To argue otherwise on the grounds that ideal thinkers cannot disagree if they have the same evidence and know they are in disagreement, is to beg the question. They disagree by virtue of their different identities.

Disagreement amongst ideal thinkers, here, is due to the asymmetric credibility of Moorean sentences. 'It is raining but I do not believe it' cannot be believed by me without susceptibility to belief-based criticism. However, you can believe that 'It is raining

but Roy Sorensen does not believe it' without this susceptibility. If an asymmetrically credible sentence is a consequence of another (possibly non-Moorean) sentence, this sentence will also be incredible. Thus Art and Bob can find (16) asymmetrically credible despite the fact that it is not a Moorean sentence. Notice that Art can believe (15)-(17) on Friday. 'It was raining but I did not believe it' can be believed by me since believing it does not make me susceptible to belief-based criticism. Just as the breakdown of the inter-agent access principle suggests the possibility of Moorean disagreement between ideal thinkers, the breakdown of the intertemporal access principle suggests the possiblity of Moorean disagreement between one's "future and past selves".

This kind of disagreement is bad news for those who find comfort in the thought that all disagreements are, at least in principle, resolvable. Here, the price of agreement is inconsistency for one of the parties. Ideal thinkers can never pay this price. Some philosophers have sought more than comfort in the remediality of disagreement. Ideal observer theories are a good example. In the past some philosophers have said that one ought to do what an ideal observer would think ought to be done. Thus moral problem solving is a matter of agreeing with the ideal observer. The possibility of Moorean disagreement amongst ideal observers raises the question 'Which of the ideal observers should I side with?'. More seriously, Moorean disagreement can arise between oneself and all ideal

In this case, it would be impossible for you to consistently follow their advice. Thus, one would have to solve one's moral problem without attempting to agree with an ideal observer. So moral problem solving cannot always be a matter of trying to agree with someone. A contemporary casualty of this consequence is Rawls' device of the original position. Rawls tries to answer certain questions concerning justice by having us imagine that a group of people convene to lay down the rules for governing society. To insure impartiality, these people are stipulated to be ignorant of their station and circumstances in the society. Rawls thinks that we should answer our questions about justice by emulating the opinions of this group in the original position. But the possibility of Moorean disagreement between us and them can make this prescription impossible to fulfill.

Although I have expressed reservations about Scriven's argument against predictive determinism, I agree that predictive determinism is false. Predictive determinism is the thesis that that any proposition about an event can be known prior to the occurrence of the event. Let 'Ape' read 'p is a true proposition about event e' and 'Let'read 'Event e occurs later than t'. Predictive determinism can then be expressed as

(18) $(p)(e)(t)(\exists x)\ (\Box(Ape\ \&\ Let)\rightarrow\Diamond Kxtp)$.

or perhaps as

(19) $(p)(e)(\exists t)(\exists x)\ (\Box(Ape\ \&\ Let)\rightarrow\Diamond Kxtp)$.

In either case, epistemic blindspots can be used as counterexamples to predictive determinism.

(19) Sometime prior to the first moment anyone knew anything, there were an even number of stars.

(20) It was first discovered that the number of stars is even in the year 1933.

(21) Although no one ever found out, all the water on Earth came from Saturn.

Propositions which are universal blindspots for all times preceding the event in question cannot be predicted. So predictive determinism is false. Of course, this does not imply that causal determinism is false. Thus the wedge which Scriven sought to drive between causal and predictive determinism has been driven by epistemic blindspots.

Similar counterexamples to retrodictive determinism can be constructed. Let 'Eet' read 'Event e occurs earlier than t'. Retrodictive determinism can then be expressed as

(22) $(p)(e)(t)(\exists x)(\Box(Ape \ \& \ Eet) \rightarrow \Diamond Kxtp)$,

(23) $(p)(e)(\exists t)(\exists x)(\Diamond(Ape \ \& \ Eet) \rightarrow \Diamond Kxtp)$.

In either case, epistemic blindspots can be used as counterexamples to retrodictive determinism.

(24) Sometime after the last moment anyone knows anything, there will be an even number of stars.

In addition to (24), (21) is a counterexample to retrodictive determinism as well as predictive determinism.

Notes

1-Richard Taylor "Deliberation and Foreknowledge", American Philosophical Quarterly, vol. 1 no. 1, (January, 1964).

2-Ibid.

3-Carl Ginet "Can the Will be Caused?", Philosophical Review, (January, 1962), p. 50.

4-Ibid., pp. 53-54.

5-Andrew Oldenquist "Causes, Predictions and Decisions", Analysis, vol 24 no. 3, (January, 1964), p. 55.

6-Ibid.

7-D. M. MacKay "On the Logical Indeterminacy of a Free Choice", Mind, vol. LXIX, (January, 1960), pp. 35-36.

8-Ibid., p. 36.

9-Ibid., p. 37.

10-Ibid., pp. 38-39.

11-Ibid., p. 39.

12-David K. Lewis and Jane Shelby Richardson, "Scriven on Human Unpredictability", Philosophical Studies, vol. XVII no. 5, (October, 1966), pp. 70-71.

13-Ibid., pp. 72-73.

14-With a few minor changes, the following story of the disagreement between Art and Bob is taken from my "Disagreement amongst Ideal Thinkers", Ratio, vol. XXII no. 2, (December, 1981), pp. 136-138. I thank the editor of Ratio for permission to use this material here.

CHAPTER VII

POST-DECISIONAL BLINDSPOTS: A SOLUTION TO NEWCOMB'S PROBLEM

Predictive determinism has also been challenged by an appeal to Newcomb's problem. This problem was first analyzed in Robert Nozick's "Newcomb's Problem and Two Principles of Choice".[1] The problem involves a chooser and a predictor. The chooser is shown two boxes. One box is transparent and contains one thousand dollars. The other is opaque and contains one million dollars just in case the predictor has long ago predicted that the chooser will take only the opaque box. However, the chooser also knows that the predictor is highly successful at predicting the chooser's decisions. Maya Bar-Hillel and Avishai Margalit suggest that the chooser might know this from having played the same game with the predictor many times before for points instead of money.[2] Newcomb's problem is the problem of determining whether the rational chooser decides to take only the opaque box or decides to take both boxes.

Nozick points out that Newcomb's problem seems to show that two principles of choice can conflict. If the chooser follows the principle of maximizing expected utility, then he will take only the opaque box since its expected utility is nearly a million dollars whereas the expected utility of taking both boxes is not much more than one thousand dollars. If the chooser follows the dominance principle, then he will take both boxes. If the million is in the opaque box, then the two-boxer has a million plus a thousand dollars while the one-boxer has only a million dollars. If the opaque box is empty, then the two-boxer has only a million dollars.

In either case, the two-boxer has more money.

Reservations have been expressed about the dominance argument since the dominance principle does not apply when the decision itself has a bearing on the probabilities of the alternative states. In "The Unpredictability of Free Choices", George Schlesinger uses a different argument.[3] Suppose Smith is a perfect well-wisher of mine who always advises me to do what is my own best interests. Smith can see whether there is any money in the opaque box. If he sees that there is a million dollars in the opaque box, then Smith surely advises me in his heart to take both boxes so that I gain $1,001,000 rather than $1,000,000. If he sees that the opaque box is empty, then he surely advises me in his heart to take both boxes so that I gain $1,000 rather than nothing. Since it is analytically true that a perfect well-wisher of mine advises me to what is in my own best interests, and he would advise me to take both boxes, then choosing both boxes is in my own best interests.

In "Perfect Diagnosticians and Incompetent Predictors", Schlesinger offers another argument for choosing both boxes: the "invinciple player's strategy argument". Suppose a player of Newcomb's game is allowed to peek inside the opaque box and thus knows whether the million dollars is there. If there is a million dollars in the opaque box, the player takes both boxes and so has $1,001,000. If the opaque box is empty, the player takes both boxes and so has at least $1,000. Thus the strategy of this invincible player is to always take both boxes. An ordinary player of Newcomb's game is unable to peek inside the opaque box. Nevertheless, he can copy the

moves of the invinciple player. Thus an ordinary player should take both boxes.

In the invincibility argument, Schlesinger tacitly appeals to the following sufficient condition for deciding to take both boxes.

(1) Necessarily, if at any time prior to either taking only the opaque box or taking both boxes a knows that there is a million dollars in the opaque box, then a finally decides to take both boxes.

If (1) is conjoined with the predictor's infallibility and the chooser's knowledge of whether there is a million dollars in the opaque box, then it follows that the predictor has not put any money in the opaque box. Given (1), it is impossible for the chooser to know the conjunction that the predictor is infallible and that there is a million dollars in the opaque box. Nevertheless, the conjunction is consistent; it is an epistemic blindspot.

Schlesinger also considers the expected utility argument for choosing only the opaque box to be valid. He argues that once the possibility of an infallible predictor is accepted, we are forced to the antinomy of a rational agent choosing and not choosing to take both boxes. To escape the antinomy we must reject the possibility of an infallible predictor. Schlesinger extends this rejection to include rejection of the predictor being even slightly reliable. For if the predictor has some degree of success, then by choosing to take only the opaque box the chooser can raise the probability of there being money in the box. Then, no matter how small the rise is, one can make the expected utility of taking only the opaque box higher

than taking both by varying the amount of money which might be in the opaque box.

Although I do not accept Schlesinger's argument, I think that his attempts to shore up the two-boxer's case suggest a reductio ad absurdum of the one-boxer's position. Both one-boxers and two-boxers agree that given that a is an ideal chooser in the Newcomb situation

(2) Either necessarily a finally decides to take only the opaque box or necessarily a finally decides to take both boxes.

A crucial premiss for one-boxers is that

(3) Necessarily, a knows that if he finally decides to take only the opaque box, then there is a million dollars in the opaque box.

The whole point of having a know that the predictor is very reliable is to ensure that (3) can explain why

(4) Necessarily, a finally decides to take only the opaque box.

One-boxers are also committed to general truths about decision-making. One might be tempted to accept the following as one such general truth.

(5) Necessarily, if a finally decides to perform a particular act, then there is some time prior to the time of the intended act at which a knows that his final decision is to perform that act

However, consider the following variation of Newcomb's problem. The chooser is shown a blue box and a brown box. He must choose exactly one of them. A predictor has put a million dollars in the blue box just in case he predicted that the chooser will choose the brown box and he has put a million dollars in the brown box just in case he predicted that the chooser will choose the blue box. The chooser

knows all this and also knows that the predictor is highly successful. Which box should the chooser take?

The answer is that the chooser has no more grounds for choosing one way rather than the other. Further, the chooser cannot know what his final decision is until he acts. If he did, then he would have sufficient grounds for changing his mind—in which case the decision could not have been final. So the chooser's final decision is an epistemic blindspot for him between the time he makes it and the time he acts on it. Someone else can know what the chooser's final decision is during this period, but the chooser cannot. If this knowledgeable outsider was given the choice between a blue and a brown box which he knew contain amounts of money equal to that contained in the like-colored first chooser's boxes, the outsider's decision would be opposite to the first chooser's final decision.

So (5) is false as it stands. But strengthening the antecedent provides a way of avoiding the counterexample.

(5') Necessarily, if a finally decides to perform a particular act because he had better grounds for performing that act rather than any other, then there is some time prior to the time of the intended act at which a knows that his final decision is to perform that act.

One might object that (5') is also too strong since an ideal chooser might not know he is an ideal chooser and so not know that his final decision must conform to the conclusion of any sound argument concerning what one should do in his position. For my purposes,

however, (5') can be further weakened to

(5'') Necessarily, if a finally decides to perform a particular act because he has better grounds for perfroming that act rather than any other, then it is logically possible that there is some time prior to the time of the intended act at which a knows that his final decision is to perform that act.

The one-boxer is committed to (1)-(4) and the above weakened principle of self-awareness. However, the conjunction of (1)-(5'') is inconsistent. By (4) and (5''), it follows that

(6) It is logically possible that there is a time prior to a's taking only the opaque box at which a knows that he has finally decided to take only the opaque box.

From (3) and (6), it follows that

(7) It is logically possible that there is a time prior to a's taking only the opaque box at which a knows that there is a million dollars in the opaque box.

However, this knowledge would activate the sufficient condition for deciding to take both boxes, (1), so

(8) It is logically possible that a finally decides to take both boxes.

But then

(9) It is not the case that necessarily, a takes only the opaque box,

which contradicts (4). So at least one of (1), (3), (4), (5'') must be rejected.

The one-boxer's position requires that the chooser's final decision be a blindspot to him between the time he makes his final decision and the time the chooser acts. In this respect, Newcomb's problem resembles the blue box/brown box problem. However, according to both one-boxers and two-boxers, there is the difference that the ideal chooser has grounds for deciding one way rather than another. They just disagree as to which way the ideal chooser decides. What makes the one-boxer's position untenable is the conjunction of his blindspot commitment and his claim that his solution is sound. For if it is a sound solution to the decision problem, it must be possible for the decision-maker to know it is sound. This is a general requirement for decision and game theory.

Showing that the one-boxer is wrong is not equivalent to showing that the two-boxer is correct. For it may be the case that a proposition that they both agree on, (2), is false. If (2) is false, then the resemblance between Newcomb's problem and the blue box/brown box problem is further strengthened. However, our chief reservations about the two-boxer's argument was the existence of the one-boxer's expected utility argument. In the face of an antinomy, one should be suspicious of both arms of the paradox. But given the collapse of the one-boxer position, and thus one arm of the antinomy, the two-boxer's dominance argument and its refinements are sufficient to establish the conclusion that the ideal chooser must finally decide to take both boxes.

Notes

1-This article appears in Essays in Honor of Carl G. Hempel, ed. by N. Rescher, (Dordrecht: Reidel, 1969).

2-This suggestion is made in their "Newcomb's Problem Revisited", British Journal for the Philosophy of Science, vol. 23, no. 3, 1972.

3-British Journal for the Philosophy of Science, vol. 25, no.3, 1974.

4-Australasian Journal of Philosophy, vol. 54, no. 3, 1976.

CONCLUDING REMARKS

Epistemic blindspots bridge Moore's problem and the prediction paradox and show that there are unfamiliar limits to knowledge. Solving the prediction paradox is a matter of familiarizing oneself with these new limits for epistemology. Although my solution may seem strange, my approach follows the pattern best exemplified by Kant and Wittgenstein. Like them, I diagnose the problems that concern me as the result of the transgression of epistemological limits. Like them, I consider the solution to lie in familiarizing oneself with these limits. Unlike them, I have not presented by solution as a philosophical panacea. Nevertheless, my analysis is ambitious. Negatively, I attempt to show, among other things, that all previous proposed solutions to the prediction paradox fail, that recent attempts to prove that decisions are uncaused fail, and that logic cannot have epistemic foundations. Positively, I attempt to solve three recent philosophical problems in a unified and open-ended fashion. Further, I think a metaphilosophical moral can be drawn. Before studying the prediction paradox, I had pictured philosophy as an impersonal affair--that being who you are is philosophically irrelevant to which position you should or could adopt. Yet, with the breakdown of the interagent access principle, who you are becomes philosophically relevant. Now, a subjectivistic shadow is cast on this picture of philosophy.

BIBLIOGRAPHY

BIBLIOGRAPHY

Alexander, P. "Pragmatic Paradoxes." Mind 59 (October 1950): 536-538.

Austin, A. K. "On the Unexpected Examination." Mind 78 (January 1979): 137.

_____. "The Unexpected Examination." Analysis 39 (January 1979): 63-64.

Bennett, Jonathan. "Review of R. Shaw's 'The Paradox of the Unexpected Examination,' G.C. Nerlich's 'Unexpected Examinations and Unprovable Statements,' Brian Medlin's 'The Unexpected Examination,' and Frederic Fitch's 'A Goedelized Formulation of the Prediction Paradox'." The Journal of Symbolic Logic 30 (June 1965): 101-102.

Binkley, Robert. "The Surprise Examination in Modal Logic." The Journal of Philosophy LXV (March 7, 1968): 127-136.

Bosch, Jorge. "The Examination Paradox and Formal Prediction." Logique et Analyse 15 (September-December, 1971): 505-525.

Butler, R. J. and Chapman, J. M. "On Quine's 'So-called Paradox'." Mind LXXIV (July 1965): 424-425.

Cargile, James. "The Surprise Test Paradox." The Journal of Philosophy LXIV (September 21, 1967): 550-563.

_____. "Review of David Kaplan's and Richard Montague's 'A Paradox Regained,' Martin Gardner's 'A New Prediction Paradox,' and K.R. Popper's 'A Comment on the New Prediction Paradox." The Journal of Symbolic Logic 30 (June 1965): 102-103.

Champlin, T. S. "Quine's Judge." Philosophical Studies 29 (May 1976): 349-352.

Cohen, L. J. "Mr. O'Connor's 'Pragmatic Paradoxes'." Mind LVIV (January 1950): 85-87.

Cole, David. "Meaning and Knowledge." Philosophical Studies 36 (October 1979): 329-331.

Deutscher, Max. "Bonney on Saying and Disbelieving." Analysis 27 (June 1967): 184-186.

Dietl, Paul J. "The Surprise Examination." Educational Theory 23 (September 1973): 153-158.

Edman, Martin. "The Prediction Paradox." Theoria XL (November 1974): 165-175.

Ellis, Brian. "Epistemic Foundations of Logic." The Journal of Philosophical Logic 5 (May 1976): 187-203.

_____. Rational Belief Systems. Oxford: Basil Blackwell, 1979.

Fitch, F.B. "A Goedelized Formulation of the Prediction Paradox." American Philosophical Quarterly 1 (April 1964): 161-164.

Gardner, Martin. "A New Prediction Paradox." The British Journal for the Philosophy of Science 13 (May 1962): 51.

Gauthier, David P. "How Decisions are Caused." The Journal of Philosophy LXIV (March 16, 1967): 147-151.

Ginet, Carl. "Can the Will be Caused?" The Philosophical Review 24 (January 1962): 49-55.

Harrison, Craig. "The Unanticipated Examination in View of Kripke's Semantics for Modal Logic." Philosophical Logic Edited by J.W. Davis et. al. Dordrecht: Reidel, 1969: 74-88.

Hintikka, Jaakko. Knowledge and Belief. Ithaca: Cornell University Press, 1962.

Kaplan, David and Montague, Richard. "A Paradox Regained." Notre Dame Journal of Formal Logic 1, (July 1960): 79-80.

Kiefer, James and Ellison, James. "The Prediction Paradox Again." Mind LXXIV (July 1965): 426-427.

Kvart, Igal. "The Paradox of Surprise Examination." Logique et Analyse 21 (June-September 1978): 337-344.

Lewis, Clarence Irving and Langford, Cooper Harold. Symbolic Logic 2nd ed. New York: Dover Publications, 1959.

Lewis, David K. and Richardson, Jan Shelby. "Scriven on Human Unpredictability." Philosophical Studies XVII (October 1960): 31-40.

Lyon, Ardon. "The Prediction Paradox." Mind LXVII (October 1959): 510-517.

MacKay, D.M. "On the Logical Indeterminacy of a Free Choice." Mind LXIX (January 1960): 31-40.

Margalit, Avishai and Bar-Hillel, Maya. "Newcomb's Problem Revisited." The British Journal for the Philosophy of Science 23 (November 1972): 295-304.

McLelland, J. "Epistemic Logic and the Paradox of the Surprise Examination." International Logic Review 3 (June 1971): 69-85.

_____ and Chihara, T. S. "The Surprise Examination Paradox." The Journal of Philosophical Logic 4 (February 1975): 71-89.

Medlin, Brian. "The Unexpected Examination." American Philosophical Quarterly 1 (January 1964): 66-72.

Nerlich, G. R. "Unexpected Examinations and Unprovable Statements." Mind LXX (October 1961): 503-513.

Nozick, Robert. "Newcomb's Problem and Two Principles of Choice." Essays in Honor of Car G. Hempel Edited by N. Rescher. Dordrecht: Reidel, 1969: 114-146.

O'Carroll, M. J. "Improper Self-Reference in Classical Logic and the Prediction Paradox" 10 (June 1967): 167-172.

O'Connor, D. J. "Pragmatic Paradoxes." Mind LVII (July 1948): 358-359.

Oldenquist, Andrew. "Causes, Predictions and Decisions." Analysis 24 (January 1964): 55-58.

Popper, K.R. "A Comment on the New Prediction Paradox." The British Journal for the Philosophy of Science 13 (May 1962): 51.

Quine, W. V. "On a so-called Paradox." Mind LXII (January 1953): 65-67.

Schiffer, Stephen R. Meaning. Oxford: Oxford University Press, 1972.

Schlesinger, George. "The Unpredictability of Free Choices." The British Journal for the Philosophy of Science 25 (September 1974): 209-221.

_____. "Perfect Diagnosticians and Incompetent Predictors." The Australasian Journal of Philosophy 54 (December 1976): 221-230.

Schoenberg, Judith. "A Note on the Logical Fallacy in the Paradox of the Unexpected Examination." Mind (January 1966): 125-127.

Scriven, Michael. "Paradoxical Announcements." Mind LX (July 1951): 403-407.

_____. "An Essential Unpredictability in Human Behavior." Scientific Psychology Principles and Approaches. Edited by Benjamin B. Wolman and Ernest Nagel. New York: Basic Books, 1964: 411-425.

Sharpe, R. A. "The Unexpected Examination." Mind LXXIV (April 1965): 255.

Slater, B. H. "The Examiner Examined." Analysis 35 (December 1974): 49-50.

Smullyan, Raymond. "Languages in which self reference is possible." The Journal of Symbolic Logic 22 (March 1957): 56-67.

Sorensen, Roy. "Disagreement amongst Ideal Thinkers." Ratio XXIII (December 1981): 136-138.

_____. "Recalcitrant Variations of the Prediction Paradox." Australasian Journal of Philosophy (forthcoming, probably December 1982).

Taylor, Richard. "Deliberation and Foreknowledge." American Philosophical Quarterly 1 (January 1964): 73-80.

Weiss, Paul. "The Prediction Paradox." Mind LXI (April 1952): 265-269.

Williams, J. A. "Moore's paradox: one or two?" Analysis 39 (June 1979): 141-142.

Windt, P. "The Liar in the Prediction Paradox." American Philosophical Quarterly 10 (January 1973): 65-68.

Wright, Crispin and Sudbury, Aiden. "The Paradox of the Unexpected Examination." The Australasian Journal of Philosophy 55 (May 1977): 41-58.

Wright, J. A. "The Surprise Exam: Prediction on the Last Day Uncertain." Mind LXXVI (January 1967): 115-117.

Wu, Kathleen Johnson. "Believing and Disbelieving." The Logical Enterprise. Edited by Alan Ross Anderson, Ruth Barcan Marcus, and R. M. Martin. New Haven: Yale University Press, 1975: 211-219.